Study of the choice behaviour of travellers in a transport network via a "simulation game"

Humberto Gonzalez Ramirez

N° d'ordre NNT: 2020LYSET008

Thèse de doctorat de l'université de Lyon

Opérée au sein de
l'École Nationale des Travaux Publics de l'État

École Doctorale N° 162
MEGA (Mécanique, Energétique, Génie Civil et Acoustique)

Spécialité / discipline de doctorat : Génie Civil

Étude des choix des usagers dans les réseaux de transport grâce à un "jeu de simulation"

Soutenue publiquement le 13 mai 2020 par :

Humberto González Ramírez

Devant le jury composé de :

Emma FREJINGER	Prof. (Université de Montréal)	Rapporteure
Francesco VITI	Prof. (Université du Luxembourg)	Rapporteur
Christine SOLNON	Prof. (INSA Lyon)	Présidente du jury
Eric GONZALES	Prof. (University of Massachusetts Amherst)	Examinateur
Ludovic LECLERCQ	Prof. (Univ. Lyon, ENTPE, IFSTTAR)	Directeur de thèse
Nicolas CHIABAUT	Prof. (Univ. Lyon, ENTPE, IFSTTAR)	Co-directeur de thèse

Thèse préparée au LICIT (Laboratoire Ingénierie Circulation Transports)

Thesis National Number: 2020LYSET008

A THESIS OF THE UNIVERSITY OF LYON

Prepared at
École Nationale des Travaux Publics de l'État

Doctoral school No. 162
MEGA (Mechanics, Energy, Civil Engineering and Acoustics)

To obtain the graduation of
PhD in Civil Engineering

Study of the choice behaviour of travellers in a transport network via a "simulation game"

Defended on May 13, 2020 by:

Humberto González Ramírez

In front of the following examination committee:

Emma FREJINGER	Prof. (Université de Montréal)	Reviewer
Francesco VITI	Prof. (Université du Luxembourg)	Reviewer
Christine SOLNON	Prof. (INSA Lyon)	Committee president
Eric GONZALES	Prof. (University of Massachusetts Amherst)	Examiner
Ludovic LECLERCQ	Prof. (Univ. Lyon, ENTPE, IFSTTAR)	Supervisor
Nicolas CHIABAUT	Prof. (Univ. Lyon, ENTPE, IFSTTAR)	Cosupervisor

LICIT lab (Transport and Traffic Engineering Lab)

Abstract

Over the past century, cities around the world have experienced rapid growth in their populations. As a consequence, the capacity of their transport networks has been exceeded, causing traffic congestion problems that are associated to economic losses, high levels of air pollution and the emission of greenhouse gases. It is in this context that it is important to implement traffic regulation strategies to mitigate congestion. Traffic simulators are essential tools for developing and testing such strategies. The traffic conditions in the transport network are consequence of the travellers' choices at any given time, but at the same time, the traffic conditions on the network influence the decisions of the travellers. This relationship between travellers' decisions and network traffic is the main idea in traffic assignment, which traffic simulators use to produce network traffic patterns. It is in this sense that understanding and approximating the behaviour of travellers is indispensable to forecast the states of a transportation network, which could help design regulations to alleviate congestion.

The objective of this thesis is to find route choice models that scale-up at network level, i.e., models that predict the choices of travellers over the diversity of situations found in a transport network. The approach in this thesis to investigate travellers' behaviour in transportation networks is through computer-based experiments at large scale, for which a platform named the Mobility Decision Game (MDG), has been developed. The MDG permits to observe the choices of the participants on a diverse set of scenarios (OD pairs and routes) with varying traffic conditions and travel time information. In this thesis, the experiments focus on the route choices of uni-modal car trips that are based on the map of the city of Lyon, France. To attain the objective of this thesis, firstly a methodology to find OD pairs that are representative of the whole network is proposed. The representative OD pairs are then used in route choice experiments to obtain data to estimate choice models that generalise to the various OD pair configurations in the network. Secondly, the choices of participants in the experiments are analysed from the rational and boundedly rational behaviour perspectives in order to establish the principle that best describe their choices. Finally, the choice models are assessed in terms of their predictive accuracy. This thesis is part of a European ERC project entitled MAGnUM: Multiscale and Multimodal Traffic Modeling Approach for Sustainable Management of Urban Mobility.

Résumé

Au cours du siècle dernier, les villes du monde entier ont connu une croissance rapide de leur population. En conséquence, la capacité de leurs réseaux de transport a été dépassée, entraînant des problèmes de congestion du trafic qui sont à leur tour associés à des pertes économiques, à des niveaux élevés de pollution atmosphérique et à l'émission de gaz à effet de serre. C'est dans ce contexte qu'il est important de mettre en œuvre des stratégies de régulation du trafic pour atténuer les encombrements. Les simulateurs de trafic sont essentiels au développement et au test de telles stratégies. Les conditions de circulation dans le réseau de transport sont la conséquence des choix des voyageurs à un moment donné, mais en même temps, les conditions de circulation sur le réseau influencent les décisions des voyageurs. Cette relation entre les décisions des voyageurs et le trafic réseau est l'idée principale de l'attribution du trafic, que les simulateurs de trafic utilisent pour produire des modèles de trafic réseau. C'est en ce sens que la compréhension et l'approximation du comportement des voyageurs sont indispensables pour prévoir les états d'un réseau de transport, ce qui pourrait aider à concevoir des réglementations pour réduire la congestion.

Lorsqu'ils voyagent dans une ville, les voyageurs sont confrontés à plusieurs décisions concernant (i) l'activité à entreprendre, (ii) la destination du voyage, (iii) le mode de transport, (iv) l'heure de départ et (v) l'itinéraire. Ces décisions individuelles façonnent les états du trafic dans le réseau de transport à un moment donné. Cependant, en même temps que les états du réseau sont la conséquence des choix des voyageurs, les caractéristiques et les états de trafic du réseau influencent les décisions des voyageurs. On pourrait penser, par exemple, aux voyageurs qui adaptent leur heure de départ et leurs choix d'itinéraire en anticipant des embouteillages sur la route habituelle. Cette relation bidirectionnelle entre le comportement des voyageurs et la dynamique du trafic est l'idée centrale des problèmes d'affectation du trafic et elle est étudiée sous deux angles différents. Le premier est lié au problème d'affectation du trafic, dans lequel les modèles de trafic au niveau du réseau sont obtenus à la suite des interactions entre les choix de tous les voyageurs dans le réseau de transport, sur la base d'hypothèses générales sur l'équilibre du réseau. Le second est lié aux choix des voyageurs, avec un accent particulier sur l'identification et la classification des facteurs qui influencent leurs choix.

La motivation de cette thèse est d'obtenir des modèles de choix à la fois cohérents avec

le comportement observé des voyageurs et généralisant bien au niveau du réseau à grande échelle. En d'autres termes, des modèles qui font des prédictions précises dans toutes les paires origine-destination (OD) dans un réseau de transport urbain, mais qui sont estimés avec des observations sur un nombre limité de paires OD. L'idée générale est de mettre en œuvre, à l'avenir, ces modèles de choix dans des algorithmes d'affectation dynamique du trafic, et ainsi d'obtenir des schémas de trafic simulés dans une perspective comportementale. Les modèles de choix sont déduits des données collectées par des expériences informatiques réalisées sur une plateforme informatique Mobility Decision Game (MDG). Cependant, comme les choix des voyageurs ne peuvent pas être observés dans un si grand nombre de paires de OD et d'itinéraires que l'on trouve dans un réseau urbain, une méthodologie pour collecter les observations des choix des voyageurs par des expériences informatiques est d'abord nécessaire. Les modèles estimés avec des données provenant de cette méthodologie devraient pouvoir évoluer, d'un petit ensemble de paires de OD observables, jusqu'au niveau du réseau complet. D'une manière générale, les objectifs de cette thèse sont :

1. **Framework for the estimation of choice models at full-scale network level.** Concerne la conception d'expériences, l'expérimentation et la généralisation des choix observés au réseau grandeur nature. Celui-ci comprend l'échantillon d'un petit ensemble de paires OD et d'itinéraires du réseau routier de la ville de Lyon en France, sur lesquels se déroulent les expériences d'itinéraires et d'heures de départ. L'échantillon doit être représentatif du réseau routier, et les choix observés sur ces paires OD et ces itinéraires doivent se généraliser au niveau du réseau à grande échelle. En d'autres termes, les modèles de choix estimés avec des données sur les petites paires et itinéraires OD représentatifs doivent approximer avec une bonne précision les choix dans l'ensemble du réseau.

2. **Spécification et sélection d'un modèle de choix.** Il s'agit de trouver, sur la base des résultats de l'analyse quantitative des résultats des expériences sur les MDG, un modèle de choix d'itinéraire et d'heure de départ qui fait des prédictions précises. L'impact des attributs des paires OD et des itinéraires, les informations de temps de trajet fournies aux participants, ainsi que l'hétérogénéité des préférences des voyageurs sont pris en compte par les modèles.

Afin d'estimer et de tester différents modèles de choix, le comportement des voyageurs doit être observé. L'approche de cette thèse pour étudier le comportement des voyageurs dans les réseaux de transport passe par des expériences informatiques à grande échelle, pour lesquelles une plateforme nommée Mobility Decision Game (MDG) a été développée. Contrairement à d'autres expériences préférence déclaré, le MDG permet d'observer les choix d'un grand nombre de participants dans différentes situations (voir Fig. 1), ce qui est essentiel pour inférer des modèles de choix au niveau du réseau à grande échelle. Le MDG est un jeu informatique basé sur le Web conçu pour confronter les participants à des problèmes de décision concernant l'heure de départ et les choix d'itinéraire pour effectuer un voyage. Les problèmes de décision sont placés sous différents scénarios hypothétiques, comprenant des paires OD reliées par trois itinéraires alternatifs avec des attributs contrastés et des conditions de trafic variables. Les scénarios se produisent dans un environnement simulé dynamiquement du réseau réel de la ville de Lyon, en France, dans lequel les conditions de circulation sont générées en temps réel par un seul simulateur microscopique dynamique. Au cours d'une expérience MDG, plusieurs paires de OD sont attribuées aux participants, ce qui permet d'observer les choix des mêmes participants dans différentes paires de OD. En outre, certains des participants peuvent recevoir des informations sur la circulation sous forme de temps de trajet.

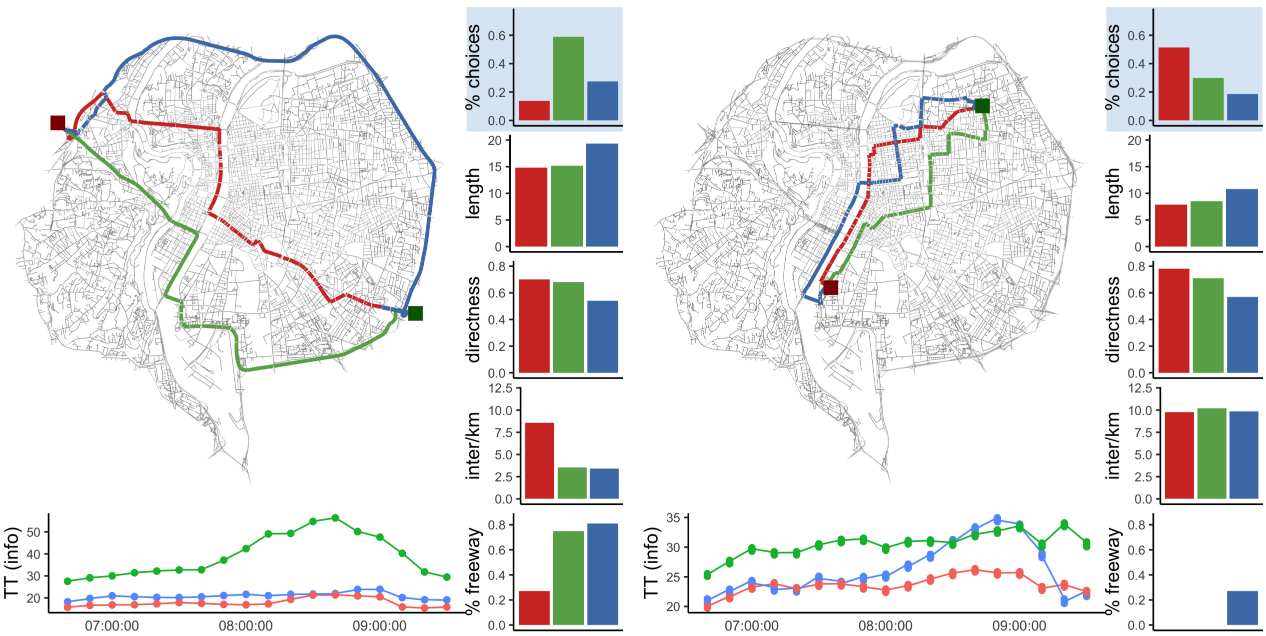

Figure 1: Deux paires OD jointes par trois voies pour l'expérience MDG. Les expériences consistent à choisir un itinéraire sur trois pour effectuer un voyage. Les conditions de circulation et les attributs d'itinéraire variables permettent d'observer les choix des voyageurs dans différentes situations sur la ville de Lyon en France.

Le modèle adopté dans cette thèse pour prédire les choix des voyageurs est le modèle mixed logit (McFadden, 1984, McFadden and Train, 2000), qui appartient au cadre de maximisation de l'utilité aléatoire (Mcfadden, 1972). Ce cadre suppose que les individus obtiennent un certain niveau d'utilité de chaque alternative dans une situation de choix, et qu'ils choisissent l'alternative avec l'utilité maximale, c'est-à-dire que les individus sont des maximiseurs d'utilité. Dans les modèles de maximisation d'utilité aléatoires, l'utilité obtenue à partir d'une alternative est liée aux attributs de l'alternative et du décideur. En conséquence, les choix des décideurs s'expliquent par des variables mesurables et qui dépendent de la paire OD et de la situation (itinéraires et conditions de trafic). De plus, le nombre d'hypothèses sur les paramètres du modèle est petit, et celles-ci peuvent être facilement déduites des données, obtenant des représentations succinctes de l'utilité et des probabilités de choix. Par conséquent, les modèles de maximisation de l'utilité aléatoire sont bien adaptés pour la prédiction des choix des voyageurs sur un réseau à grande échelle. Une sélection rigoureuse des scénarios sur lesquels évoluent les expériences des MDG permet d'estimer des modèles de choix qui se généralisent bien au niveau du réseau à grande échelle, et de tester et sélectionner la spécification du modèle qui représente le mieux les choix des participants aux expériences.

Plan de la thèse

Cette thèse comprend quatre études sur le comportement des voyageurs. Ces études sont toutes basées sur les données obtenues par des expériences informatisées de parcours et d'heures de départ, réalisées avec la plateforme Mobility Decision Game. Le deuxième chapitre de la thèse présente cette plateforme. Par la suite, les quatre études sont présentées, organisées en deux parties : la première consacrée au comportement de choix d'itinéraire et la seconde au comportement de choix d'itinéraire commun et d'heure de départ. Une brève description des chapitres de la thèse est donnée ci-dessous.

Chapitre 2. Ce chapitre présente le MDG, une plate-forme informatique (développée par le laboratoire LICIT) pour étudier le comportement des voyageurs dans les réseaux de transport à grande échelle. Le MDG est un jeu informatique basé sur le Web conçu pour confronter les participants à une variété de problèmes de décision concernant le mode, l'heure de départ et les choix d'itinéraire pour effectuer un voyage. Les problèmes de décision sont placés sous différents scénarios hypothétiques, c'est-à-dire des paires OD jointes par trois itinéraires alternatifs qui contrastent dans leurs attributs et avec des temps de trajet variables. Cela permet d'étudier les déterminants des décisions des participants dans différentes circonstances.

Part I: Choix d'itinéraire

Chapitre 3. Ce chapitre présente une brève revue de la littérature sur le comportement des voyageurs. L'objectif est de fournir une classification grossière des lignes de recherche sur les études qui se trouvent dans la littérature, en mentionnant certains des aspects du comportement des voyageurs qui ont été largement étudiés. Ensuite, une revue de la littérature sur les modèles de choix discrets est donnée. L'accent est mis sur la dérivation et l'estimation du modèle logit mixte, qui est l'approche de modélisation choisie dans cette thèse.

Chapitre 4. Dans un réseau à grande échelle, les voyages sont effectués dans des milliers de paires origine-destination (OD) reliées par plusieurs itinéraires, résultant en un grand nombre d'alternatives aux caractéristiques diverses qui influencent le comportement de choix d'itinéraire des voyageurs. Par conséquent, pour prédire avec précision les choix des utilisateurs à l'échelle du réseau, un modèle de choix d'itinéraire doit être évolutif pour s'adapter à toutes les configurations possibles qui peuvent être rencontrées. Dans ce chapitre, une nouvelle méthodologie pour obtenir un tel modèle est proposée. L'idée principale est d'utiliser partitionnement de données pour obtenir un petit ensemble de paires et d'itinéraires OD représentatifs qui peuvent être étudiés en détail par des expériences de choix d'itinéraire informatique pour recueillir des observations sur le comportement des voyageurs. Les résultats sont ensuite mis à l'échelle vers toutes les autres paires OD du réseau. Il a été constaté que 9 configurations de paires OD sont suffisantes pour représenter le réseau de Lyon, en France, composé de 96 096 paires OD et 559 423 itinéraires. Les observations, recueillies sur ces neuf configurations de paires OD représentatives, ont été utilisées pour estimer trois modèles logit mixtes. La précision prédictive des trois modèles a été testée par rapport à la précision prédictive des mêmes modèles (avec la même spécification), mais estimée sur des configurations de paires OD sélectionnées au hasard. Les résultats obtenus montrent que les modèles estimés avec les paires OD représentatives sont supérieurs en précision prédictive, suggérant ainsi l'extension à l'ensemble du réseau des choix des participants sur les configurations de paires OD représentatives, et validant la méthodologie de cette étude.

Chapitre 5. Des études empiriques récentes ont révélé que les choix d'itinéraire des voyageurs s'écartent de la rationalité parfaite, en montrant que les déplacements urbains ne suivent pas nécessairement les itinéraires les plus courts (Papinski et al., 2009, Thomas and Tutert, 2010, Zhu and Levinson, 2015, Hadjidimitriou et al., 2015, Yildirimoglu and Kahraman, 2018b). Cependant, il n'y a pas de consensus sur la mesure dans laquelle le comportement de choix d'itinéraire des voyageurs s'écarte de l'hypothèse rationnelle parfaite. L'objectif de cette étude est de contribuer à la compréhension de la façon dont les voyageurs traitent le temps de trajet lorsqu'ils font des choix d'itinéraire et de quantifier dans quelle mesure les utilisateurs sont des minimiseurs de temps de trajet stricts ou si une rationalité limitée est observée. La question de savoir si les voyageurs évaluent les différences de temps de voyage en termes absolus ou relatifs est également abordée, et l'hétérogénéité du comportement de choix d'itinéraire des voyageurs a été étudiée. Les résultats des expériences informatiques de choix d'itinéraire, axés sur les choix d'itinéraire dans diverses paires OD et conditions de trafic, sont analysés. Il a été constaté que les voyageurs évaluent les différences relatives plutôt qu'absolues dans le temps de trajet. Dans 60,5% des voyages, les participants ont choisi l'itinéraire le plus rapide, mais ce pourcentage est de 80% lorsque le temps de trajet des alternatives est au moins 30% supérieur à l'itinéraire le plus rapide. Seulement 10% des individus ont choisi l'itinéraire le plus rapide dans tous les voyages, confirmant l'hypothèse de rationalité limitée. Les participants présentaient des bandes d'indifférence hétérogènes en termes de temps de trajet : au moins 70% d'entre eux ne considéreraient pas les itinéraires avec des temps de trajet 1,5 fois plus lents que l'alternative la plus rapide ; le participant moyen était indifférent aux différences de temps de voyage relatives inférieures à 31%.

Chapitre 6. Dans ce chapitre, un modèle de choix qui prend en compte le comportement rationnel limité dans la génération des ensembles de choix des individus pour le choix de l'itinéraire est développé (modèle BRCS). Dans le BRCS, la distribution des bandes d'indifférence est déduite de manière endogène en estimant conjointement la génération d'ensembles de choix et les choix de route. Le modèle est proposé comme alternative au modèle logit mixte (MXL) dans le Chapitre 5, où les bandes d'indifférence estimées individuellement sont estimées de manière exogène et entrent ainsi dans le modèle en tant que variables indépendantes. Le modèle BRCS est comparé, en termes de précision prédictive, au modèle MXL en utilisant des données synthétiques et réelles, obtenues à partir des expériences avec la plate-forme MDG. Les résultats montrent que le modèle BRCS est capable d'inférer la distribution des bandes d'indifférence pour les données générées synthétiques. De plus, pour ces données, le BRCS montre une précision prédictive plus élevée que le modèle MXL. Dans le cas des données sur les MDG, le modèle BRCS présente une précision prédictive plus élevée que le MXL et le MXL avec des bandes d'indifférence estimées de manière exogène du Chapitre 5.

Conclusions de la partie I: sélection du modèle de choix d'itinéraire. En conclusion de la partie choix de l'itinéraire de la thèse, la précision prédictive est évaluée pour différentes spécifications du modèle MXL, ainsi que du modèle BRCS.

Part II: Choix d'itinéraire et d'heure de départ

Chapitre 7. Dans un réseau unimodal (trajets en voiture), l'itinéraire et l'heure de départ sont deux des principales décisions que les voyageurs prennent pour faire un voyage. Au niveau agrégé, cela implique que les modèles de trafic dans le réseau unimodal s'expliquent principalement par la somme de ces deux choix individuels. Dans ce chapitre, le comportement de choix de l'itinéraire et de l'heure de départ des voyageurs est étudié. À cette fin, une expérience qui considère les deux décisions simultanément a été réalisée à l'aide de la plate-forme MDG. L'objectif est de comprendre quelles variables influencent les choix conjoints d'itinéraire et d'heure de départ des voyageurs, et de tester la pertinence du modèle MXL pour expliquer et prédire ces choix. Le modèle conjoint proposé ici introduit des corrélations dépendantes du temps dans la spécification d'un modèle MXL. Il est important de mentionner qu'il s'agit d'une enquête en cours et que les résultats ne sont ni définitifs ni complets.

Contributions

Méthodologie pour l'estimation des modèles de choix d'itinéraire au niveau du réseau à grande échelle. Une méthodologie, basée sur le partitionnement de données, est proposée pour déterminer un sous-ensemble optimal de paires OD sur lesquelles les expériences de choix sont effectuées. Un centroïde de cluster, étant l'élément qui minimise la distance euclidienne à tous les éléments de cluster, peut être sélectionné pour représenter son groupe, et l'ensemble de tous les centroïdes de cluster comme représentatifs de l'ensemble du réseau. Les paires OD obtenues avec cette méthodologie ont trois propriétés qui les rendent adaptées aux expériences de choix de route. Premièrement, ils couvrent une grande variabilité des paires OD et des routes du réseau, c'est-à-dire qu'ils sont représentatifs du réseau à grande échelle. Deuxièmement, la variance de leurs attributs est élevée, ce qui permet d'identifier l'influence de chaque attribut dans les choix des participants. Troisièmement, le nombre de configurations OD est petit, ce qui permet de collecter suffisamment de données dans chaque paire OD. Les première et deuxième propriétés impliquent qu'un modèle de choix estimé avec des données sur les paires OD représentatives se généralisera bien au reste des paires OD dans le réseau, dans le sens que les choix dans n'importe quelle paire OD dans le réseau sont prédits avec précision. La troisième propriété implique une estimation robuste des paramètres du modèle, c'est-à-dire des erreurs d'estimation plus petites. Les modèles estimés avec les paires OD représentatives ont montré une précision prédictive plus élevée que les modèles estimés avec différents ensembles de paires OD.

Résultats du comportement de choix d'itinéraire des voyageurs. La première constatation principale est que les voyageurs évaluent les différences relatives plutôt qu'absolues dans le temps de trajet. Cela signifie qu'une différence de 5 minutes dans le temps de trajet pèse différemment pour des trajets de 10 et 30 minutes. Ce résultat a des implications pratiques pour l'estimation des modèles de choix d'itinéraire, et donc dans l'affectation du trafic, où l'expression du temps de trajet en termes relatifs pourrait accroître le réalisme des prévisions. Une deuxième constatation est qu'au niveau individuel, un petit pourcentage des participants (10%) a toujours choisi la voie la plus rapide, ces participants peuvent être considérés comme parfaitement rationnels. Le comportement des autres participants peut être mieux expliqué par une rationalité limitée. À cet égard, il a été constaté que les participants sont hétérogènes par rapport à leur bande d'indifférence, et qu'au moins 70% d'entre eux ne considéreraient pas les itinéraires avec des différences de temps de trajet 1,5 fois plus lent que l'alternative la plus rapide. Le participant moyen n'a pas considéré les itinéraires avec des différences de temps de trajet 1,3 fois plus lents que l'alternative la plus rapide.

Modèle de choix d'itinéraire pour un comportement rationnellement limité. Un modèle de choix d'itinéraire rationnellement limité a été proposé: le modèle de rationalité limité pour la génération d'ensemble de choix (BRCS). Cette modèle considère (i) un processus de génération d'un ensemble de choix rationnel limité dans lequel les itinéraires alternatifs avec des différences de temps de trajet au-dessus d'un certain seuil (bande d'indifférence) sont supprimés des alternatives disponibles, et (ii) un processus de choix rationnel pour les alternatives dans l'ensemble de choix généré. L'ensemble de choix et le choix de route sont tous deux estimés conjointement, ce qui permet au BRCS d'inférer implicitement la distribution de la bande d'indifférence de la population latente. Le processus de génération d'ensembles de choix suppose une distribution paramétrique pour les bandes d'indifférence. La capacité du modèle à déduire la distribution des bandes d'indifférence a été testée à l'aide de données synthétiques et réelles. Dans les deux cas, le modèle a donné une précision prédictive supérieure au modèle logit mixte classique.

Données sur l'itinéraire des voyageurs et le comportement de choix de l'heure de départ. Une dernière contribution concerne les données sur les choix des voyageurs collectées pour l'élaboration de cette thèse. Les données proviennent de 16 expériences, dont 9 ont été utilisées dans l'élaboration de cette thèse, totalisant 7 510 choix effectués par 717 participants différents sur 41 paires de DO. Les données seront accessibles au public.

Cette thèse fait partie d'un projet européen ERC intitulé MAGnUM: Approche de modélisation du trafic multi-échelle et multimodale pour la gestion durable de la mobilité urbaine.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Ludovic Leclercq and Prof. Nicolas Chiabaut who gave me the opportunity to work with them at the LICIT laboratory, and then supported me throughout the three and a half years of my PhD. In short, I could say that this thesis is the result of maturing the ideas and points of view discussed during our long (very long) meetings around pages of half-baked analyses depicted in a graph. Also, I recognise the scientific freedom that I had to investigate and try these ideas.

I acknowledge all the people who made the MDG experiments possible, both from the technical and organisational tasks. Here, I mention again my supervisors, who conceived the tool and help on all tasks related to the experiments. I warmly thank Cécile Becarie and Jean Krug, who dedicated a lot of time and effort in the developing and configuration of the MDG. This thesis would not have been possible without your work! I also thank the rest of the colleagues who helped to organise and carry out the experiments. And, of course, I thank all of the people who participated in the experiments.

I would like to express my gratitude to Prof. Emma Frejinger and Prof. Francesco Viti who accepted to review my thesis. Your comments, helped not only to improve my manuscript, but also to contemplate points of view that I would hardly have considered otherwise. I am also thankful to the rest of the members of the examination committee, Prof. Christine Solnon and Prof. Eric Gonzales. I enjoyed the discussion that your questions sparked, not only about the content of my thesis but also of the many interesting future directions that might originate from it.

I am grateful to have worked in the LICIT laboratory. Its members, past members and visitors made my journey an enriching and pleasant experience. The long coffee breaks and discussions about any subject (environment, economics, politics, music, movies and traffic, of course), the Thursdays of chapati, the conferences, the nights out, the MAGnUM meetings and *sorties*. There is always something else to learn. It was a pleasure to work with you

Finalmente, quiero dar las gracias a mi familia, Elsa, Humberto, Daniela, Céline y Yah-in, y a mis amigos, quienes de cerca o de lejos me han dado el apoyo emocional e intelectual para completar esta fase de mi vida.

This thesis received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 – MAGnUM project).

Table of Contents

List of Figures

List of Tables

1 Introduction

During the last century, the world's population has witnessed an unprecedented process of urbanisation. In 1950, 30% of the world's population lived in cities, by 2018 they are home to half of the people worldwide and this percentage is projected to increase by 68% by the year 2050 (U.N., 2019). Notwithstanding the benefits that cities represent for their inhabitants, such as access to labour, health, education and other services, there are drawbacks that present serious challenges for their future viability. Traffic congestion is amongst them. Traffic congestion is the cause of monetary losses as well as the emission of pollutants that are, in turn, the cause of premature death amongst their inhabitants and a major contributor to the warming of our planet. It is in this context that strategies that help reduce congestion become relevant. To test the feasibility of different traffic reduction strategies, researchers and network managers rely on traffic simulation. In traffic simulation, travellers are assumed to follow a behavioural model that determines their travel choices, these choices are then considered altogether during a period of time in order to produce the traffic patterns in the networks. This allows to anticipate the impact of traffic control strategies under different scenarios: demand levels, network topology and traffic facilities. It is in this sense that understanding and approximating the behaviour of travellers is indispensable to forecast the states of a transportation network, in both normal and hypothetical scenarios.

The experimental studies on travellers' behaviour usually focus on the particular determinants of travellers' choices, and thus consider limited environments in their experiments: few OD pairs and routes and traffic conditions. As a consequence, the choice models derived from data coming from these experiments play a more interpretative than predictive role: their purpose is to understand the role of different traits that influence travellers' behaviour, rather than predicting their choices across the different situations found in the road network. In this thesis, choice models are studied from the predictive perspective. The objective is to formulate models capable of predicting the route choices at a full-scale urban network level, which is necessary for simulators to realistically represent the traffic conditions. For this purpose, the behaviour of travellers is studied through computer-based experiments. The experiments in this thesis comprise a large variety of situations (OD pairs, routes and traffic conditions) that resemble those that travellers find when travelling in an urban environment. This variety of situations must be taken into account in order to obtain choice models capable of scaling-up at the network level.

1.1 Background and motivation

When travelling in a city, travellers face several decisions regarding the (i) activity to engage, (ii) the destination of the trip, (iii) the mode of transportation, (iv) the departure time, and (v) the route to complete the trip. These individual decisions altogether shape the traffic states in the transportation network at a given time. However, at the same time that the states of the network are the consequence of travellers' choices, the characteristics and traffic states of the network influence travellers' decisions. One might think, for example, of travellers adapting their departure time and route choices anticipating congestion on the usual road. This bidirectional relationship between travellers' behaviour and traffic dynamics is the central idea in network loading problems, and it is studied from two different perspectives. The first is related to the traffic assignment problem, in which the traffic patterns at network-level are obtained as the result of the interactions between the choices of all travellers in the transportation network, based on general assumptions about network equilibrium. The second is related to the choices of the individual travellers, with particular emphasis on the identification and classification of the factors that influence their choices.

In traffic assignment, the demand from all origins to all destinations is assigned to the roads in the network by assuming a general behavioural principle governing the choices of travellers. The first and simpler of such principles is the well-known Wardrop's first principle or user equilibrium (UE) (Wardrop, 1952), defined as the state of the network in which no user may lower his transportation cost through unilateral action. UE assumes that all users in the network are utility maximisers (or cost minimisers) and thus the traffic patterns in the network are the result of all travellers minimising their travel cost simultaneously. Since then, considering other models of behaviour for the travellers, alternative definitions of network equilibrium have been proposed, such as stochastic user equilibrium (SUE) (Daganzo and Sheffi, 1977), which is based on random utility theory, and boundedly rational user equilibrium (BRUE) (Mahmassani and Chang, 1987), based on the bounded rationality hypothesis. In traffic assignment, the emphasis is placed on how the various behavioural principles derive in different traffic patterns, and on the methods to solve the network loading problem to determine static or dynamic traffic states at large scale. The choice models that dictate the individual behaviour of the travellers are, in this case, considered as simplified mathematical abstractions that permit to compute the assignment under the different behavioural principles. Moreover, since the aim is to derive the full path flow distributions at the network level, little attention is paid to the fine calibration of the individual user behaviour.

Studies, from the perspective of the individual behaviour of travellers are mainly concerned with the specific factors that influence their choices. Travellers' behaviour is a process which involves psychological and cognitive mechanisms through which travellers perceive and evaluate the states of the network, and then make decisions accordingly (Bovy and Stern, 1990, Ben-Akiva et al., 1999). Although this definition may appear simple, there are many factors, associated to both the traveller and the environment, that intervene in this process, making it a complex problem. These factors are heterogeneous (as heterogeneous as individuals can be), and they interact in ways that are not easily observable to produce the choices. In the route choice context, studies have been made to understand the learning process of travellers (Iida et al., 1992, Bogers, 2005, Selten et al., 2007) and its relationship to the formation of habit, familiarity and the exploration of the alternative routes (Srinivasan and Mahmassani, 2000, Prato et al., 2012, Kaplan and Prato, 2012). The reliability of the travel time in alternative routes has been found to play a major role in route choice. Travellers (or the great majority of them) exhibit risk-aversion, and thus they prefer slower reliable routes rather than fast unreliable ones (Abdel-Aty et al., 1997, Avineri and Prashker, 2005, Ben-Elia and Shiftan, 2010). However, risk aversion is heterogeneous amongst travellers, and between trip purposes (Ramos, 2015), and depends on the value of time of the travellers (de Palma and Picard, 2005). Another line of research largely treated in literature is the effect of travel information on the route choices, with special interest on the impact of information on the travel time minimisation behaviour (Bonsall, 1992, Ben-Elia and Shiftan, 2010, Adler and McNally, 1994, Abdel-Aty et al., 1997) and its effects on the social cost (Mahmassani and Jayakrishnan, 1991, Ben-Akiva et al., 1992, Rapoport et al., 2014a, Ben-Elia and Avineri, 2015). Studies about variables other than travel time that explain the choices of travellers can be found in Bovy and Stern (1990), Ramming (2002). The above-mentioned studies are predominantly based on laboratory-like experiments, where participants' choices are observed on simple scenarios (two or three routes in few OD pair configurations) that do not cover the multiplicity of situations that are found in a city-scale transportation network. These simplifications are made in order to guarantee the internal validity of the experiments. For example, if the objective is to measure the impact of travel time reliability in the route choices of travellers, then it suffices to present choice problems to the participants with only two alternative routes; the shape and attributes of the routes are irrelevant. This is justified because including more routes in the experiment or showing extra information on the alternatives render the analysis of the results more complicated, as the effect of the variable being investigated is confounded with the effect of the rest of the attributes of the alternatives.

Since the objective of these studies is to understand the determinants of travellers' behaviour, the choice models in these works play a mainly interpretative role. At full-scale network level, the choices of travellers are made over thousands of OD pair configurations that consist of short and long trips as well as routes that differ in their attributes. In the case concerning this thesis, the city of Lyon in France, the network has 19,967 links and 19,697 nodes. Therefore, these choice models may not generalise well to the amount of situations that are found in an urban network, in the sense that their predictions may not be accurate in OD pairs with different enough characteristics.

The motivation of this thesis is to obtain choice models that are both consistent with the observed behaviour of travellers, and that generalise well at full-scale network level. In other words, models that make accurate predictions in all OD pairs in an urban transportation network, but that are estimated with observations on a limited number of OD pairs. The general idea is to implement, in the future, these choice models in dynamic traffic assignment algorithms, and thus obtain traffic patterns simulated from a behavioural perspective. To attain consistency with travellers' behaviour, the choice models are inferred from data collected through computer-based experiments. However, since the choices of travellers cannot be observed in such a large number of OD pairs and routes that are found in an urban network, first a methodology to collect observations of the choices of travellers through computer-based experiments is required. Models estimated with data coming from this methodology should be able to scale-up, from a small set of observable OD pairs, to the full-network level. Broadly speaking, the objectives of this thesis are (i) building a methodology that allows to estimate choice models over a full-scale road network, and (ii) selecting an appropriate model specification to be applied to a large-scale network, taking into account different aspects of travellers behaviour.

1.2 Research approach

In order to estimate and test different choice models, the behaviour of travellers need to be observed. There are two recognised methods to collect data on travellers' behaviour: stated preference (SP), in which subjects are faced with hypothetical choice situations, and revealed preferences (RP), that are based on direct observation of the choices of travellers in real-world situations or based on surveys asking for actual travel behaviour. In the route choice context, an example of a SP experiment might be a survey in which the respondents are faced with choosing the route that they would, given different attributes and congestion scenarios in the alternative routes. If the same route experiment were based on the RP method, then the actual route choices of travellers would be observed (through GPS, for example), or asked via a survey. Examples of studies about travellers' behaviour based on SP data are Iida et al. (1992), Adler and McNally (1994), Mahmassani and Liu (1999), Ben-Elia and Avineri (2015); and examples based on RP data are Bierlaire and Frejinger (2008), Ramos et al. (2012), Zhu and Levinson (2015), Yildirimoglu and Kahraman (2018b). The SP and RP methods have advantages and disadvantages (Kroes and Sheldon, 1988, Bovy and Stern, 1990, Train, 2003). On the one hand, SP suffers from external validity of the responses, understood as the discrepancy between the stated responses and the actual behaviour of the respondents in real-life situations. The lack of external validity is caused by respondents stating intentions or opinions and not their actual behaviour, which is explained by the absence of context in the experiment (e.g. change in traffic conditions due to bad weather) or by the inadequacy of the incentives given to the participants. On the other hand, the main disadvantages of the RP method are the lack of awareness of the alternatives considered by the decision-makers and the lack of variability in the scenarios in which the choices are made. The lack of awareness on the alternatives implies that the researcher does not have information on the alternatives against which the decision-makers evaluate their choices, inevitably introducing a new source of uncertainty and reducing the internal validity of the experiments. The advent of the information technologies has facilitated new forms to carry out behavioural experiments, notably, computer-based experiments, that belong to the SP type. Although the computer-based SP experiments also suffer from the lack of validity of traditional SP methods, they possess some characteristics that could attenuate this issue. First, computer-based experiments permit to define more complex scenarios that resemble real-life situations, and to present them to the participants in a more realistic and intuitive way, providing more context to the situations in which the choices are made (Chen and Mahmassani, 1993, Koutsopoulos et al., 1994, 1995). For example, by showing the alternative routes over a map of the city. Second, by presenting the choice situations in a more intuitive way, the cognitive burden of the participants is eased, allowing to increase the amount of choices that can be collected from each individual thus reducing the costs of the experiments. Third, the computer-based SP experiments make it possible to introduce consequences, in the form of a score (score design), to the choices of the participants, allowing participants to earn points as if it were a game. This could enhance the engagement of participants, although the design of the score may also influence the respondents' strategies (Bogers et al., 2005).

The approach in this thesis to investigate travellers' behaviour in transportation networks is through computer-based experiments at large scale, for which a platform named the Mobility Decision Game (MDG) has been developed. Contrary to other SP experiments, the MDG allows to observe the choices of a large number of participants under different situations (see Fig. 1.1), which is essential for inferring choice models at full-scale network level. The MDG is a web-based computer game designed to confront the participants in the experiments with decision problems regarding the departure time and the route choices to complete a trip. The decision problems are placed under different hypothetical scenarios, comprising of OD pairs joined by three alternative routes with contrasting attributes and varying traffic conditions. The scenarios occur in a dynamically simulated environment of the real network of the city of Lyon, France, in which the traffic conditions are generated in real-time by a single dynamic microscopic simulator. During a MDG experiment, multiple OD pairs are assigned to the participants, allowing to observe the choices of the same participants in different OD pairs. Furthermore, some of the participants may receive traffic information in the form of travel time.

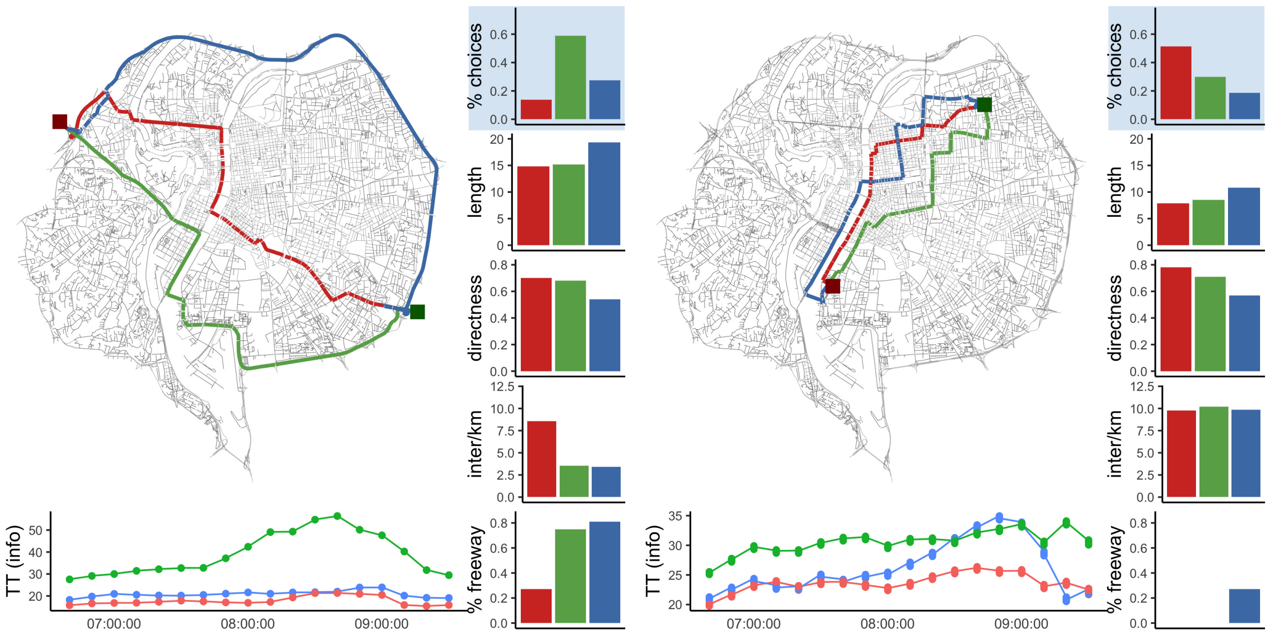

Figure 1.1: Two OD pairs joined by three routes for the MDG experiment. The experiments consist in choosing one out of three alternative routes to complete a trip. The varying traffic conditions and route attributes allow to observe the choices of travellers in different situations over the city of Lyon in France.

The model adopted in this thesis to predict the choices of travellers is the mixed logit model (McFadden, 1984, McFadden and Train, 2000), which belongs to the random utility maximisation framework (Mcfadden, 1972). This framework assumes that individuals obtain a certain level of utility from each alternative in a choice situation, and that they choose the alternative with the maximum utility, i.e., individuals are utility maximisers. In random utility maximisation models, the utility obtained from an alternative is related to the attributes of both the alternative and the decision-maker. As a consequence, the choices of the decision-makers are, to some extent, explained by variables that can be measured and that depend on the OD pair and on the situation (OD pair and traffic conditions). Moreover, the number of assumptions about the parameters of the model is small, and these can be easily inferred from the data, obtaining succinct representations of the utility and the choice probabilities. Therefore, random utility maximisation models are well-suited for the prediction of travellers' choices at full-scale network. A careful selection of the scenarios over which the MDG experiments evolve, permits to estimate choice models that generalise well to the full-scale network level, and to test and select the model specification that best represent the choices of the participants in the experiments.

1.3 Research objectives and questions

The objective of this thesis is to propose and estimate choice models that predict accurately the choices of travellers over the diversity of situations found in a transport network. To attain this objective, empirical evidence on travellers behaviour is collected through computer-based SP experiments that permit to observe the choices of the participants on a diverse set of scenarios (OD pairs and routes) with varying traffic conditions and travel time information. The experiments focus on the route choices of uni-modal car trips that are based on the map of the city of Lyon, France. The number of scenarios on which the choices of the participants can be observed through the MDG is still limited, compared to the entire set of situations found on a full-scale network. Therefore, the scenarios used for the experiments need to be determined in order to represent the diversity of scenarios found on the entire network. This leads to the first specific objective of this thesis, the estimation of choice models at full-scale network level, which deals with the design of the choice experiments with the MDG. The second specific objective of the thesis, the choice model selection, concerns the finding of route and departure time models based on the collected empirical evidence. The specific objectives are stated in a succinct manner below, together with the research questions that need to be answered for their achievement.

Framework for the estimation of choice models at full-scale network level. Concerns the design of experiments, the experimentation and the generalisation of the observed choices to the full-scale network. This comprises the sample of a small set of OD pairs and routes of the road network of the city of Lyon in France, over which the route and departure time experiments are held. The sample must be representative of the road network at route attribute level, and the observed choices over these OD pairs and routes must generalise at full-scale network level. In other words, choice models estimated with data on the small representative OD pairs and routes must approximate with good accuracy the choices in the entire network.

- *Is it possible to represent the OD pairs and routes in a network with a small set of OD pairs?*
- *Are choices in the representative OD pair set also representative of the choices in other OD pairs?*

Choice model selection. Regards the finding of an appropriate route and departure time choice model based on the findings of the quantitative analysis of the results of the MDG experiments. The impact of the attributes of the OD pairs and routes, the travel time information provided to the participants, as well as the heterogeneity of the preferences of the travellers are accounted for by the models. Boundedly rational behaviour is considered in the development of the models.

- *What is the impact of the travel time information and route attributes on the choices of travellers?*
- *Are travellers' choices best explained from a perfect rational or boundedly rational paradigm?*

1.4 Contributions

Methodology for the estimation of route choice models at full-scale network level. A methodology, based on the cluster analysis, is proposed to determine an optimal subset of OD pairs over which the choice experiments are carried out. A cluster centroid, being the element that minimises the Euclidean distance to all the cluster elements, can be selected to represent its group, and the set of all the cluster centroids as representative of the whole network. The OD pairs obtained with this methodology have three properties that make them suitable for the route choice experiments. First, they cover a large variability of the OD pairs and routes of the network, i.e., they are representative of the full-scale network. Second, the variance of their attributes is high, allowing to identify the influence of each attribute in the choices of the participants. Third, the number of OD configurations is small, which allows to collect enough data in each OD pair. The first and second properties imply that a choice model estimated with data over the representative OD pairs will generalise well to the rest of the OD pairs in the network, in the sense that the choices in any OD pair in the network are accurately predicted. The third property implies a robust estimation of the model parameters, i.e., smaller estimation errors. The models estimated with the representative OD pairs showed a higher predictive accuracy than models estimated with different sets of OD pairs.

Travellers' route choice behaviour findings. The first main finding is that travellers evaluate relative rather than absolute differences in travel time. This means that a 5 minute difference in travel time weights differently for trips of 10 and 30 minutes. This result has practical implications for the estimation of route choice models, and thus in traffic assignment, where expressing the travel time in relative terms could increase the realism of the predictions. A second finding is that at individual level, a small percentage of the participants (10%) chose always the fastest route, these participants can be considered as perfect rational. The behaviour of the rest of the participants can be better explained by bounded rationality. In this regard, it was found that the participants are heterogeneous with respect to their indifference band, and that at least 70% of them would not consider routes with travel time differences 1.5 times slower than the fastest alternative. The average participant did not consider routes with travel time differences 1.3 times slower than the fastest alternative.

Route choice model for boundedly rational behaviour. A boundedly rational route choice model was proposed: the bounded rational choice set generation mixed logit model (BRCS), which considers (i) a bounded rational choice set generation process in which the alternative routes with travel time differences above a certain threshold (indifference band) are removed from the available alternatives, and (ii) a rational choice process for the alternatives in the generated choice set. Both the choice set and the route choice are jointly estimated, allowing for the BRCS to implicitly infer the latent population's indifference band distribution. The choice set generation process assumes a parametric distribution for the indifference bands. The ability of the model to infer the distribution of the indifference bands was tested using synthetic and real data. In both cases the model resulted superior in predictive accuracy than the classical mixed logit model.

Data on travellers' route and departure time choice behaviour. A fourth contribution is the data on travellers choices collected for the elaboration of this thesis. The data comes from 16 experiments, from which 9 were used in the elaboration of this thesis, totalling 7,510 choices made by 717 different participants over 41 OD pairs. The data will be publicly available.

1.5 Thesis outline

This thesis comprises four studies about travellers' behaviour. These studies are all based on the data obtained through computer-based route and departure time experiments, carried out with the Mobility Decision Game platform. The second chapter in the thesis introduces this platform. Thereafter, the four studies are presented, organised into two parts: the first dedicated to route choice behaviour and the second to joint route and departure time choice behaviour. A short description of the thesis' chapters is given below, followed by a diagram explaining the relationship between them (Fig. 1.2).

Chapter 2. Presentation of the mobility decision game (MDG), the computer-based SP experimental tool used to collect data on travellers' behaviour at large scale. This chapter has as a secondary objective to document the functionalities of the tool.

Part I: Route choice behaviour

Chapter 3. Review of the literature on route choice and random utility models. The selection of the mixed logit model (MXL) as the modelling approach for the route choice behaviour of travellers in this thesis is justified in this chapter.

Chapter 4. This chapter is concerned with the experimental design for the MDG, specifically, the methodology to select the OD pairs for the route choice experiments in the MDG. The selected OD pairs need to be representative of those found in the road network, such that the choices collected over them generalise at full-scale network level.

Chapter 5. The factors that influence the route choices of travellers are analysed in this chapter. In particular, the effect that travel time information and the road attributes that can be observed over the road map of the city are studied. The bounded rationality of participants with respect to travel time information is assessed, giving estimates on the size and the heterogeneity of the indifference band.

Chapter 6. Based on the findings of Chapter 5, a model for route choice that assumes boundedly rational behaviour is introduced: the bounded rational choice set (BRCS) model. The model considers a bounded rational choice set formation process and a rational choice process. The predictive accuracy of the model is tested against that of the mixed logit model using synthetic and real data.

Conclusions of part I: route choice model selection. As a conclusion of the route choice part of the thesis, the predictive accuracy is assessed for different specifications of the MXL model, as well as the BRCS model.

Part II: Route and departure time choice behaviour

Chapter 7. This chapter presents the first results of an investigation on simultaneous route and departure time choice behaviour. A quantitative analysis of the results of the experiments is done. A model is proposed for joint route and departure time choice where the alternatives are correlated in time. This chapter is still an investigation in progress.

Chapter 8. This chapter concludes the thesis and provides future research perspectives.

1.6 List of publications

Peer-reviewed journal articles

- González Ramírez, H., Leclercq, L., Chiabaut, N., Becarie, C., Krug J. (2019). Unravelling travellers' route choice behaviour at full-scale urban network by focusing on representative OD pairs in computer experiments. PLOS ONE 14(11): e0225069. https://doi.org/10.1371/journal.pone.0225069

- González Ramírez, H., Leclercq, L., Chiabaut, N., Becarie, C., Krug J. (2019). Travel time and bounded rationality in travellers' route choice behaviour: a computer route choice experiment. Submitted to Travel Behaviour and Society - Under minor review after second revision.

Conference proceedings and presentations

- González, H., Leclercq, L., Chiabaut, N., Bécarie, C., Krug, J. (2018). A massively multiplayer simulation game framework to study dynamic route choice behavior #18-00902. Proceedings of the 97th Transportation Research Board Annual Meeting (TRB), January 6-11, Washington, (USA), Transportation Research Board, 16 p.

- González, H., Leclercq, L., Chiabaut, N. (2017). A massively multiplayer online mobility game for data collection on travellers' behaviour. Conference on Traffic & Granular Flow (TGF), 19-22 July, Washington, DC. (USA).

Working papers

- González, H., Leclercq, L., Chiabaut, N. (2020). Bounded rational choice set generation MXL model for route choice. Working paper.

- Mariotte, G., Leclercq, L, González Ramírez, H., Krug J., Becarie, C. (2020). Assessing traveler compliance to social optimum: A stated preference study. Submitted to Travel Behaviour and Society.

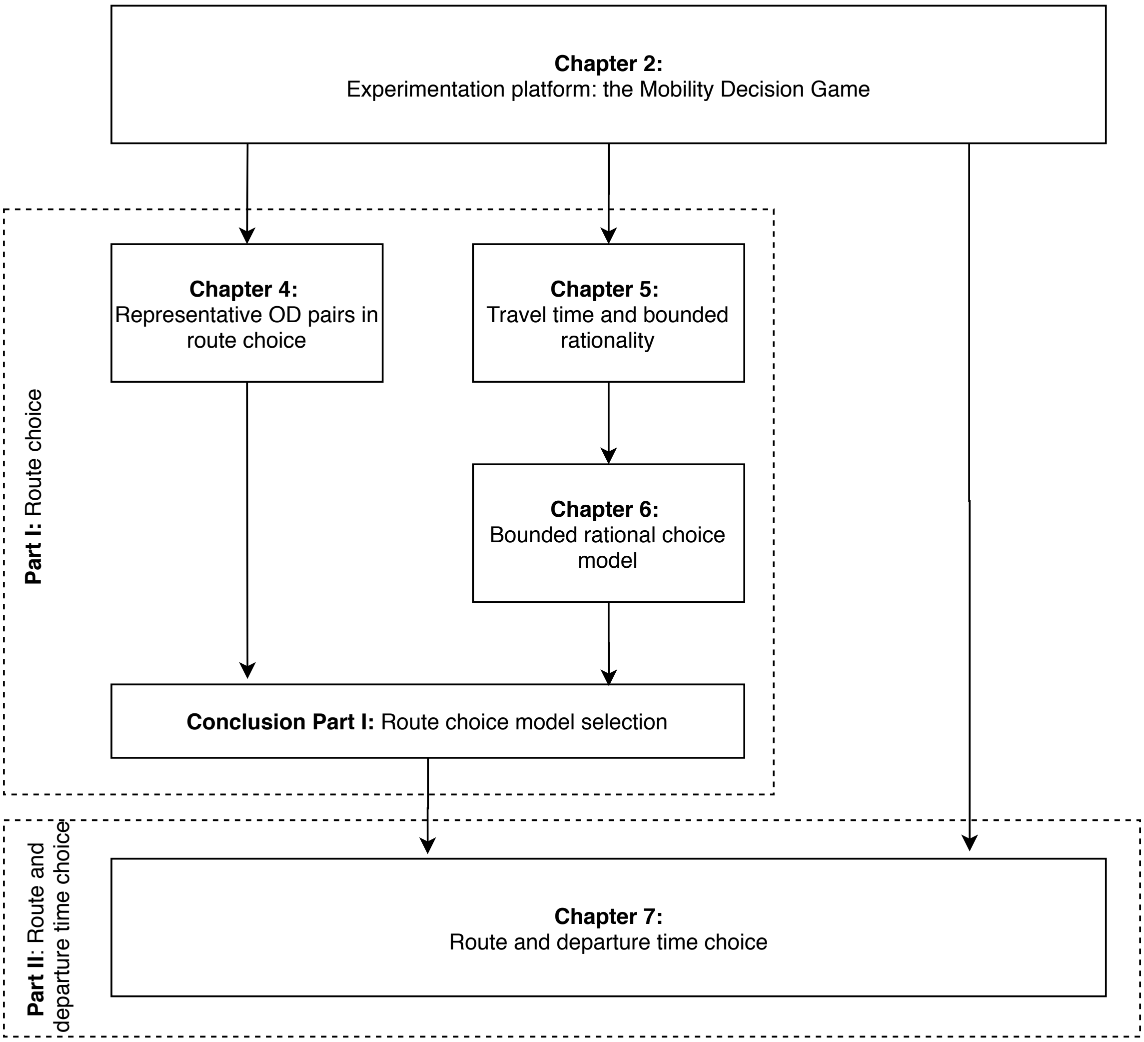

Figure 1.2: Thesis outline. Dependence relationship between the chapters of this thesis.

2 Mobility decision game

This chapter introduces the mobility decision game (MDG), a computer platform (developed by the LICIT laboratory) to investigate travellers' behaviour in transportation networks at large-scale. The MDG is a web-based computer game designed to confront the participants with a variety of decision problems regarding the mode, the departure time and the route choices to complete a trip[1]. The decision problems are placed under different hypothetical scenarios, i.e., OD pairs joined by three alternative routes that contrast in their attributes and with varying travel times. This allows to investigate the determinants of the participants' decisions under different circumstances.

2.1 Motivation

There are two recognised methods to collect data on travellers' behaviour: stated preference (SP), in which subjects are faced with hypothetical choice situations, and revealed preferences (RP), that are based on direct observation of the choices of travellers in real-world situations or based on surveys asking for actual travel behaviour. In the route choice context, an example of a SP experiment might be a survey in which the respondents are faced with choosing the route that they would use, given different attributes and congestion scenarios in the alternative routes. If the same route experiment were based on the RP method, then the actual route choices of travellers would be observed (through GPS, for example), or asked via a survey.

The SP and RP methods have advantages and disadvantages (Kroes and Sheldon, 1988, Bovy and Stern, 1990, Ben-Akiva et al., 1992, Earnhart, 2002, Train, 2003). In the case of SP, the most important aspect impacting the reliability of the data is the validity of the responses, understood as the discrepancy between the stated responses and the actual behaviour of the

[1]Even though the MDG is intended for experiments regarding the three choices, the studies in this thesis concern only the route and departure time choices; the mode choice is not studied.

respondents in real-life situations. This discrepancy introduces biases in the data that might lead to wrong predictions or interpretations. The lack of validity is caused by respondents stating intentions or opinions and not their actual behaviour, which is explained by the absence of context in the experiment (e.g. change in traffic conditions due to bad weather) or by the inadequacy of the incentives given to the participants. An example of this discrepancy can be found in Ramos et al. (2012), where the authors observe that participants in a combined RP and SP route choice experiment stated their preference for reliable routes, but that they barely chose it as shown by the RP data. The lack of validity is not an issue for the RP method, as choices are observed in the context in which they actually evolve, i.e., there are no hypothetical situations that need to be described. The main disadvantages of the RP method, with respect to the SP method, are (i) the lack of awareness of the alternatives and the values that the variables describing the alternatives take at the moment of the decision, and (ii) the lack of variability in the scenarios in which the choices are made, caused by the limited situations in which the choices may occur. In the route choice problem, (i) implies that the researcher does not have information on the alternative routes that are considered by the traveller to complete a trip and, more important, the traffic conditions at the moment of the choice are not known with certainty. Therefore, introducing a new source of uncertainty. For example, in the GPS-based route choice experiments in Frejinger and Bierlaire (2007), Frejinger et al. (2009), Zhu and Levinson (2015), Yildirimoglu and Kahraman (2018b), de Moraes Ramos et al. (2020), the alternative routes joining each OD pair in the network had to be generated, and their travel times inferred from the data. The implications of (ii) are that the route choices may be observed under the same traffic conditions, causing multicolinearity in the attributes that describe the alternatives (Ben-Akiva et al., 1992, Train, 2003). Following the example of route choice, a low variability of the travel times on the alternative routes during the experiment will make the effect of this variable indistinguishable. Considering the advantages and disadvantages of both methods, the SP method can be regarded as a controlled laboratory experiment that, nonetheless, introduces some biases in the data, and the SP method can be regarded as observations in real-life situations which, on the other hand, are difficult to control.

The advent of the information technologies has facilitated new forms to carry out behavioural experiments, notably, computer-based experiments, that belong to the SP type. Although the computer-based SP experiments also suffer from the lack of validity of traditional SP methods, they possess some characteristics that could attenuate this issue. First, computer-based experiments permit to define more complex scenarios that resemble real-life situations, and to present them to the participants in a more realistic and intuitive way, providing more context to the situations in which the choices are made (Chen and Mahmassani, 1993, Koutsopoulos et al., 1994, 1995). For example, by showing the alternative routes over a map of the city. Second, by presenting the choice situations in a more intuitive way, the cognitive burden of the participants is eased, allowing to increase the amount of choices that can be collected from each individual thus reducing the costs of the experiments. Third, the computer-based SP experiments make it possible to introduce consequences, in the form of a score, to the choices of the participants, letting participants earn points as if it were a game. This could enhance the engagement of participants, although the design of the score may also influence the respondents' strategies.

In view of the aforementioned advantages and disadvantages of both SP and RP methods, the decision to use one method over the other (or using both) responds to the purpose of the study and availability of data. On the one hand, if the interest is the study of mobility patterns without regarding the causes of travellers' choices or the characteristics of the individuals, then RP may be a better choice. Examples of this kind of studies are Brockmann et al. (2006), González et al. (2008), Song et al. (2010), Tachet et al. (2017), which study general mathematical patterns of human mobility by using GPS or cell phone data. When the number of observations is large enough to infer the conditions of the environment and cost is not a limitation, then RP technique would also be more adequate. The studies of Zhu and Levinson (2015), Yildirimoglu and Kahraman (2018b) are examples in which the high density of observations allows to estimate the travel times in the network. On the other hand, if the interest is the study of the determinants of travellers' choices, then SP may be more adequate. Examples of SP in the route choice context found in literature, including computer-based experiments, are the studies on the learning process of travellers of Iida et al. (1992), Bogers (2005), Selten et al. (2007), the impact of advanced travel information systems (ATIS) in travellers route choices in Adler and McNally (1994), Lotan (1997), Mahmassani and Liu (1999), Ben-Elia and Shiftan (2010), Ben-Elia and Avineri (2015), Abdel-Aty et al. (1997), Srinivasan and Mahmassani (2000), and the effect of travel time variability and risk attitudes of travellers in De Moraes Ramos et al. (2013), Avineri and Prashker (2005), de Palma and Picard (2005), to mention some.

2.2 Definition of the MDG

In an experiment with the MDG, participants are confronted with decision problems under different scenarios. A decision problem consists in travelling from an origin to a destination, for which the participants are required to make several choices related to the departure time, the mode of transportation, and the route choice. A scenario is the environment in which the decision problems are placed, notably: the transportation network, the OD pairs and routes where the decisions are made, and the traffic conditions that determine the travel times on each of the alternatives. The MDG platform is in charge of both creating the scenarios and presenting the decision problems to the participants. To generate the scenarios in which the choices are made, the MDG interacts in real-time with a single dynamic microscopic simulator. In the MDG, the participants access simultaneously to the experiment through a dedicated web interface, showing the map of an urban network (see Fig. 2.1). During a MDG session, multiple OD pairs are assigned to the participants, allowing to observe the choices of the same participants in different OD pairs. Furthermore, some of the participants may receive traffic information in the form of travel time estimates or congestion maps. Thus, the MDG permits to investigate the determinants of the participants' decisions under different transportation and traffic information conditions.

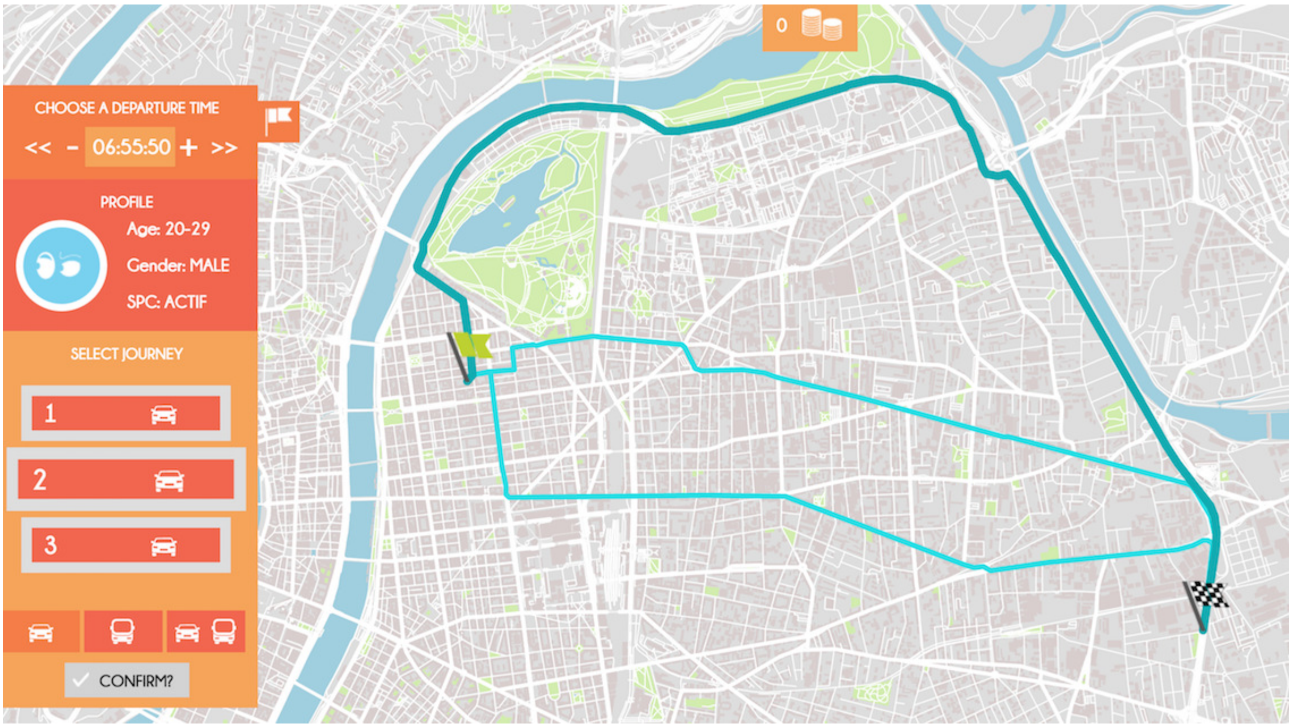

Figure 2.1: MDG interface. An OD pair and three alternative routes are shown over a section of the Lyon network in the MDG. The left menu allows participants to choose the mode of transportation, the departure time and the route choice.

2.2.1 Scenarios

Scenarios in the MDG are defined as the environment or the conditions in which the decision problems faced by the participants happen. They consist in an urban road network, its traffic conditions, and the OD pairs and routes where participants' choices are made. Therefore, a scenario can be defined more precisely as the road traffic conditions that the participants face when making a trip between a specific OD pair in a specific urban network. An important characteristic of the MDG is its ability to present participants with several and diverse scenarios. The scenarios in the experiments can be configured to study different aspects of travellers' behaviour.

Network description. The network description is the map representation of the road network of an urban area. This representation allows the participants in the experiments to observe some of the characteristics of the proposed routes, such as the functional class, depicted as the width of the links; and the number of intersections. Other characteristics of the network are depicted using different colours, as parks (green) and water bodies (blue). Two road networks, based on the real road network of the city of Lyon, were used in the experiments: Lyon-full network, the complete road network inside the peripheral ring, and the Lyon-36V network, a subnetwork of the former that includes a segment of the peripheral ring. The former network is composed of 3,663 links, whereas the second of 19,967 links (see Fig. 2.2). The trips' origins and destinations come from the zoning defined by the National Institute of Statistics and Economic Studies (INSEE) (Institut national de la statistique et des études économiques, 2018), and the major entry/exit points to the network. The zones are the geostatistic units used for the trip demand estimations and represent the origins and destinations of the trips generated or terminated inside the network. The entry and exit points represent the origins and destinations of the demand coming or going outside the network. In the Lyon-full network, there are 285 zones, 29 entry points and 28 exit points. The total number of origins is 313 and of destinations 310 (this quantity does not correspond exactly to the sum of zones plus entries/exits as there are zones that may have no outgoing or incoming trips), giving a total number of 96,096 OD pairs (see Fig. 2.2). The routes joining the origins to the destinations are derived with the A* algorithm looking for the k-shortest paths in free-flow travel time. The routes in a random sample of 100 OD pairs were compared to those obtained in Google Maps®, finding a good match between them. The total number of routes in the network, obtained with this algorithm, is 559,423, with an average number of 5.82 routes per OD pair. The network of Lyon-36V is composed of 71 zones, 14 entry points and 13 exit points (see Fig. 2.2). The total number of origins and destinations are, respectively, 85 and 84, giving a total number of 9,494 OD pairs. The total number of routes defined for this network is 40,938, with an average of 4.3 routes per OD pair. The characteristics of the road networks are summarised in Table 2.1.

Table 2.1: Characteristics of the Lyon-full and Lyon-V36 road networks.

Network	No. links	No. zones	No. entries	No. exits	No. ODs	No. routes	No. routes/OD
Lyon-full	19,967	285	29	28	96,096	559,423	5.8
Lyon-V36	3,663	71	14	13	9,494	40,938	4.3

Traffic conditions. The traffic conditions in the network are dynamically generated by a microscopic traffic simulator, based on the LWR traffic model (Leclercq, 2007, Laval and Leclercq,

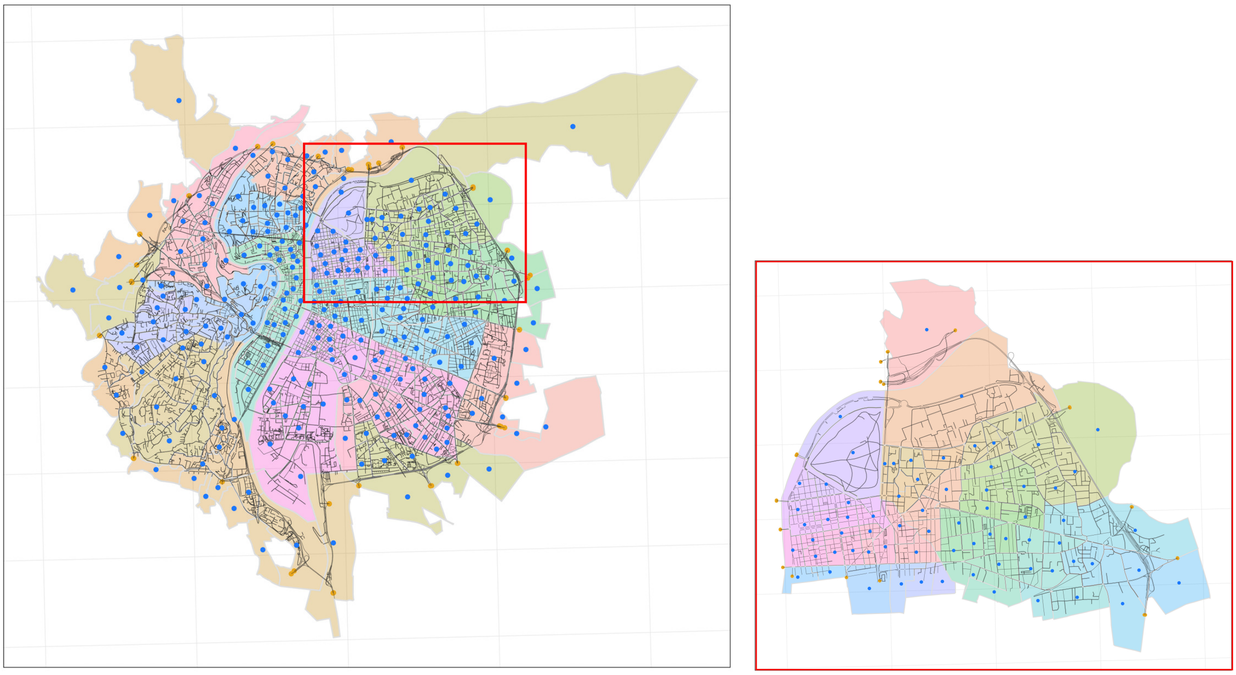

Figure 2.2: Lyon-full and Lyon-V36 OD pairs. The zones are depicted in different colours with their centroids in blue. The entry/exit points to the network are depicted with yellow points.

2008, 2010). The simulator generates and handles all the trips that populate the transportation network. The choices of the participants are considered as updated trip specifications by the microscopic simulator. This alters the traffic conditions in the network. To produce different traffic patterns in the network, the global demand (network level) as well as the local demand (OD pair and alternative level) are predefined. The demand, both global and local, can change over the simulated time periods, so the traffic conditions in the network vary throughout the experiment. The global demand is given by the trip rate (number of trips generated per second) in the whole network. These trips are then distributed locally amongst all the OD pairs in the network following the OD matrix, which contains the share of the global trips in each OD pair. The distribution of the travel demand has been built upon the estimation of the real dynamic OD matrix (Krug et al., 2019) with adequate modifications to increase the diversity of travel time configurations for the OD pairs where decisions of participants are made. It corresponds to the travel distribution during the morning rush hour. Finally, the local trips are assigned to each of the routes connecting the OD pairs. In the experiments, the global demand and the assignment in the playable OD pairs are designed to be time-dependent, with the objective to obtain different traffic conditions on the same OD pairs (see Fig. 2.3 and Fig. 2.4). This allows observe the change in route choice behaviour when the fastest alternative route is switched.

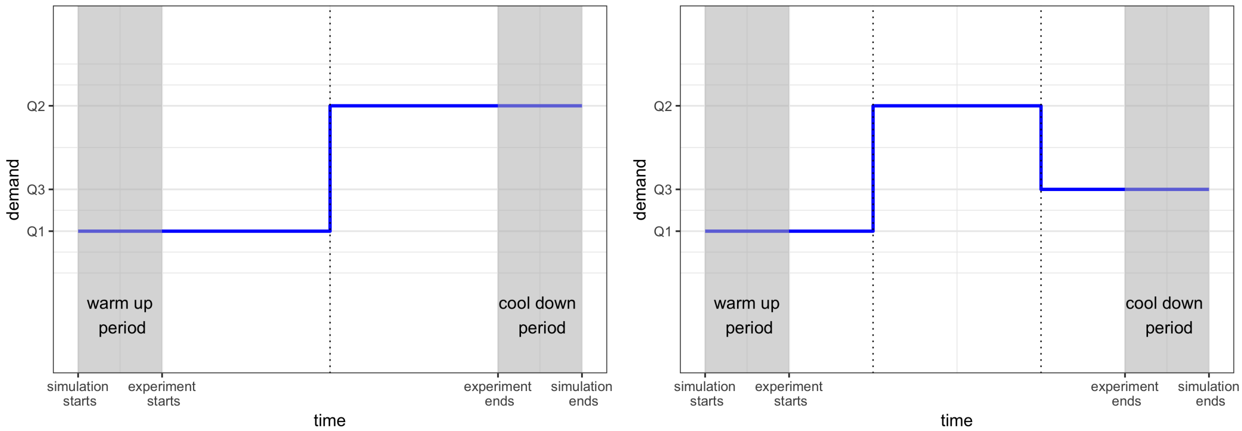

Figure 2.3: Typical global demand profiles in the MDG experiments. The global demand has an impact in the whole network traffic conditions. During the warm up period the network is populated in order to arrive to the desire traffic conditions. The cool down period allows the started trips to finalise.

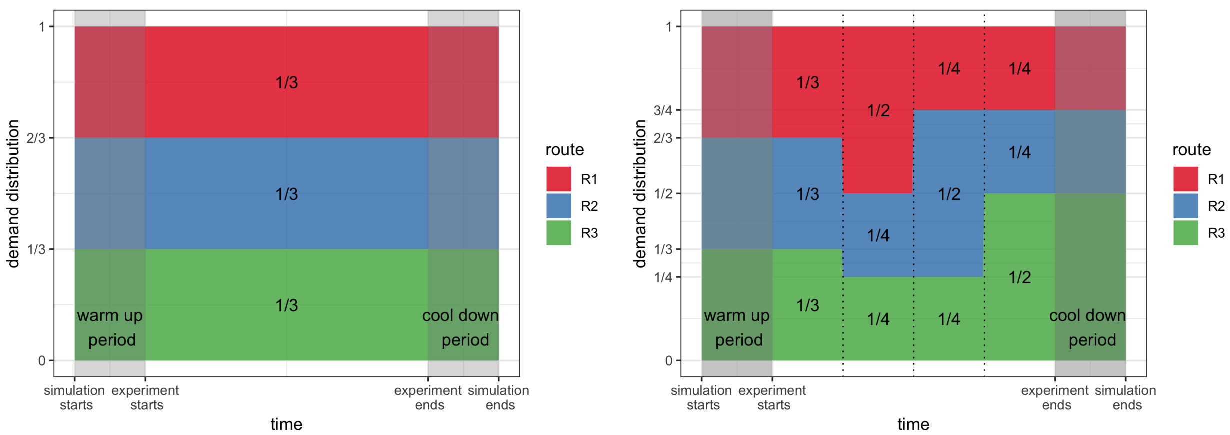

Figure 2.4: Typical trip assignment profiles in the MDG experiments. The local demand is distributed among the alternative routes to change their traffic conditions. On the left, the assignment profile with constant trip distribution. On the right, the trip distribution on the alternatives change to produce congestion in one of the three alternatives.

Playable OD pairs and alternative routes. The set of OD pairs with three connecting routes in which the decision problems are posed to participants, i.e., the origins and destinations of the trips that the participants are asked to complete. These OD pairs are predefined in the experiments. Three routes are proposed to the participants to complete a trip between an OD pair in a decision problem; these routes constitute the choice set for the participants' route choice. The playable OD pairs and alternative routes can be a subset of the OD pairs and routes in the network, or can be manually defined. The limitation to three alternative routes in the route choice problems (coming from technical limitations of the MDG) does not restrict the scope of the experiments or diminishes the quality of the results for two reasons. First, choice sets with many alternatives may be burdensome for participants as they may have trouble identifying the differences between the routes. Second, the low variability between routes attributes due to the small number of alternatives in the choice set is compensated by the presence of many OD pairs.

2.2.2 Decision problem

A decision problem in the MDG consists in completing a trip on a particular OD pair, with a given purpose and an objective arrival time. Participants are required to make a series of choices regarding the mode of transportation, the departure time, and the route. Once a trip is started, participants may also change the initially chosen route. The consequences of participants' choices are translated as a score, obtained after finishing a trip, that depends on whether the mission was completed near the objective arrival time. More formally, a decision problem in the MDG is composed of three components that are compulsory for the definition and description of the problem, and one nonessential (optional) component. The three essential components are: the statement of the problem, the choices of the respondents, and the score; the nonessential component is the travel time information that is provided to the participants as a support for their choices. It is worth mentioning that travel time information is of high importance to study the influence of travel time on the choices of participants. The four components and their description are listed below; they are also depicted in the flow chart presented in Fig. 2.5.

Statement of the problem. The problem faced by the participants is to complete a trip on time on a given OD pair. To this end, the following pieces of information are made available to the participants.

- Objective arrival time. It is the target arrival time of the trip; participants are required to complete the trip before the objective arrival time. The objective arrival time is computed by the MDG, depending on (i) the time at which the decision problem is presented to the

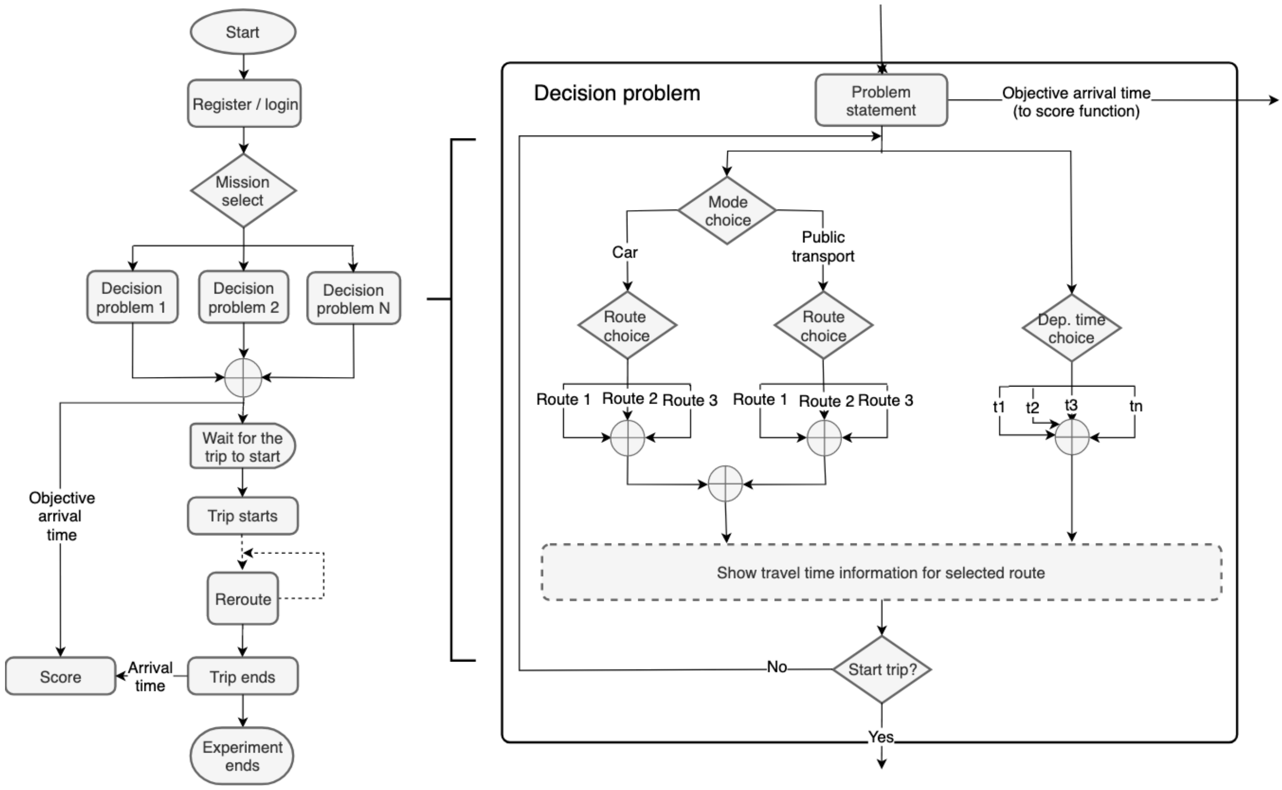

Figure 2.5: Flow chart of a MDG experiment (left) and decision problems (right).

participant, (ii) the travel time of that trip in the original microscopic simulation used by the MDG, and (iii) an added margin with the purpose of giving more time to participants to arrive on time. If we denote the arrival time of trip k in the simulation without humans (unmodified by the players' choices) as ta_k, then the objective arrival time, informed to the participant who selected that trip, is given by $to_k = ta_k + \delta$, where δ is a (predefined) constant.

- Purpose of the trip. Induces participants to value their time and their losses (late arrivals) and to change their behaviour accordingly. The purposes can be, for example, "going to work", "going shopping" or "catching a train".

Participants' choices. Given the statement of the problem, the participants make one or more choices to complete the trip. The choices that are allowed to the participants depend on the objectives of the experiment. In some experiments, the choices can be limited to route choice, whereas in other experiments the departure time and route choices might be considered together. The choices that the participants can make in the MDG are:

- Planning phase. Choices faced by the participants before starting a trip (pre-trip choices).

  - Departure time choice. Participants can change the departure time of the trip to select earlier or later departures to accomplish the objectives of the trip.
  - Mode of transportation choice. Participants can choose between public and private transportation modes to complete the trips.
  - Route choice. Refers to the initial route choice of the participants. Depending on the mode choice, different alternative routes are presented to the participants. From these alternatives they are required to choose one.

- En-route phase. Choices faced by the participants once the trip is started.

  - Rerouting. The participants can change the current route when they consider the proposed alternatives are better.

Score. At the end of each trip, participants receive a numerical score that depends on the difference between the objective arrival time of the trip (informed in the statement of the problem) and the actual arrival time. The score is designed in order to capture the attention of participants in the experiments. However, as noted by Bogers et al. (2005), providing a score may influence the choices of participants, deviating their choices from what they would actually choose in a real situation towards a score maximising behaviour. The score is produced by a *score function*, a piece-wise linear function that can be customised to (i) penalise only late arrivals, or (ii) penalise arrivals outside an interval centered at the objective arrival time. To formally define the score functions, let t^*_k and t^{arr}_k be the objective and observed arrival time of trip k. Then, the score function, $\text{score}_1(t^{arr}, t^*_k)$, that penalises only late arrivals is given by

$$\text{score}_1(t^{arr}_k, t^*_k) = \begin{cases} S & t^{arr}_k \leq t^*_k \\ 0 & \text{otherwise} \end{cases}, \tag{2.1}$$

and the score function, $\text{score}_2(t^{arr}, t^*_k)$, that penalises arrivals outside the interval $[t^*_k - h_1, t^*_k + h_1]$ is given by

$$\text{score}_2(t^{arr}_k, t^*_k) = \begin{cases} 0 & t^{arr}_k < t^*_k - h_2 \\ \frac{S}{h_2 - h_1}(t^{arr}_k - (t^*_k - h_2)) & t^*_k - h_2 \leq t^{arr}_k < t^*_k - h_1 \\ S & t^*_k - h_1 \leq t^{arr}_k < t^*_k + h_1 \\ -\frac{S}{h_2 - h_1}(t^{arr}_k - (t^*_k + h_2)) & t^*_k + h_1 \leq t^{arr}_k < t^*_k + h_2 \\ 0 & t^*_k + h_2 \leq t^{arr}_k \end{cases}. \tag{2.2}$$

The plots of these score functions can be seen in Fig 2.6.

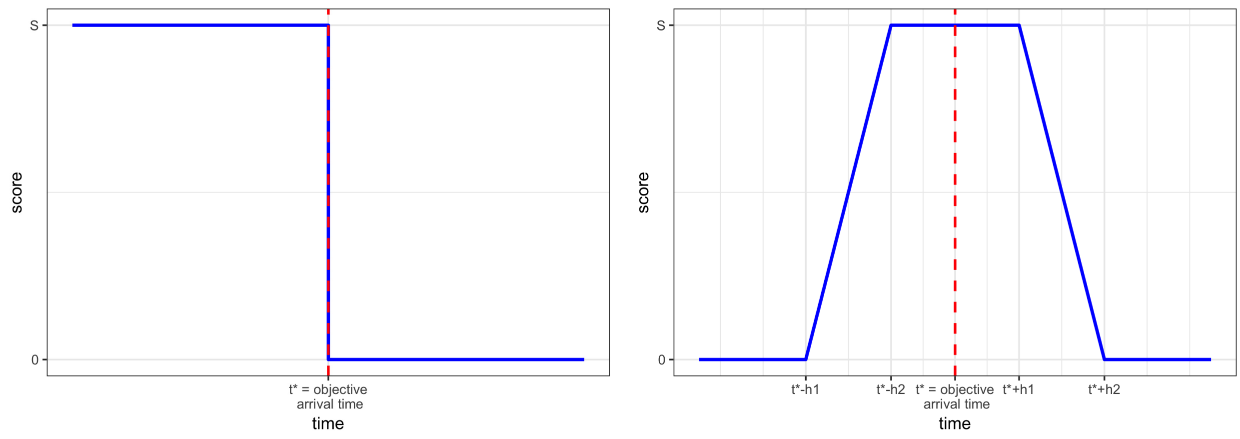

Figure 2.6: Shapes of the score function. Plots of the two score functions $score_1$ (left) and $score_2$ (right).

Traffic information. Extra information that can be provided to some or all the participants in order to study the influence that it might have in the choices of the participants.

- Messages. Are messages that can be sent to groups of participants with information that may influence their choices. This option was not used for the experiments in this thesis.

- Travel time information. It is the estimate of the travel time in the alternative routes. Travel time information allows the study of the influence of travel time on the choices of participants. Depending on whether the information is consulted during the trip planning phase or whether it is consulted when the trip is in the en-route phase, the estimates of the travel time correspond to historical information or real-time information, respectively. The historical information comes from a trip simulation with the same setting used in the experiment, while the real-time information comes from the simulation running in the experiment. The historical information is needed in the planning phase, since it is not possible to easily estimate the travel times for future trips. In both cases, the travel time information is obtained as the average travel time during a period of time on the alternative routes. The average travel time is computed at link level by measuring the average speed of the vehicles crossing the link during a period of time. Then, the travel time information on the alternative routes is the sum of the average travel times in the links that conform the routes. The travel time that is shown to the participants during the planning phase, is obtained for the currently selected departure time period. For example, if the currently selected departure time is set to 8:03, then the estimated travel time information corresponds to the average travel time of the trips following the same route that departed between 8:00 and 8:10; if the departure travel time is currently 8:16, then the information shown is that of the period between 8:11 and 8:20. The real-time travel time information is only shown at the points where the participants can re-route, and it corresponds to the average travel times on the alternative routes during the last time period. In both cases, the travel time information given to the participants is uncertain and no information on the distribution is provided.

- Congestion maps. Colours indicating the congestion state of the links in the network (green, orange, red). They are produced using the average speeds in the links. As for travel time information, the congestion maps in the planning phase come from historical information, and in the en-route phase from real-time information. This option was not used for the experiments in this thesis.

2.3 MDG gameplay

From the point of view of the participants, an experiment with the MDG consists in playing *missions*, which correspond to trips from an origin to a destination with a given purpose and an objective arrival time. When a player selects a mission, the target arrival time of the trip is displayed. Then, the player chooses the mode of transportation, the departure time and, finally, the route to complete the trip. Once the trip has started, the participants can re-route at predefined points. A mission is finished when the destination is reached before the target arrival time or when the target arrival time is not accomplished. A score, which is determined by a score function and depends on whether the objective of the mission is accomplished, is earned and informed after each mission. After a mission is finished, the participants may play another one. The actions that the participants can perform during a MDG session and their description are listed below; they are depicted in Fig. 2.7 and Fig. 2.8 with the corresponding MDG interface.

1. Select mission. A mission corresponds to a trip with a purpose, such as going to work or to the train station, and an objective arrival time. The origin and the destination of the trip are placed on the map interface, and the objective is given to the participant.

2. Choice of departure time. The players can adjust their desired departure time to earlier or later time.

3. Mode choice. The participants can choose several modes to complete their trip. An estimated cost for each mode is given to the travellers.

4. Route choice. The participant chooses one of the three alternative routes that are proposed in the MDG interface.

5. Reroute. Participants can change route once the trip has already started in predefined choice points. New alternatives are proposed to the player, and, in the case they are

receiving traffic information, the travel time estimate is shown along the proposed alternatives.

When travel time information is available to the participant, it is updated according to the choices that are being explored (mode of transportation, route and departure time), i.e., before the choices are made.

2.3.1 Interface

The interface of the MDG consists in the full map of a real road and public transport network, such as Lyon, and the control panels with which the participants interact with the game. Also, messages shown on the screen are part of the MDG interface. The elements of the MDG interface are described below.

1. Log-in and registration screen.
2. Main interface. Composed of the map of the road and public transport network of the city, displayed on the whole screen.
3. Mission menu. It is located at the bottom-right of the main interface and allows the participants to select the mission they want to play.
4. Choice menu. Once a mission is selected, the choice menu appears on the left of the main interface. All the choices pertaining the departure time, the mode, and the route choice are made using this panel.
5. Messages. The objective of the missions and the score after a trip is finished are presented to the participant in the form of messages placed at the centre of the screen.
6. Traffic information. The traffic information is presented as number near the origin of the trip (or the reroute point) or and as a congestion map.

2.4 Experimental design

The design of an experiment is led by the particular behavioural traits that are being investigated. This is achieved by the adequate definition of the decision problems and scenarios in the MDG. The scenarios for the route choice experiments were selected based on the variables that influence the route choices of travellers (Bovy and Stern, 1990). These variables, however, are based on real-world choices where decisions are the result of both navigational and map-reading tasks. Since the choices in the MDG are limited to map-reading tasks, many of these

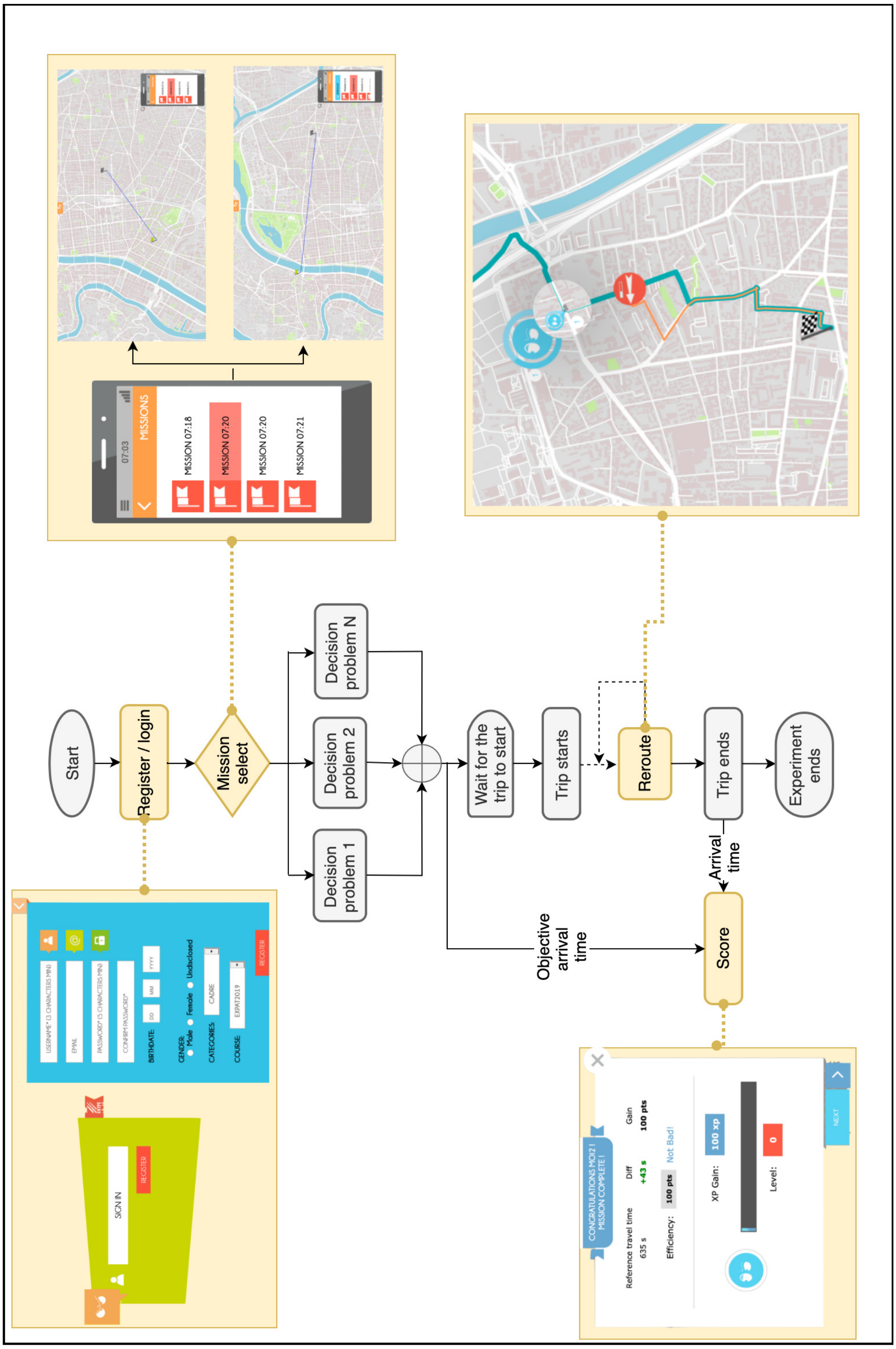

Figure 2.7: MDG gameplay with interface description. The flow chart of the experiment is presented alongside its interface.

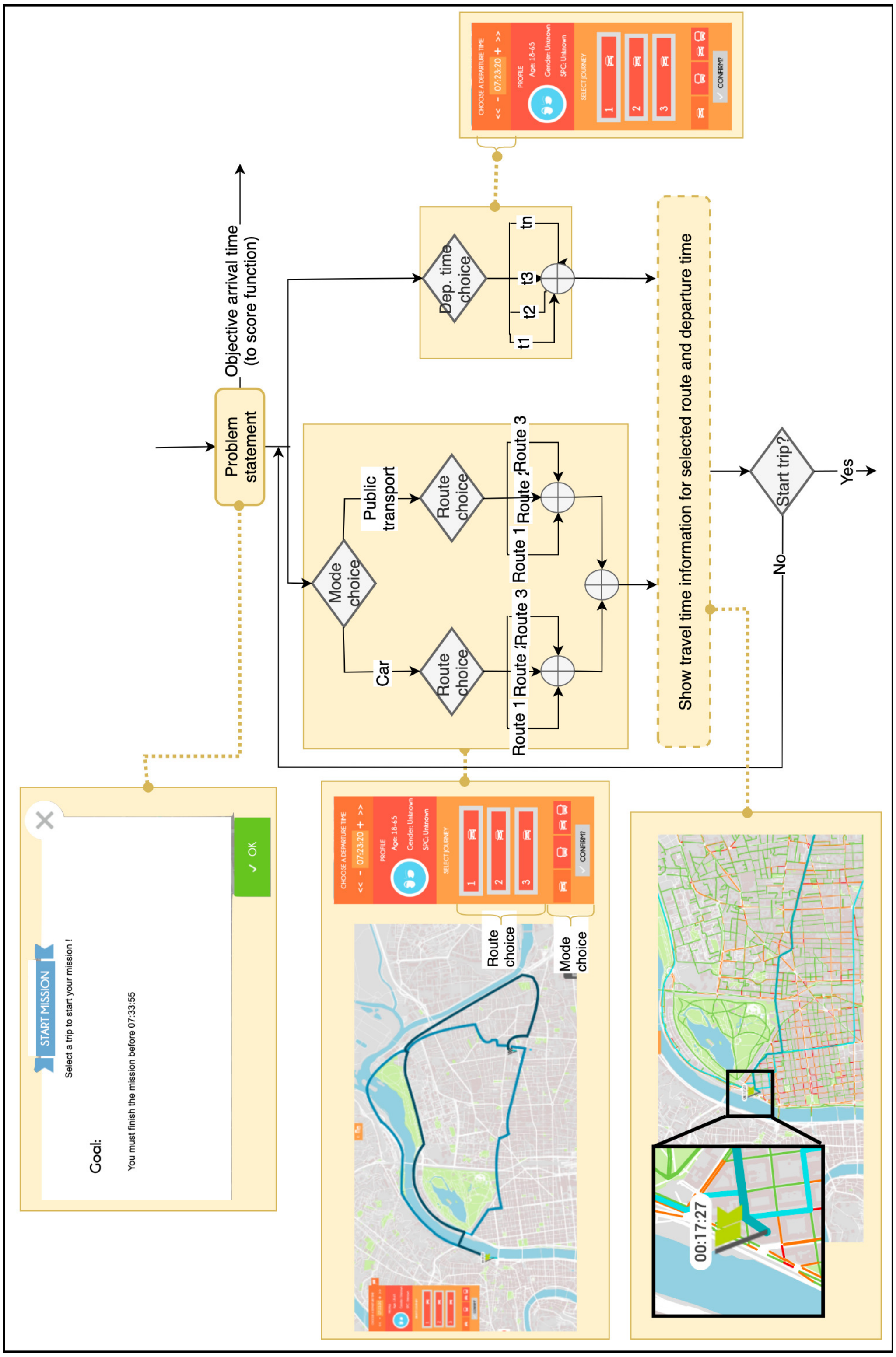

Figure 2.8: MDG gameplay with interface description. The flow chart of the decision problem is presented alongside its interface.

variables are irrelevant for the design of the experiments because they cannot be experienced by the participants. For example, the road surface, the lighting and the weather. The variables relevant for the design of the scenarios in the MDG experiments are listed below.

- $EDIST_{od}$ - Euclidean distance from origin to destination.
- LEN_j - Length of the route j in kilometres.
- DIR_j - Directness of the route j, the directness is defined as $DIR_j = EDIST_{od}/LEN_j$.
- TNR_j - Number of turns per kilometre in route j.
- INT_j - Number of intersections per kilometre in route j.
- FRW_j - Freeway composition of the route j, defined as the percentage of the route length that is composed of freeway segments.
- ITT_{js} - Informed travel time in the route j in choice problem s.

The OD pairs and routes were selected so that the values of the above-mentioned attributes show a significant variation across routes, while the routes remain plausible alternatives. This was achieved by defining short and long OD pairs from and to the cardinal directions and using the route planning of Google Maps® to obtain three candidate alternative routes. Some of the alternative routes were modified with the purpose of obtaining more variation in the attributes, for example by forcing one of the alternatives to use the ring road, or to have many turns. This was done by placing intermediate points in the planned routes. The traffic conditions in the selected OD pairs were obtained by modifying the local travel demand and by modifying the assignment of the trips in the three routes. As in the selection of the playable OD pairs, the local demands and the assignments were chosen such that the travel times, and hence the travel time information between the routes vary, both between the alternative routes and the period of the simulation. For example, in some OD pairs the demand was configured such that the ring-road alternative was the fastest during all the experiment, while in other OD pairs or experiments it was the fastest in some periods but not in others. Since the MDG in its actual state does not permit to study the learning behaviour of participants, travel time information acts as a proxy for the travel times on the alternative routes. The reaction of participants to the travel time is therefore studied through travel time information.

In total, 41 playable OD pairs were defined for the MDG experiments, 15 OD pairs in the Lyon-V36 network and 26 OD pairs in the Lyon-full network.[2] The values of the attributes of

[2]Nine of the OD pairs in the Lyon-full network were obtained following a different methodology, described in Chapter 4.

these OD pairs and routes are summarised in Fig. 2.9, where they are compared to the values of the attributes of all the OD pairs and routes in the network. The attributes of the 41 OD pairs and routes are included in appendix 2.A, and their maps in appendix 2.B.

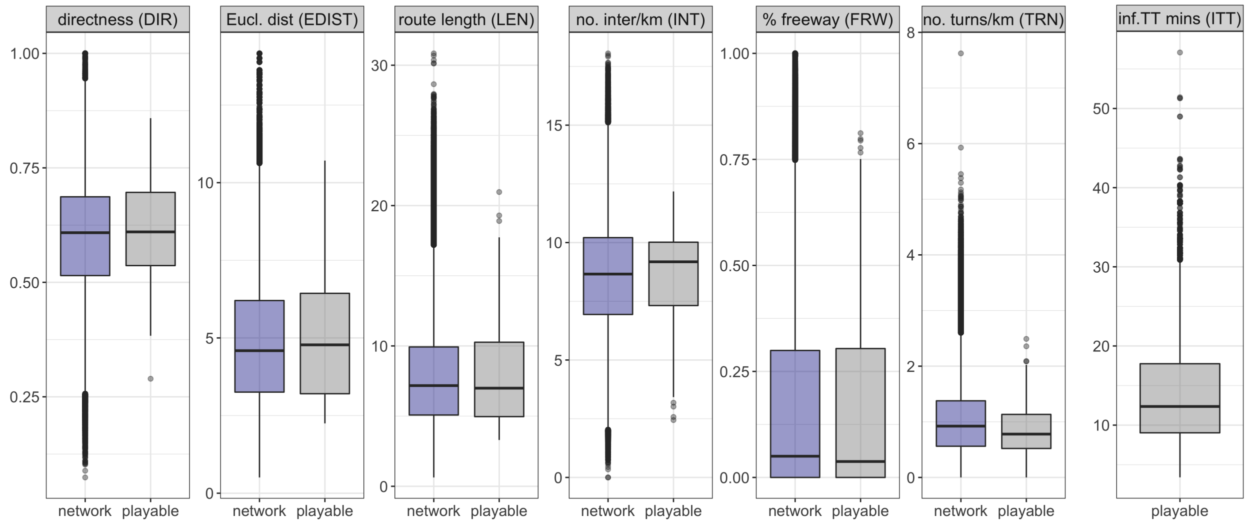

Figure 2.9: Summary of the attributes of the OD pairs and routes. Distributions of the attributes experienced by the travellers in the OD pairs and three alternative routes defined for the MDG experiments.

2.4.1 List of experiments with the MDG

The data on route choice behaviour comes from 9 MDG experiments carried out between February 2018 and May 2019. The participants in the experiments were mainly students at the University of Lyon attending the courses of traffic theory, staff from the IFSTTAR (French Institute of Science and Technology for Transport, Development and Networks) and other universities, who received an invitation by e-mail to remotely join the experiments via a web browser. All participants signed an informed consent form (see appendix 2.C) describing the objectives of the study, the data collection and processing, and the confidentiality rules. Participants could opt out of the experiment at any time. No personal data were mandatory to participate in the experiment, as people had the opportunity to identify themselves by a login of their choice. Finally, all data were fully anonymised and processed as such. At the beginning of the experiment, the participants were briefed about the objective of the experiment and the interface of the experimental platform; for the participants that joined the experiments via web, a document with the instructions was shared. The summary of the experiments is presented in Table 2.2.

Table 2.2: Summary of the MDG experiments.

No.	Date	Purpose	Public	Network	No. OD	Participants	No. choices
1	12/04/2017	RC	Univ. Lyon	Lyon-V36	3	201	743
2	12/04/2017	RC	Univ. Lyon	Lyon-V36	2	15	199
3	12/04/2017	RC	Univ. Lyon	Lyon-V36	2	25	362
4*	15/02/2018	RC	IFSTTAR/Other. Univs.	Lyon-V36	15	76	2,591
5*	26/04/2018	RC	IFSTTAR/Other. Univs.	Lyon-full	15	53	216
6*	02/05/2018	RC	Univ. Lyon	Lyon-full	15	193	1,186
7*	22/06/2018	RC	Univ. Lyon	Lyon-V36	10	108	940
8	04/09/2018	RR/DT	General(hEART Conf.)	Lyon-V36	3	67	175
9	10/10/2018	DT	Univ. Lyon	Lyon-V36	2	11	147
10	13/10/2018	TP	General (Fête de la Science)	Lyon-V36	3	<10	22
11	08/11/2018	MC/RC	General (TUBA)	Lyon-full	5	<10	96
12*	20/11/2018	RC	IFSTTAR	Lyon-full	12	27	166
13*	04/12/2018	RC	Other. Univs.	Lyon-full	12	19	337
14*	25/01/2019	RC	Univ. Lyon	Lyon-full	9	25	243
15*	01/02/2019	RC	Univ. Lyon	Lyon-full	9	17	185
16*	10/05/2019	RC/DT	Univ. Lyon	Lyon-full	5	199	1,646

CODE: RC=route choice, DT=departure time choice, MC=mode choice, TP=trip purpose, RR=reroute

NOTE: Data used in this thesis come from the experiments marked with *.

2.A Attributes of the OD pairs

Table 2.3: Attributes of OD pairs and routes in Lyon-V36 network. Attributes experienced by participants in the MDG experiments.

		Map-reading variables						Travel time info.			
OD	Route	*EDIST*	*LEN*	*DIR*	*TNR*	*INT*	*FRW*	min	max	mean	s.d.
	R1	5.20	6.40	0.80	0.31	10.11	0.04	12.80	35.90	18.50	5.20
O01D01	R2	5.20	7.00	0.74	0.72	10.01	0.03	15.70	20.90	18.00	1.50
	R3	5.20	8.40	0.62	0.12	4.18	0.57	10.20	19.60	12.40	2.00
	R1	4.70	5.90	0.78	0.51	9.96	0.04	11.80	30.30	16.90	4.30
O02D01	R2	4.70	6.10	0.76	0.33	9.82	0.04	13.40	17.20	15.10	1.30
	R3	4.70	8.90	0.52	0.00	5.05	0.54	11.30	20.00	13.60	1.70
	R1	2.40	3.70	0.60	0.81	11.82	0.00	7.10	9.30	7.90	0.60
O03D03	R2	2.40	3.80	0.58	1.05	10.81	0.00	6.70	12.90	9.30	1.80
	R3	2.40	3.80	0.58	0.78	10.65	0.00	7.50	29.30	11.90	4.70
	R1	3.60	4.50	0.76	0.67	9.21	0.00	10.80	16.00	13.20	1.30
O04D03	R2	3.60	4.70	0.72	1.07	8.16	0.00	7.70	10.10	9.30	0.70
	R3	3.60	5.10	0.66	0.39	5.29	0.31	9.00	11.30	10.10	0.70
	R1	4.40	5.90	0.54	0.85	8.62	0.03	11.60	23.20	16.40	3.20
O05D04	R2	4.40	6.30	0.51	0.00	3.18	0.75	5.10	6.10	5.60	0.30
	R3	4.40	6.60	0.49	0.61	9.84	0.03	10.40	15.10	11.80	1.30
	R1	5.20	5.90	0.86	0.17	9.22	0.04	12.30	15.40	13.80	0.80
O06D05	R2	5.20	6.80	0.74	0.88	10.13	0.04	14.60	35.30	20.40	5.00
	R3	5.20	9.60	0.53	0.21	5.43	0.50	15.50	24.30	17.80	2.10
	R1	3.50	5.50	0.60	0.91	8.60	0.03	12.40	31.40	17.30	4.50
O07D06	R2	3.50	6.40	0.52	0.79	3.62	0.77	6.00	6.60	6.30	0.20
	R3	3.50	6.40	0.51	1.25	9.19	0.03	9.30	11.40	10.00	0.60
	R1	2.90	3.30	0.69	0.60	3.59	0.28	3.40	4.00	3.50	0.10
O08D07	R2	2.90	3.40	0.68	1.17	10.27	0.00	5.80	6.70	6.20	0.30
	R3	2.90	4.00	0.59	1.52	8.59	0.00	6.30	7.10	6.60	0.20
	R1	3.10	4.00	0.78	0.75	8.48	0.05	7.50	16.90	9.40	2.20
O09D08	R2	3.10	4.30	0.73	0.70	8.38	0.05	9.50	12.50	10.80	0.90
	R3	3.10	5.10	0.62	0.20	9.41	0.38	10.00	11.40	10.90	0.40
	R1	3.10	4.90	0.59	2.02	8.49	0.04	7.40	9.40	8.00	0.60
O10D02	R2	3.10	5.20	0.56	0.58	8.14	0.04	9.80	14.40	11.70	1.20
	R3	3.10	5.40	0.54	0.93	8.96	0.04	9.30	11.10	10.20	0.50
	R1	2.70	3.80	0.66	1.84	11.30	0.00	8.20	11.10	9.50	0.90
O11D11	R2	2.70	4.00	0.63	2.01	9.80	0.00	9.10	11.80	10.20	0.80
	R3	2.70	4.70	0.53	0.85	8.07	0.00	8.20	9.00	8.60	0.10
	R1	2.70	3.90	0.59	0.77	7.48	0.00	7.80	9.10	8.40	0.30
O12D12	R2	2.70	4.20	0.54	1.91	9.56	0.00	8.00	9.40	8.50	0.40
	R3	2.70	4.60	0.50	1.09	9.81	0.00	8.50	10.60	9.50	0.50
	R1	3.20	3.90	0.74	0.78	9.84	0.00	9.70	12.70	11.40	0.80
O13D13	R2	3.20	4.30	0.66	2.09	11.62	0.00	8.60	13.50	9.60	0.80
	R3	3.20	5.00	0.57	1.41	9.23	0.00	11.00	13.60	12.40	0.70
	R1	2.80	3.40	0.62	2.36	10.02	0.00	6.30	9.80	7.70	1.30
O14D14	R2	2.80	3.80	0.55	1.57	9.92	0.00	6.30	11.70	8.00	1.20
	R3	2.80	4.00	0.52	0.99	10.88	0.00	7.90	10.80	9.10	0.70
	R1	4.10	4.60	0.84	0.87	10.93	0.00	9.80	25.00	13.20	3.10
O15D15	R2	4.10	5.40	0.71	0.74	9.94	0.00	12.70	36.00	16.60	5.30
	R3	4.10	6.20	0.63	0.49	11.85	0.00	13.40	15.10	14.10	0.50

Table 2.4: Attributes of OD pairs and routes in Lyon-full network. Attributes experienced by participants in the MDG experiments.

		Map-reading variables						Travel time info.			
OD	Route	*EDIST*	*LEN*	*DIR*	*TNR*	*INT*	*FRW*	min	max	mean	s.d.
	R1	10.70	13.70	0.78	0.15	5.57	0.79	16.30	18.10	16.40	0.30
O16D16	R2	10.70	18.90	0.56	0.16	3.02	0.78	16.20	18.30	17.20	0.70
	R3	10.70	21.00	0.51	0.14	2.58	0.73	22.30	27.70	23.80	1.40
	R1	10.50	14.80	0.70	0.41	8.58	0.27	27.60	32.80	28.90	1.00
O17D17	R2	10.50	15.20	0.68	0.00	3.55	0.75	15.80	17.60	16.50	0.40
	R3	10.50	19.30	0.54	0.26	3.42	0.81	18.30	21.00	19.50	0.80
	R1	6.40	7.90	0.78	0.66	9.79	0.00	19.70	24.00	21.60	1.30
O19D19	R2	6.40	8.50	0.71	1.21	10.19	0.00	20.80	24.70	22.40	1.00
	R3	6.40	10.80	0.57	1.14	9.84	0.27	25.10	31.20	27.40	1.60
	R1	6.10	9.00	0.67	0.57	12.18	0.25	17.00	21.40	18.10	1.00
O20D20	R2	6.10	10.10	0.60	0.61	9.90	0.02	23.30	27.30	25.00	1.00
	R3	6.10	15.70	0.38	0.19	4.64	0.65	17.80	19.10	18.60	0.30
	R1	4.10	5.00	0.76	1.05	10.20	0.00	12.90	15.70	14.40	0.90
O21D21	R2	4.10	5.40	0.70	0.97	11.42	0.25	13.80	17.90	16.10	1.20
	R3	4.10	6.40	0.60	0.81	10.94	0.00	12.90	17.00	14.60	1.10
	R1	8.20	10.20	0.80	0.40	9.60	0.18	19.70	31.90	22.00	3.20
O22D22	R2	8.20	11.20	0.73	0.73	9.12	0.00	27.30	43.60	30.40	3.00
	R3	8.20	17.60	0.46	0.46	4.70	0.45	22.60	29.70	25.60	2.20
	R1	6.40	8.40	0.78	1.10	9.29	0.66	13.30	18.20	14.60	1.20
O23D23	R2	6.40	10.10	0.65	1.32	9.52	0.03	19.80	22.20	20.50	0.70
	R3	6.40	10.50	0.62	1.07	7.68	0.00	23.70	26.30	24.00	0.50
	R1	9.20	14.20	0.62	1.29	8.95	0.00	32.10	42.60	34.20	2.20
O24D24	R2	9.20	14.60	0.60	0.56	8.49	0.24	27.80	39.60	31.70	3.10
	R3	9.20	17.10	0.52	0.53	5.97	0.37	27.30	51.40	33.60	6.50
	R1	5.90	6.70	0.83	0.93	8.68	0.00	15.10	17.30	16.10	0.50
O25D25	R2	5.90	8.10	0.69	1.15	9.18	0.00	16.30	19.60	18.50	1.30
	R3	5.90	10.70	0.52	0.57	3.84	0.56	13.20	15.50	14.40	0.70
	R1	7.20	8.90	0.80	0.46	12.09	0.10	20.60	33.60	23.50	2.40
O26D26	R2	7.20	12.60	0.57	0.49	8.67	0.50	19.90	25.80	22.00	1.40
	R3	7.20	14.20	0.51	0.07	4.30	0.56	15.80	20.60	17.20	0.90
	R1	5.60	8.20	0.67	0.63	6.99	0.35	12.50	16.70	14.70	1.30
O27D27	R2	5.60	9.40	0.58	1.09	8.62	0.00	19.50	23.50	21.20	1.10
	R3	5.60	9.40	0.58	0.54	9.33	0.27	18.10	20.20	19.30	0.70
	R1	10.70	13.00	0.82	0.08	6.86	0.21	18.30	21.90	19.20	0.80
O28D28	R2	10.70	17.40	0.61	0.87	8.77	0.15	38.00	57.10	38.90	2.30
	R3	10.70	17.70	0.60	0.29	5.47	0.29	30.90	40.20	32.20	1.30
	R1	5.80	8.10	0.72	0.38	7.15	0.00	14.20	18.30	15.60	0.90
O29D29	R2	5.80	8.60	0.68	0.96	7.12	0.00	17.70	22.20	19.50	1.00
	R3	5.80	9.00	0.65	0.23	2.44	0.80	9.10	11.30	10.00	0.60
	R1	8.00	10.50	0.75	0.68	8.74	0.32	20.40	22.30	21.00	0.40
O30D30	R2	8.00	10.80	0.73	1.42	12.09	0.00	25.60	33.80	28.50	2.00
	R3	8.00	12.90	0.61	0.87	6.99	0.14	26.10	30.50	27.70	1.10
	R1	6.30	8.40	0.70	0.86	9.42	0.25	15.20	22.00	18.00	2.50

Continuation of Table 2.4...

		Map-reading variables						Travel time info.			
OD	Route	*EDIST*	*LEN*	*DIR*	*TNR*	*INT*	*FRW*	min	max	mean	s.d.
CC1	R2	6.30	8.70	0.68	1.19	11.21	0.09	17.10	21.40	19.00	1.20
	R3	6.30	10.30	0.57	0.79	8.03	0.35	19.90	22.70	21.40	1.00
	R1	2.20	3.30	0.66	1.96	10.90	0.00	8.10	9.60	9.10	0.30
CC2	R2	2.20	3.60	0.60	2.08	8.59	0.00	8.60	9.90	9.40	0.40
	R3	2.20	3.90	0.56	1.39	11.68	0.00	8.90	10.50	10.00	0.40
	R1	6.80	9.90	0.66	0.72	10.67	0.29	19.10	34.10	26.30	4.30
CC3	R2	6.80	10.30	0.64	0.70	9.02	0.14	23.00	31.40	26.10	2.10
	R3	6.80	11.10	0.59	0.65	9.28	0.13	22.80	28.10	24.80	1.30
	R1	4.80	6.50	0.71	1.13	10.67	0.00	12.50	13.40	13.00	0.30
CC4	R2	4.80	7.10	0.65	1.17	10.31	0.00	16.00	17.70	16.90	0.60
	R3	4.80	10.10	0.46	0.41	6.92	0.32	18.00	20.50	19.40	0.80
	R1	6.60	9.90	0.67	0.73	11.33	0.07	24.10	34.60	28.40	3.20
CC5	R2	6.60	12.30	0.54	0.66	9.03	0.35	22.90	32.80	25.80	2.30
	R3	6.60	15.70	0.42	0.77	5.14	0.50	22.50	26.70	24.40	1.20
	R1	7.90	10.00	0.69	0.31	7.86	0.24	16.00	20.00	18.10	1.30
CC6	R2	7.90	13.30	0.52	0.31	6.48	0.32	20.00	26.90	23.10	1.10
	R3	7.90	14.50	0.47	0.35	7.64	0.49	22.00	27.80	24.40	1.40
	R1	3.80	5.40	0.68	1.16	11.43	0.15	14.50	17.80	15.90	0.70
CC7	R2	3.80	8.00	0.46	0.78	5.91	0.33	10.20	21.90	14.20	3.50
	R3	3.80	9.50	0.39	0.65	6.97	0.52	13.20	16.20	14.40	0.90
	R1	3.30	4.30	0.75	0.74	11.12	0.00	10.40	16.10	13.00	1.70
CC8	R2	3.30	4.80	0.67	1.31	9.92	0.00	8.70	16.50	12.40	2.70
	R3	3.30	5.60	0.58	1.32	9.72	0.00	12.10	14.30	13.20	0.60
	R1	3.00	5.20	0.55	1.43	11.06	0.00	11.60	14.30	12.50	0.80
CC9	R2	3.00	5.80	0.49	0.90	8.95	0.12	12.10	14.90	13.30	0.80
	R3	3.00	6.90	0.42	1.06	9.18	0.29	11.80	15.10	12.90	0.70
	R1	2.50	3.30	0.68	1.97	9.70	0.00	7.10	7.80	7.40	0.10
CR2	R2	2.50	3.80	0.59	1.97	9.73	0.00	6.80	7.30	7.10	0.20
	R3	2.50	4.70	0.48	2.49	10.07	0.00	8.40	9.10	8.60	0.30
	R1	4.80	8.80	0.52	1.52	10.09	0.00	21.10	23.10	21.90	0.60
CR7	R2	4.80	9.40	0.49	0.65	8.07	0.00	20.40	22.70	21.20	0.70
	R3	4.80	15.90	0.29	0.77	4.65	0.49	21.50	26.90	22.60	1.40
	R1	3.60	5.70	0.51	1.09	6.26	0.19	11.30	12.40	11.80	0.30
CR9	R2	3.60	6.40	0.46	1.14	9.86	0.00	12.50	13.60	13.10	0.40
	R3	3.60	7.40	0.39	1.11	7.65	0.00	16.80	20.90	17.70	1.10

2.B Maps of the OD pairs

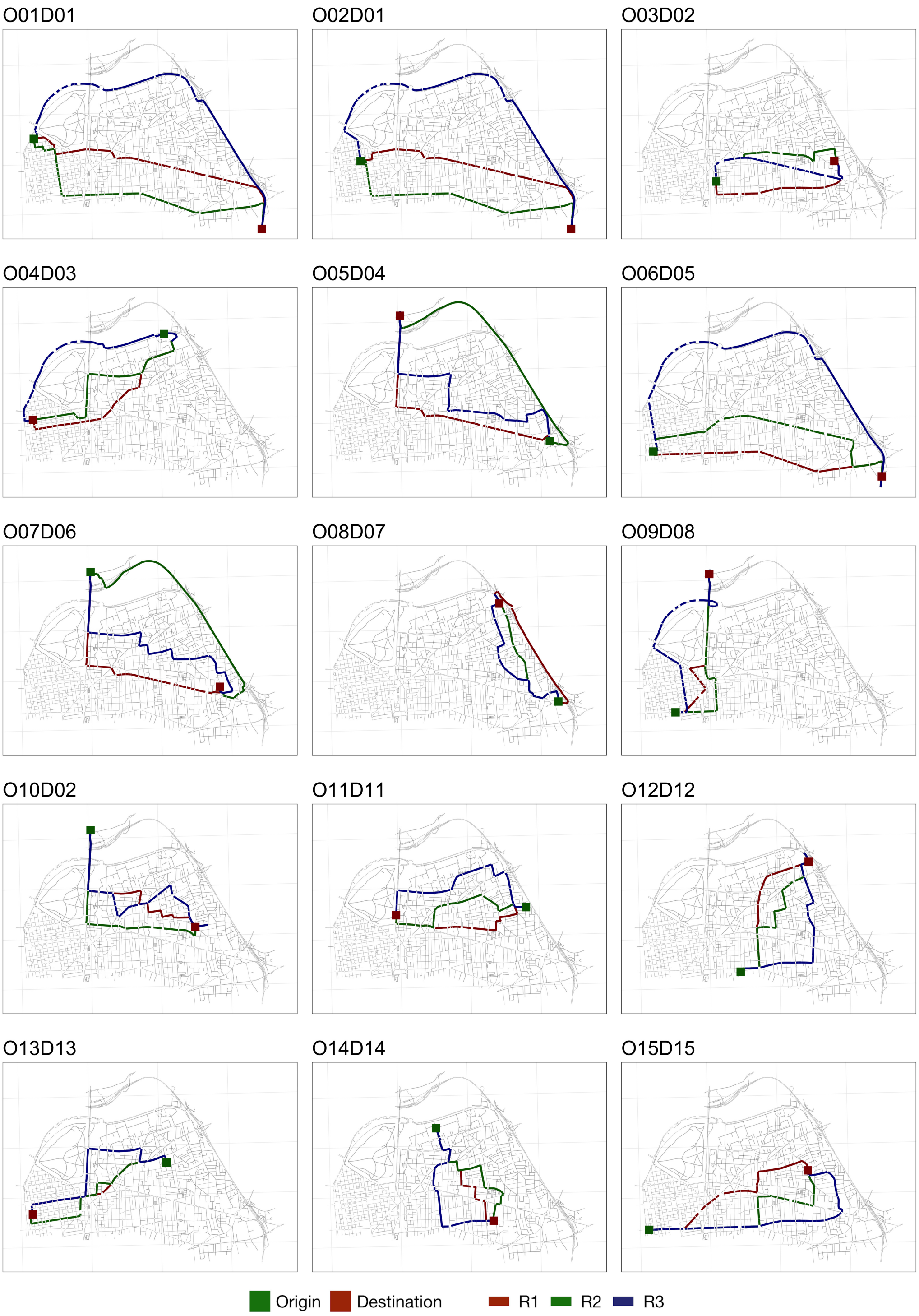

Figure 2.10: Maps of the playable OD pairs and routes in Lyon-V36 network.

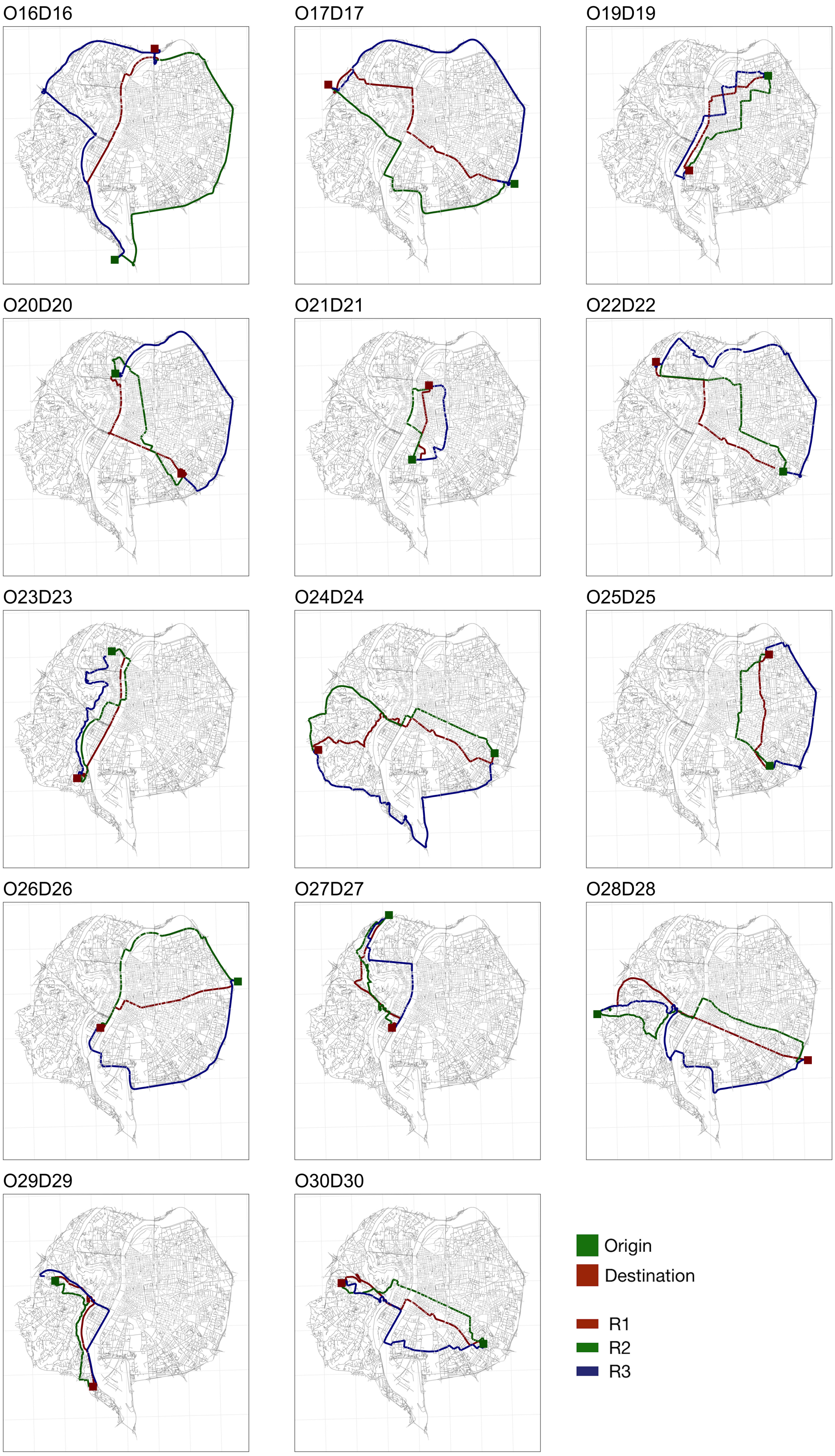

Figure 2.11: Maps of the playable OD pairs and routes in Lyon-full network.

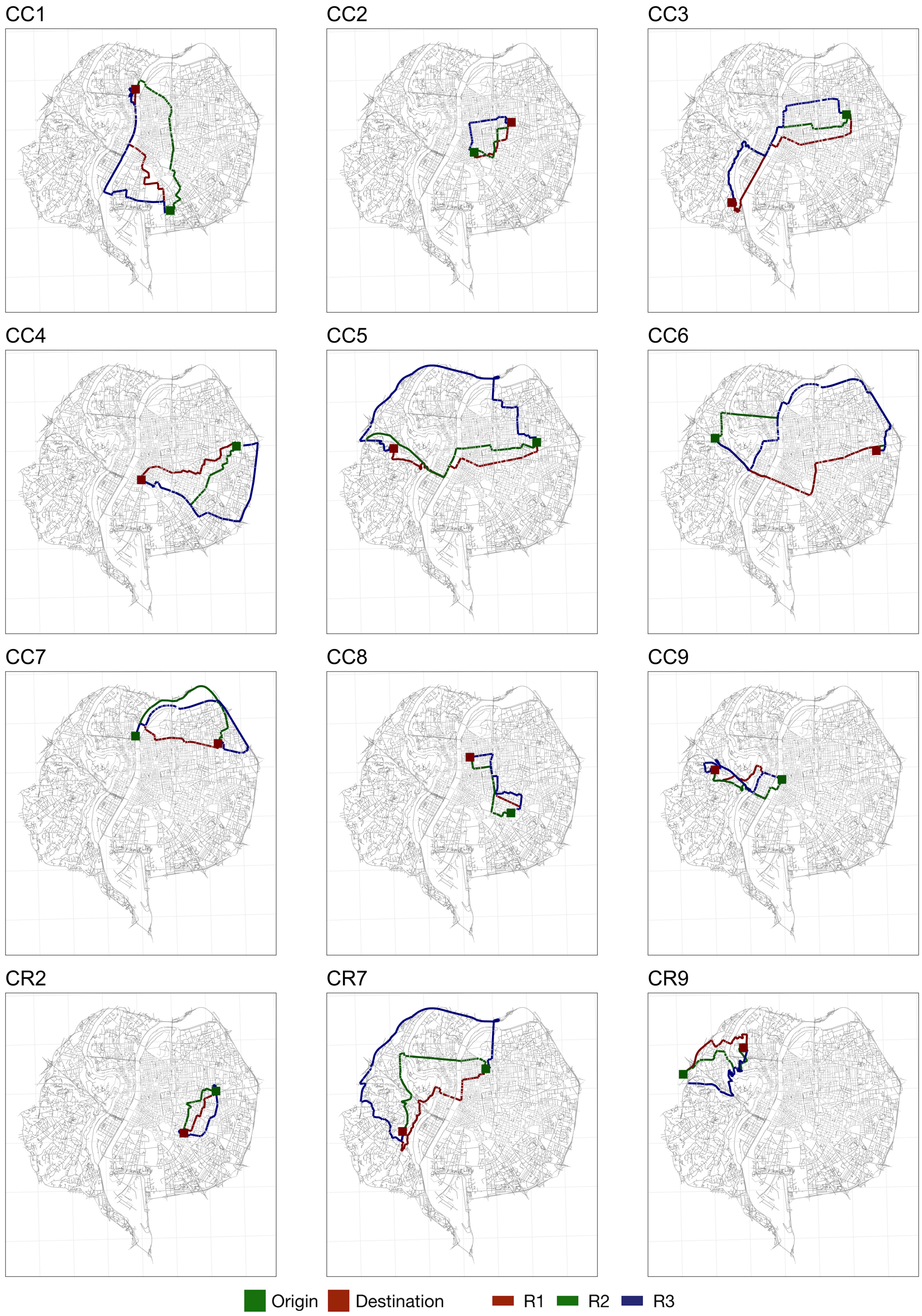

Figure 2.11: Maps of the playable OD pairs and routes in Lyon-full network.

2.C Consent form for the experiments' participants

IFSTTAR
INSTITUT FRANÇAIS
DES SCIENCES
ET TECHNOLOGIES
DES TRANSPORTS,
DE L'AMÉNAGEMENT
ET DES RÉSEAUX

ENTPE
L'école de l'aménagement durable des territoires

INFORMED CONSENT FORM

From Mrs, Mr Ludovic Leclercq, Principal Investigator (PI).
PI's Address:
COSYS – LICIT / ENTPE, Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France.

I have been invited to get involved into an IFSTTAR research study regarding a mobility simulation game based on traffic simulation in urban network. I have been free to accept or refuse.

I declare to be an adult with respect to the legislation of my country and to be at least 18 years of age.

I have received and I have understood following information:
The study concerns a real-time traffic simulator. I will play the role of a virtual agent described with agent characteristics (Socio-Professional Characteristics, Age interval, Gender). I will be assigned one or several mission(s) standing, for each mission, in moving from one origin area to one destination area, within a given time interval and others specific constraints (fuel consumption, pollution, etc.). I may be suggested to achieve secondary objectives. Before starting, I may be allowed to select my route and transportation mode and the departure time of my travel. During the displacement, I will be allowed to adjust some settings, depending on my knowledge of the traffic state. Information related to my choices and their consequences on global traffic will be recorded and studied. PI may invite me to several game sessions, and I will be free to participate. All these tasks will be carried on through a laptop, tablet or smartphone-based Internet application. The collected information will be post-treated, in order to evaluate the impact of user choices (regarding route choice, departure time and mode) on traffic dynamics at urban scale. The laboratory members involved in the study constitute data recipients.

I consent voluntarily to participate to this research.

I have well-informed that:
The information collected from this research project will be kept confidential. Personal information collected during the research will be only accessible to the PI and two core members of the research project. As acted in the law no78-17 of 6 January 1978 on Data Processing, Data Files and Individual Liberties, I have rights of access and of rectification with regard to the processing of personal data, that is applicable by asking PI Prof. Leclercq L., whom address is mentioned above.

My consent does not free researchers from their responsibilities. I keep all the rights with respect to the law. I know that I am entitled the object to the processing of my personal data, on legitimate grounds. At last, I am informed that this study is funded by the European commission, within the ERC-2014-CoG MAGnUM, ref 646592 (a Multiscale and Multimodal Traffic Modelling Approach for Sustainable Management of Urban Mobility), lead by Prof. Ludovic Leclercq.

Date ………………………

Name, Firstname

…………………………………..

Signature (read and approved)

Figure 2.12: Consent form signed by the participants.

Part I

Route choice behaviour

3 Literature review on route choice behaviour

This chapter presents, first, a short review of the literature on travellers' behaviour. The objective is to provide a coarse classification of the lines of research on studies that are found in the literature, mentioning some of the aspects of travellers' behaviour that have been extensively investigated. Then, a review of the literature on discrete choice models is given. The emphasis is placed on the derivation and estimation of the mixed logit model, which is the modelling approach chosen in this thesis.

3.1 Review of route choice studies

The study of travellers' behaviour can be divided in three non-exclusive categories: (i) the theoretical studies about the general laws describing human mobility, (ii) the empirical studies that try to prove or disprove a theory, and (iii) the studies about the factors that influence the mobility choices of travellers. To the first category belong the works of Brockmann et al. (2006), González et al. (2008), Song et al. (2010), Peng et al. (2012), Wang et al. (2012), Zhao et al. (2015) and Tachet et al. (2017) who describe the patterns of human mobility by physical or mathematical models, or the work of Marchetti (1994) that investigates these patterns from an anthropological point of view. To the second category belong the research of Iida et al. (1992), Bekhor et al. (2006), Selten et al. (2007), Papinski et al. (2009), Thomas and Tutert (2010) and Zhu and Levinson (2015) who try to see if travellers follow the shortest-time paths in their trips, or the work of Yildirimoglu and Kahraman (2018b) in which the authors try to determine if the UE hypothesis holds in road networks. The studies in the first two categories are mainly concerned with observation of the actual choices of travellers at large scale and thus the specific factors or determinants that influence the individual travel behaviour are, in general, unimportant. The third category is related to the factors, associated either to the traveller or the environment, that have an effect on the choices of the travellers.

Travellers' behaviour is a process which involves psychological and cognitive mechanisms through which travellers perceive the states of the network, and then make decisions accordingly (Bovy and Stern, 1990, Ben-Akiva et al., 1999). Although this definition may appear simple, there are many factors that intervene in this process, making it a complex problem. The determinants of travellers' behaviour are associated to both the traveller and the environment in which the trip is made. In the case of the traveller, the cognitive capacities, such as memory and spatial navigation skills, and risk attitudes have influence in the perception and thus the travel behaviour. In the case of the environmental factors, variables such as the travel time, travel time reliability, the attributes of the routes and purpose of the trip have an effect on the choices of the travellers. These factors are heterogeneous (as heterogeneous as individuals can be), and they interact in ways that are not easily observable to produce the choices. Therefore, the vast amount of studies on the aspects of travel behaviour that can be found in literature is not surprising.

Learning, familiarity and habit. Learning is an iterative process in which the outcomes experienced in previous trips, as well as the external information, build up travellers' subjective knowledge or perceptions on which they base their choices. However, as pointed by Bogers et al. (2005), Bogers (2005), the travellers' ability to learn from the past experiences is influenced by their memory and skills to process information; and the travellers' perceptions are altered by their risk attitudes. The type of information that travellers obtain can be classified into (i) experiential, (ii) descriptive and prospective (Ben-Elia and Avineri, 2015). The former is when the travellers' only source of information is their past experience; the latter refers to the information obtained from external sources. The two types of information are not mutually exclusive, for example, commute travellers are experienced in their daily trips and they may also consult traffic information to support their choices. Bogers (2005) performed a computer-based SP experiment, finding that memory helps in constituting more accurate beliefs and that travellers' choices rely more on the latest information. This observation is also made by Iida et al. (1992). The learning process involves exploring the different alternatives and/or consulting external travel information. Nonetheless, from a psychological point of view, acquiring information, either by experience or external information, is cognitively costly. Hence, to reduce the cognitive burden, travellers form habits. This explains why inexperienced travellers tend to explore different alternatives in order to gain knowledge on the states of the network, and when a satisfactory route is found they stick to it, behaviour observed in the route choice experiments by Adler and McNally (1994), Lotan (1997), Selten et al. (2007), Ben-Elia and Shiftan (2010), Vreeswijk et al. (2014), De Moraes Ramos et al. (2013). The cognitive skills and risk attitudes that explain, to some extent, the choices of travellers are not observable characteristics. However, they are expressed in other aspects of the behaviour that can be measured, in other words, they are latent variables. Taking this into account, Prato et al. (2012) and Kaplan and Prato (2012) propose a route choice model and perform experiments, finding that latent constructs for mnemonic, spatial and time saving abilities have a positive correlation with the preferences of individuals, suggesting that individuals with these characteristics tend to search better alternatives and tend to remember them. On the other hand, latent constructs of familiarity and habit have negative correlation, which suggests that individuals with these characteristics do not tend to search for better alternatives, even if their current choice is suboptimal. This observation is regarded as one of the main causes that support boundedly rational choices in travellers, as noted by Mahmassani and Chang (1987). Through a computer-based SP experiment, Srinivasan and Mahmassani (2000) conclude that increasing congestion may change the habitual route choice of travellers and that, when external information is available, the habit decreases.

Travel time reliability and risk aversion. From the external factors affecting the learning process, travel time reliability has been found to play a mayor role. In a SP study recording choices in two routes, Abdel-Aty et al. (1997) notice that the higher the travel time variation in the fastest but uncertain route, the less the travellers are willing to choose it. In other words, travellers exhibit risk aversion. The authors also observed that the larger the mean travel time difference between the two routes, the more people choose the fastest but uncertain. This implies that travellers recognise mean travel times and their variability, and that they try to minimise their travel times, while maintaining the variation is acceptable. A risk averse attitude towards travel time was also found in a study by Avineri and Prashker (2005), but adding that when the variance of the slow route is increased, participants chose it more. This last behaviour is in accordance with Ben-Elia and Shiftan (2010). Another observation in Avineri and Prashker (2005) is that, the higher the variance in travel time is, the lower is the traveller's sensitivity to travel time differences and the slower the rate of learning. A preference for reliable routes was explicitly stated by participants in Bogers (2005), Ramos (2015), however, in the latter study, GPS-based data showed that the chosen routes were actually amongst the least reliable. The above are general conclusions and describe the average behaviour, however risk aversion is heterogeneous amongst travellers, and between trip purposes. Risk-aversion and the preference for reliable routes depends on the value of time of the travellers, as pointed out by de Palma and Picard (2005), who noticed that the surveyed individuals in a SP and RP experiment were grouped in risk-averse and risk-neutral travellers (66% of their sample) and risk-seekers (33% of their sample), and that these numbers where related to the socioeconomic factors of the individuals. In Bogers (2005), De Moraes Ramos et al. (2013) it was found, as expected, that the proportion of participants who chose the reliable option is higher when the purpose of the trip is important. In a SP experiment, Bogers et al. (2006) also found that participants are risk-averse, adding that high variance in travel time has a larger effect on the choices of participants than extreme low frequency events. An extended review on the travel time variability and the risk attitudes of travellers is given in Taylor (2013).

Travel time information. The rapid development of information technologies in the last two decades has significantly increased the availability of travel time information, which travellers can incorporate in their decision process. The uncertainty of travellers on the traffic conditions of the network can be alleviated by the incorporation of external traffic information into their decision-making process, thus improving their choices Bonsall (1992). In a computer-based SP experiment, Ben-Elia and Shiftan (2010) found that participants who received real-time information showed higher levels of travel time minimisation compared to those who received no information, and that the information reduced their exploration rate. This suggests that information contributed to expediting learning, as suggested in Bogers (2005). Moreover, the authors observed that the information encouraged risk-seeking behaviour in the participants; result aligned with those in Abdel-Aty et al. (1997) and De Moraes Ramos et al. (2013). Ramos (2015) observed that travellers comply more to pre-trip than to en-route information, and that they use the information to plan their routes more than their departure times; behaviour also found in Mahmassani and Liu (1999). A travel time minimisation behaviour is also noticed in Adler and McNally (1994), Lotan (1997), Abdel-Aty et al. (1997), Srinivasan and Mahmassani (2000), Selten et al. (2007) and De Moraes Ramos et al. (2013). Nonetheless, the minimisation behaviour in the presence of information is attenuated by familiarity and habit, as pointed out by Lotan (1997), Liu and Mahmassani (1998) and De Moraes Ramos et al. (2013), who found that familiar travellers tend to consult less information. This result seems reasonable, since the travel time uncertainty is lower for familiar travellers, reducing the need to consult information. Obtaining and evaluating information (either experiential or descriptive or prospective) on a daily basis is cognitively costly, implying that habitual choices take more importance. However, the former can be more cognitively costly to obtain as it requires travellers to repeatedly explore the alternative routes and to remember the outcomes. This explains why travellers rely more on recent experiences when no information is available and, when information is available, travellers base their choices on both short and long-term experiences (Ben-Elia and Shiftan, 2010). A consequence of the travel time minimisation behaviour in the presence of travel time information is a deterioration of the network performance (Mahmassani and Jayakrishnan, 1991, Ben-Akiva and De Palma, 1991, Rapoport et al., 2014b, Ben-Elia and Avineri, 2015). This follows since travellers decisions are closer to be based on perfect information, thus moving the system towards the UE. However, there is current research on how information systems can also be used in order to move travellers' behaviour towards the system optimum, thus reducing the social and environmental costs (van Essen et al., 2016, Vreeswijk et al., 2015). A complete review on the response of travellers to travel information can be consulted in Ben-Elia and Avineri (2015).

3.2 Random utility models

The quantitative models of human choice behaviour have their origins in Expected Utility Theory. Proposed first by Bernoulli (1954), and later endowed with an axiomatic basis by Neumann and Morgenstern (1944), EUT allows an ordering of preferences over alternatives with uncertain outcomes. Consider an individual, labeled by i, facing a choice problem over a set of J alternatives, indexed by $j = 1, 2, ..., J$. For each alternative j, denote as Ω_j the set of (uncertain) outcomes if the alternative j were chosen. Since the outcomes Ω_j are uncertain, we assign to them a probability distribution, $Pr(\omega)$. EUT assumes that the decision-maker i evaluates the possible outcomes Ω_j resulting of choosing the alternative j via an utility function, $u_i : \Omega_j \to \mathbb{R}$. Thus, the expected utility that decision-maker obtains from choosing alternative j is given by

$$\mathbb{E}\left[u_i(\Omega_j)\right] = \int u_i(\omega) Pr(\omega) d\omega.$$

If the decision-maker satisfy the preference axioms of EUT, which are often referred to as *axioms of rational choice*, then he/she will prefer alternative j over alternative k if and only if $\mathbb{E}\left[u_i(\Omega_j)\right] > \mathbb{E}\left[u_i(\Omega_k)\right]$. In other words, EUT assumes that decision-makers will choose the alternatives that maximise their expected utility.

A generalisation of the EUT model was proposed by Savage (1954), by letting the probability $Pr(\omega)$ to represent the subjective beliefs of the decision-makers, rather than objective verifiable information. Thus, enabling the model to consider situations in which the decision-makers have incomplete information about the process generating the (objective) outcomes. In this case, the decision-makers are also considered rational, because they are maximising their utility based on their subjective beliefs. Although, from an individual perspective the EUT constitutes a well defined framework for individual decision making, it has some limitations when the choices are studied by an external observer. This is because neither the utilities that individuals get form the outcomes, u_i, nor the subjective probabilities that they assign to them, $Pr(\omega)$, can be easily observed by a bystander. The uncertainty that the external observer has on the utility of individuals led to the development of the random utility theory, based on the work of Thurstone (1927), Luce (1959) and Marschak (1959), and in the later development of the discrete choice Random Utility Maximisation (RUM) models by Mcfadden (1972). As in EUT, in RUM models it is assumed that individuals obtain a certain level of utility from each alternative in a choice situation, and that they choose the alternative with the maximum utility, i.e., individuals are utility maximisers. However, in contrast to the Expected Utility Theory, in RUM models the utilities cannot be directly observed by a bystander; what can be observed by the bystander are the actual choices, and the attributes of both the alternatives and the decision-makers.

To formally derive the RUM models, let a decision-maker, labelled i ($i \in I$), face a choice amongst J_i alternatives. The set of the alternatives is known as the *choice set*, and it must meet three requirements. First, the alternatives must be mutually exclusive, meaning that the decision-maker only chooses one alternative from the choice set. Second, the choice set is exhaustive, this is, the decision-maker necessarily chooses one of the alternatives. Three, the number of alternatives must be finite. Let the alternatives in the choice set be labelled by $j = 1, \ldots, J_i$. The decision-maker i obtains a level of utility from each alternative j, this utility is denoted by U_{ij}. Finally, it is assumed that the decision-maker i chooses the alternative that maximises her utility, i.e., the decision-maker chooses j if $U_{ij} > U_{ik}$ for all $k \neq j$, $k, j \in J_i$. However, contrary to EUT, in the RUM models the utilities U_{ij} are not observed by the bystander. Therefore, from a bystanders' point of view, the utility that an individual i obtains from alternative j, is a random variable, written as

$$U_{ij} = V(\mathbf{x_{ij}}; \boldsymbol{\beta}) + \varepsilon_{ij}, \tag{3.1}$$

where $\mathbf{x_{ij}}$ is a vector of explanatory variables describing individual i and alternative j; $\boldsymbol{\beta}$ is a vector of unknown coefficients; $V_{ij} = V(\mathbf{x_{ij}}; \boldsymbol{\beta})$ is a function of the explanatory variables $\mathbf{x_{ij}}$ and the coefficients $\boldsymbol{\beta}$; and ε_{ij} is a random disturbance for i and j, with $\mathbb{E}[\varepsilon_{ij}] = 0$ for all $i \in I$ and $j \in J_i$.

The random variable ε_{ij} captures the unobserved factors that affect the utility, but that are not present in V_{ij}. Note that if no random term ε_{ij} were included in Eq. (3.1), then the utility that two different individuals, i and i', get from an alternative j would be equal if $\mathbf{x_{ij}} = \mathbf{x_{i'j}}$. Nonetheless, this assumption is unrealistic. First, because not all the characteristics of individuals influencing their choices can be objectively measured; and second, because it would be expected that two different individuals present variations on the utilities that they get from the same alternative. Therefore, the function V is referred to as the systematic part of the utility, as it captures the expected utility that an individual with characteristics $\mathbf{x_{ij}}$ gets from an alternative j, i.e.,

$$\mathbb{E}[U_{ij}] = V(\mathbf{x_{ij}}; \boldsymbol{\beta})\,,$$

and the random variable ε_{ij} is known as the idiosyncratic term, as it reflects the deviations of different individuals from the expected utility due to unobserved (measured) attributes or idiosyncrasy.

To determine the utilities it is necessary to estimate the coefficients in the vector $\boldsymbol{\beta}$ that better fit the observations. This is a latent regression problem, since the utilities U_{ij} cannot be directly observed: they are random variables. What can be observed, however, are the measurable characteristics of the individuals and alternatives, $\mathbf{x_{ij}}$ and the actual choices y_{ij}, the later variable being binary and taking the value $y_{ij} = 1$ when individual i chose alternative j, and $y_{ij} = 0$ otherwise. Since individuals are assumed to be utility maximisers, then the (observed) chosen alternative is the one that maximizes the utility, i.e.,

$$y_{ij} = \begin{cases} 1, & U_{ik} < U_{ij} \; \forall \; k \neq j \\ 0, & \text{otherwise.} \end{cases}$$

This relationship relates the unobserved utilities to the actual choices of the individuals. Nevertheless, since the utilities are random variables, the relationship is given through the probabilities

$$\begin{aligned} Pr(y_{ij} = 1) &= Pr(U_{ik} < U_{ij} \; \forall \; k \neq j) \\ &= Pr(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik} \; \forall \; k \neq j). \end{aligned} \tag{3.2}$$

In words, the above expression is read as *the probability that individual i chooses alternative j is equal to the probability of i obtaining the highest utility from j.*

The expression in Eq. (3.2) is the general form of the additive RUM models, and it corresponds to the cumulative distribution of the joint distribution of ε_{ij} for all $j = 1, \ldots, J_i$. The RUM models are specified by defining both, the form of the systematic part of the utility, V_{ij}, and the distribution function of the disturbances ε_{ij}. Usually, the systematic part of the utility is specified to be linear in the parameters $\boldsymbol{\beta}$, i.e., $V(\mathbf{x_{ij}}; \boldsymbol{\beta}) = \mathbf{x_{ij}}^T \boldsymbol{\beta}$. In this case, the estimation of $\boldsymbol{\beta}$ is a convex non-linear optimisation problem, which is computationally efficient to solve. The interpretation of the coefficients β_p is the change in the expected utility by a unit change in the attribute x_{ijp}. In terms of behaviour, the coefficient β_p is referred to as the taste or preference of the decision-maker with respect to the attribute x_{ijp}.

The assumption over the distribution of the random variables ε_{ij} gives rise to different random utility models and, hence, to the different forms of the probabilities. The multinomial logi (MNL) model, result of the works of Luce (1959), Marschak (1959) and Mcfadden (1972), is obtained by assuming that the disturbances ε_{ij} are independent and identically distributed (i.i.d.) extreme value random variables. The MNL model has the advantage of having a closed form for the probabilities $Pr(y_{ij} = 1)$ in Eq. (3.2). However, the i.i.d. assumption of the disturbances ε_{ij} implies that the unobserved factors are uncorrelated for the different alternatives; assumption that, while convenient in some situations, it can be restrictive in others. The generalized extreme value models (GEV), credited to Williams (1977) and Mcfadden (1978), were developed to overcome this drawback. The GEV model considers correlations between the unobserved part of the utilities, maintaining the close form of $Pr(y_{ij} = 1)$. The probit model (Thurstone, 1927, Marschak, 1959, Daganzo, 1979, Sheffi et al., 1982), obtained by assuming that the vector $\varepsilon = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})^T$ is multivariate normally distributed $\mathcal{N}_J(\mathbf{0}, \Sigma)$, also allows for the alternatives to be correlated. Nevertheless, it lacks of a closed form for the probabilities $Pr(y_{ij} = 1)$, which makes the probit model computational expensive to estimate. More recently, the mixed logit model (MXL) has been introduced and developed thanks to the work of McFadden (1984), McFadden and Train (2000) and Ben-Akiva et al. (2001), Walker and Ben-Akiva (2002). This model is based on the MNL model, considering the coefficients $\boldsymbol{\beta}$ as random variables, allowing for the utilities to be correlated. Furthermore, the MXL model is a general model, as it can approximate any discrete choice model (Train, 2003). The models used in this work are of this last type.

3.2.1 Multinomial logit model

The simplest random utility model is the multinomial logit (MNL) model. In the MNL model, the disturbances ε_{ij} in Eq. (3.1) are assumed independent and identically distributed extreme value random variables for all $i \in I$ and $j \in J_i$. The extreme value distribution is also known as the Gumbel distribution, in honor of the mathematician and political activist Emil Julius Gumbel, who established the extreme value theory (Gumbel, 1954) that deals with the probability of extreme events, such as the river floods. The shape of the Gumbel distribution has no particular behavioural interpretation in the context of discrete choice models. However, its mathematical properties relate the logit formula, that possess desirable behavioural features and can easily infer the probabilities from observed choices (Luce, 1959), to the random utility maximization hypothesis (Marschak, 1959). To be more specific, the choice probabilities follow the logit formula if and only if the random disturbances of a random utility model are i.i.d. Gumbel random variables. The implication is credited to Marschak (1959), while the converse to Mcfadden (1972), who also proposed a model in which the utilities of alternatives depended on their measured attributes. His works in discrete choice, awarded him the Nobel Prize in Economics in the year 2000.

To derive the probability $Pr(y_{ij} = 1)$, first note that, conditioning on ε_{ij}, the expression in Eq. (3.2), can be written as

$$
\begin{aligned}
Pr(y_{ij} = 1) &= \int Pr(\varepsilon_{ik} < V_{ij} - V_{ik} + \varepsilon_{ij} \;\forall\; k \neq j \,|\, \varepsilon_{ij} = u) Pr(\varepsilon_{ij} = u) du \\
&= \int Pr(\varepsilon_{ik} < V_{ij} - V_{ik} + u \;\forall\; k \neq j) Pr(\varepsilon_{ij} = u) du \\
&= \int \prod_{k \neq j} Pr(\varepsilon_{ik} < V_{ij} - V_{ik} + u) Pr(\varepsilon_{ij} = u) du \,,
\end{aligned}
\tag{3.3}
$$

where the last equality is because the random variables ε_{ik} are assumed independent. The density and cumulative probability functions of a Gumbel random variable, ε, are, respectively,

$$f_\varepsilon(u) = e^{-u}e^{-e^{-u}}$$
$$F_\varepsilon(u) = e^{-e^{(-u)}} .$$

By substituting the density and cumulative probability functions in Eq. (3.3), and factorising for $e^{-e^{-u}}$,

$$\begin{aligned} Pr(y_{ij}=1) &= \int e^{-u} \prod_{k\neq j} e^{-e^{-u}[e^{-(V_{ij}-V_{ik})}+1]} du \\ &= \int e^{-u} e^{-e^{-u}\sum_{k\neq j}[e^{-(V_{ij}-V_{ik})}+1]} du \\ &= \frac{1}{\sum_{k\neq j}[e^{-(V_{ij}-V_{ik})}+1]} . \end{aligned} \tag{3.4}$$

The last inequality is obtained by noting that $\frac{d}{du}\left[e^{-\alpha e^{-u}}\right] = \alpha e^{-u}e^{-\alpha e^{-u}}$, and integrating for all the values of u. Finally, by multiplying and dividing Eq. (3.4) by e^{Vij}, the standard expression for the probabilities in the MNL model are obtained,

$$Pr(y_{ij}=1) = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} . \tag{3.5}$$

The simplicity of the MNL model, resulting from the closed form of the choice probabilities (Eq. (3.5)), makes it the most used discrete choice model, and the starting point in the development of more complex models. Moreover, the log-likelihood of the MNL model is globally concave for linear-in-parameters utility, and thus it is computationally efficient to solve for the parameters $\boldsymbol{\beta}$. Moreover, in spite of its simplicity, the MNL model allows to represent systematic taste variations between individuals, and the dynamics of repeated choice through a correct specification of the functions V_{ij}. However, the MNL model exhibits the independence of irrelevant alternatives (IIA) property, which may not be suitable in some choice situations. Furthermore, as a result of the i.i.d. assumption of the unobserved factors ε_{ij}, the MNL model is not able to represent random taste variation of individuals, nor their correlation in repeated choice situations. These limitations, and their implications in route choice context, are elaborated in the next sections.

3.2.1.1 Independence from irrelevant alternatives (IIA)

This property states that the preference between two alternatives j and k is independent from the rest of the alternatives in the choice set. This property can be easily observed for the MNL model by taking the ratio

$$\frac{Pr(y_{ij}=1)}{Pr(y_{ik}=1)} = \frac{e^{V_{ij}}}{e^{V_{ik}}} . \tag{3.6}$$

As the proportion in this expression does not depend on the rest of the alternatives, the IIA property implies that the introduction of a new alternative k' in the choice set does not change the relative preference between j and k. Moreover, changes in the attributes of the alternative k' do not affect the preferences either. In other words, the alternative k' is *irrelevant* for the preference between two other alternatives.

The IIA property has undesirable consequences in the context of route choice. To illustrate this, consider the problem of route choice between two different routes j and k, k being a route crossing through the inner city, and j being a freeway in the peripheral with $Pr(y_j = 1) = 2/3$ and $/Pr(y_k = 1) = 1/3$, i.e., $Pr(y_j = 1)/Pr(y_k = 1) = 2$. If a third route, k', is now included as an alternative, such that its attributes are very similar to the alternative k, then $Pr(y_k = 1)/Pr(y_{k'} = 1) \approx 1$. However, since the probabilities of the three routes must sum one, now $Pr(y_j = 1) \approx 1/2$ and $Pr(y_k = 1) \approx Pr(y_{k'} = 1) \approx 1/4$. This situation is unrealistic, as the preference for the peripheral route k would be expected to remain unchanged, while the preference for the inner city route split into the two alternatives. In fact, the probability of choice of the two first alternatives decreases by the same proportion when k' is available. A common situation in which this occurs is when routes are highly overlapping, this is illustrated in Fig (3.1).

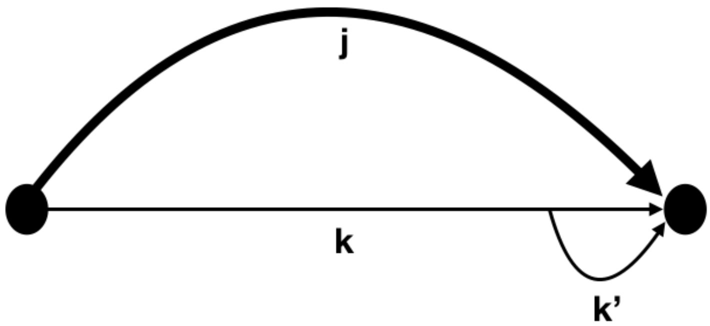

Figure 3.1: Independence of irrelevant alternatives. The inclusion of a new route k', similar to one of the routes already present in the choice set, will cause the same proportional decrease in the choice probabilities of j and k.

Extensions to the MNL model have been proposed to overcome the problems that the IIA may cause in some situations. The nested logit, and then the generalised extreme value model (Ben-Akiva, 1973, Williams, 1977, Mcfadden, 1978, Vovsha, 1997, Bekhor, 2016) is a relaxation of the MNL model that divides the choice set into a hierarchy of *nests*, such that the alternatives are correlated within the nests, but uncorrelated between nests. This implies that the IIA property holds for the alternatives in a nest, but not for alternatives in different nests. The C-logit (Cascetta et al., 1996) and the path-size logit (Ben-Akiva and Bierlaire, 1999, Ramming, 2002) were proposed to account for the overlapping of alternative routes by specifying the amount of overlapping in the systematic part of the utility of a MNL model. An approach to capture the correlation between roads in large networks is proposed by Frejinger and Bierlaire (2007), Frejinger (2008) and Bierlaire and Frejinger (2008).

3.2.1.2 Taste variation between individuals

In MNL models, the taste of the decision-makers with respect to the different attributes are represented by the coefficients $\boldsymbol{\beta}$, which are fixed for all decision-makers i. Nevertheless, this assumption may not be adequate in situations where variations in the tastes of different individuals is suspected. To exemplify this, consider the route choice problem in which the travel time in route j, TT_j, is part of the explanatory variables. The utility that individual i gets from the alternative route j is specified as

$$U_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta} + \beta_{TT} TT_j + \varepsilon_{ij} \, . \tag{3.7}$$

The coefficient β_{TT} is negative as higher travel times reduce the utility that a traveller gets from an alternative. In this case, the taste for the travel time β_{TT} is fixed and equal for all decision-makers. However, this may not be true, as the taste for the travel time depends on the value of time of each individual: the disutility caused by the travel time in a route is higher for travellers with higher value of time. Denote the value of time of individual i as VOT_i. The taste variation can be then represented through the systematic part of the utility, i.e.,

$$\begin{aligned} U_{ij} &= \mathbf{x_{ij}}^T \boldsymbol{\beta} + \beta_{i,TT} TT_j + \varepsilon_{ij} \\ \beta_{i,TT} &= \gamma VOT_i \, , \end{aligned}$$

where γ is a negative coefficient. In this last expression the value of $\beta_{i,TT}$ varies as a linear function of the value of time. Therefore, the above model can be written as

$$U_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta} + \gamma (VOT_i TT_j) + \varepsilon_{ij} \, . \tag{3.8}$$

Note that in this last expression the unobserved part of the utility is given by the disturbances ε_{ij}, which are i.i.d. Gumbel random variables, and thus the model is a MNL model.

In Eq. (3.8) the variable taste variation is specified as part of the systematic part of the utility through its relationship with another variable, VOT_i. However, what happens if the value of time of the travellers is unknown? In this case, it makes sense to consider the coefficients $\beta_{i,TT}$ as random variables. Let β_{TT} be the mean of the distribution of the coefficients $\beta_{i,TT}$. Then, $\beta_{i,TT} = \beta_{TT} + \upsilon_i$, where the values υ_i are i.i.d. with $\mathbb{E}[\upsilon_i] = 0$. With this assumption, the utilities are specified as

$$\begin{aligned} U_{ij} &= \mathbf{x_{ij}}^T \boldsymbol{\beta} + \beta_{i,TT} TT_j + \varepsilon_{ij} \\ &= \mathbf{x_{ij}}^T \boldsymbol{\beta} + \beta_{TT} TT_j + \left(\upsilon_i TT_j + \varepsilon_{ij}\right). \end{aligned} \tag{3.9}$$

In this last expression, the unobserved part of the utility is $\varepsilon^*_{ij} = \upsilon_i TT_j + \varepsilon_{ij}$. Note that for the decision-maker i and two alternatives j and k, $cov(\varepsilon^*_{ij}, \varepsilon^*_{ik}) = TT_j \cdot TT_k var(\upsilon_i)$. Thus, the i.i.d. assumption of the unobserved part of the utility is violated, and the model is no longer a MNL model. In the context of this work, no data from the attributes of the individuals that participated in route choice experiment was collected, thus making it more adequate to treat the variation on the preferences of the individuals as random variables.

3.2.1.3 Panel data

Panel data or repeated choice is when the choices of the same individuals are observed in several choice situations. In this case, the utility that decision-maker i gets from an alternative j in choice situation s, with $s = 1, \ldots, S_i$, is given by

$$U_{ijs} = V(\mathbf{x_{ijs}}; \boldsymbol{\beta}) + \varepsilon_{ijs} \,,$$

where $\boldsymbol{\beta}$ is fixed for all decision-makers, and the disturbances ε_{ijs} are i.i.d. Gumbel random variables. These assumptions are necessary to guarantee that the model is a MNL model. Nonetheless, the utilities that a decision-maker gets from the same alternative in two different choice situations s and s' are not correlated, and therefore, neither are the choice probabilities. This may be unrealistic, as choices from the same individuals are expected to be correlated to some extend.

As in the case for heterogeneous taste, treated in the previous section, some aspects of the repeated choices of individuals can be captured through the specification of the systematic part of the utility. When dynamics in the choices are taken into account, it is possible, for example, to introduce lagged independent variables $x_{ij(s-1)}$, or lagged dependent variables $y_{ij(s-1)}$ to account for the past attributes or past choices of the decision-makers. However, introducing variables or functions of variables in the systematic part of the utility may be difficult when the temporal aspect of the choices is not relevant, and when the choice sets in the different choice situations are not the same. This is the case in route choice, when the observations are the choices of the same individuals on different OD pairs. In this problem, it would be difficult to explicitly specify a relationship between the choices in the different OD pairs as part of the systematic part of the utility.

3.2.2 Mixed logit model

In order to overcome the limitations of the MNL model, it has been extended to the more flexible mixed logit model (MXL), developed after the works of McFadden (1984), McFadden and Train (2000) and Walker and Ben-Akiva (2002). In the MXL model, the coefficients $\boldsymbol{\beta}$ are considered random variables, while maintaining the independent and identically Gumbel distributed assumption for the unobserved idiosyncratic errors ε_{ij}. Therefore, conditioning on $\boldsymbol{\beta}$, the probabilities have the same closed form as in Eq. (3.5),

$$Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta}) = \frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta})}}{\sum_{k=1}^{J} e^{V(\mathbf{x_{ik}};\boldsymbol{\beta})}} \,. \tag{3.10}$$

Since $Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta})$ is a function of the random vector $\boldsymbol{\beta}$, the conditional probability in this last expression is a random variable. Thus, to obtain the (unconditional) choice probability, the expression in Eq. (3.10) has to be integrated over all the possible values of $\boldsymbol{\beta}$, i.e.,

$$\begin{aligned} Pr(y_{ij} = 1) &= \int_{\Omega(\boldsymbol{\beta})} Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta} = \mathbf{u}) Pr(\boldsymbol{\beta} = \mathbf{u}) d\mathbf{u} \\ &= \int_{\Omega(\boldsymbol{\beta})} \frac{e^{V(\mathbf{x_{ij}};\mathbf{u})}}{\sum_{k=1}^{J} e^{V(\mathbf{x_{ik}};\mathbf{u})}} f_{\boldsymbol{\beta}}(\mathbf{u};\boldsymbol{\theta}) d\mathbf{u} \,, \end{aligned} \tag{3.11}$$

where $f_{\boldsymbol{\beta}}(\cdot;\boldsymbol{\theta})$ is the probability density function of $\boldsymbol{\beta}$ parametrised by $\boldsymbol{\theta}$. This last expression is the general form of the MXL models.

Independence from irrelevant alternatives (IIA). Note that in the case of the MXL model, the ratio between the choice probabilities in Eq. (3.6) depends on the attributes of all the alternatives, and thus the IIA property does not hold. Moreover, the MXL model permits to specify the change in the choice probabilities as a result of a new alternative being included in the choice set, or as a result of the change of the attributes of an irrelevant alternative. This is done trough the correlation structure of the random coefficients $\boldsymbol{\beta}$. In contrast to the MNL model, in the MXL model the utilities that an individual i gets from two alternatives, j and k, are correlated. If the systematic part of the utility is assumed to be linear in the coefficients, i.e., $V_{ij} = \mathbf{x_{ij}}^T\boldsymbol{\beta}$, then the correlation is given by $Cov(U_{ij}, U_{ik}) = \mathbf{x_{ij}}^T \Sigma_{\boldsymbol{\beta}} \mathbf{x_{ik}}$, where $\Sigma_{\boldsymbol{\beta}}$ is the covariance matrix of the random coefficients $\boldsymbol{\beta}$.

Taste variation. Since in MXL model the coefficients $\boldsymbol{\beta}$ are considered random variables, they can capture the taste variation between individuals. To make this fact explicit, consider $\boldsymbol{\beta}_\mathbf{i}$ the vector of coefficients associated to individual i, then, the utility is written as

$$U_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta}_\mathbf{i} + \varepsilon_{ij} \,.$$

If the coefficients $\boldsymbol{\beta}_\mathbf{i}$ were observed by the bystander, then the probability that individual i chooses alternative j is given by the conditional probability $Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta}_\mathbf{i})$, which is a MNL model, as it can be seen in Eq. (3.10). However, the coefficients $\boldsymbol{\beta}_\mathbf{i}$ are not known, and thus they need to be integrated out to obtain the unconditional probability

$$Pr(y_{ij} = 1) = \int_{\Omega(\beta)} Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta} = \boldsymbol{\beta}_\mathbf{i}) Pr(\boldsymbol{\beta} = \boldsymbol{\beta}_\mathbf{i}) d\boldsymbol{\beta}_\mathbf{i} \,, \tag{3.12}$$

which is equivalent to the formulation of the MXL model in Eq. (3.11).

If the coefficients $\boldsymbol{\beta}_\mathbf{i}$ are assumed to be identically distributed between individuals, then the probability distribution $f_\beta(\cdot\,;\boldsymbol{\theta})$ can be interpreted as the distribution of the tastes or preferences in the population. In this sense, a larger variance of $f_\beta(\cdot\,;\boldsymbol{\theta})$ means that the tastes of the population are highly heterogeneous. However, in some cases the assumption that the coefficients are identically distributed may be restrictive. In this case, more complex representations of the utility can be obtained by letting the parameters of the distribution $f_\beta(\cdot\,;\boldsymbol{\theta})$ vary according to the observed values on the individuals, i.e., $\boldsymbol{\theta_i} = g(\mathbf{z}_i)$, where $\mathbf{z}_i$ is a vector of observed values on the individual i. To illustrate this, remember the example in section 3.2.1.2, but now considering the random coefficient $\beta_{i,TT} = \beta_{TT} VOT_i + \sqrt{1/VOT_i}\upsilon_i$. The mean and variance of $\beta_{i,TT}$ are then $\mathbb{E}[\beta_{i,TT}] = \beta_{TT} VOT_i$ and $Var(\beta_{i,TT}) = (1/VOT_i)\sigma_\upsilon^2$. In this specification, the dissutility caused by higher travel times is larger for travellers with higher value of time (as long as $\beta_{TT} < 0$), but also the variance diminishes, capturing the fact that travellers with higher value of time are less heterogeneous.

Panel data. In the MXL models, when the decisions of individuals are observed in several choice situations, the utility is written as

$$U_{ijs} = \mathbf{x_{ijs}}^T \boldsymbol{\beta}_\mathbf{i} + \varepsilon_{ijs} \,,$$

where $s = 1, \ldots, S_i$ indexes the choice situation in which the observation is made; the disturbances ε_{ijs} are i.i.d. Gumbel random variables; and $\boldsymbol{\beta}_\mathbf{i}$ is a vector of random coefficients. In its simplest form, the coefficients are allowed to vary over individuals, but to be constant over choice situations, i.e., $\boldsymbol{\beta}_\mathbf{i}$. In other words, the tastes or preferences of individuals vary between individuals but they remain constant for the same individual in different choice situations. Under this hypothesis, the choices of a same individual are correlated, as $Cov(U_{ijs}, U_{iks'}) = \mathbf{x_{ijs}}^T \Sigma \boldsymbol{x_{ijs'}}$.

To obtain the expression of the choice probability, let j_s represent an alternative in choice situation s. Note that, since the random errors ε_{ijs} are i.i.d., the conditional probabilities $Pr(y_{ij_s s} = 1 \,|\, \boldsymbol{\beta} = \boldsymbol{\beta}_\mathbf{i})$ are independent for $s = 1, \ldots, S_i$. Therefore, the joint conditional probability of observing the sequence of choices $j_1, j_2, \ldots, j_{S_i}$ from individual i is given by

$$Pr(y_{ij_s s} = 1,\ s = 1, \ldots, S_i \,|\, \boldsymbol{\beta} = \boldsymbol{\beta_\mathrm{i}}) = \prod_{s=1}^{S_i} Pr(y_{ij_s s} = 1 \,|\, \boldsymbol{\beta} = \boldsymbol{\beta_\mathrm{i}})\,,$$

and the joint unconditional probability is obtained integrating over $\boldsymbol{\beta_\mathrm{i}}$,

$$Pr(y_{ij_s s} = 1,\ s = 1, \ldots, S_i) = \int_{\Omega(\beta)} \prod_{s=1}^{S_i} \frac{e^{V(\boldsymbol{x_{ij_s s}};\boldsymbol{\beta}_\mathrm{i})}}{\sum_{k=1}^{J} e^{V(\boldsymbol{x_{iks}};\boldsymbol{\beta}_\mathrm{i})}} f_\beta(\boldsymbol{\beta_\mathrm{i}};\boldsymbol{\theta}) d\boldsymbol{\beta_\mathrm{i}}\,.$$

3.2.3 Joint random utility models

Joint RUM models arise in situations in which decisions of the same individuals are observed in several related choice problems, and correlation among their decisions is suspected. This is the case of surveys, where the answers of individuals to different questions may be correlated; or in route choice, with decisions of travellers in different OD pairs. Joint estimation of RUM models is not new in literature. It has its origins in the work of Ashford and Sowden (1970) with the introduction of the multivariate probit model (MPM); a generalisation of the probit model that permits the joint estimation of several binary response regressions, accounting for the correlations that may exist between them. In their work, though not related to choice behaviour modelling, Ashford and Sowden (1970) estimate a MPM to simultaneously model the binary response of various physiological systems to a given stimuli. The problem that the authors address with the MPM is the relation that may exist between the different systems; relation that would be neglected if the responses were estimated independently. Regarding the modelling of choices, the MPM has been used in situations when the decision-makers can choose more than one alternative in a choice set, but the choices are suspected to be related. Golob and Regan (2002) estimate a MPM to understand the adoption of information technologies in the trucking industry, where the alternative technologies could be competing or complementary, and thus present correlation patterns. In the transportation context, Srinivasan and Mahmassani (2000) used a probit model jointly estimated for two groups of participants to study the inertia and compliance to travel time information. In these studies, the problem pertains the modelling of binary response subproblems that are correlated. Thus, is reasonable the choice of a MPM as a modelling approach, as the correlations can be captured by the error structure of the probit model. However, in situations where the responses are discrete instead of binary, the correlations within and between subproblems need to be estimated, problem that can easily grow intractable with the number of alternatives in the choice sets.

A special case in joint RUM models is when the choice problems share part of their variables. In this situation, the coefficients of the shared variables in the model can be assumed to be equal across the choice problems. This problem is encountered when combining revealed preference (RP) and stated preference (SP) data into a single model, with the objective to exploit the advantages of each type of data, while mitigating their weaknesses. The techniques to jointly estimate models combining SP and RP data were first developed by Ben-Akiva and Morikawa (1990), who used them to estimate a MNL model to study the transportation mode switching of travellers. Other related works can be found in Bradley and Daly (1991), Adamowicz et al. (1998, 1994), Hensher and Bradley (1993), Earnhart (2002). When combining SP and RP data, two MNL models are estimated, one for RP and the other for SP data,

$$\begin{aligned} U_{ij}^{RP} &= \mathbf{x_{ij}}^T \boldsymbol{\beta} + \boldsymbol{z}_{\boldsymbol{ij}}^T \boldsymbol{\gamma} + \varepsilon_{ij}^{RP} \\ U_{ij}^{SP} &= \mathbf{x_{ij}}^T \boldsymbol{\beta} + \boldsymbol{w}_{\boldsymbol{ij}}^T \boldsymbol{\eta} + \varepsilon_{ij}^{SP} \,. \end{aligned} \tag{3.13}$$

Since some of the dependent variables are measured in both SP and RP data, their coefficients are set to be equal in the joint model. Thus, the available SP data participates in the estimation of the coefficients of the RP model, and vice versa. In this case, special care has to be taken with differences in the scale parameters (variance of the unobserved part of the utility ε_{ij}^{RP} and ε_{ij}^{SP}) associated to the two data collection techniques; consideration that can be neglected when data comes from one source. Note that when the joint model shares all of its variables ($\boldsymbol{z}_{\boldsymbol{ij}} = \boldsymbol{w}_{\boldsymbol{ij}}$), then it consists of one representation of the utility, given by the variables and their respective coefficients. The joint estimation of MNL models has been extended to the more flexible MXL models (McFadden and Train, 2000, Bhat and Castelar, 2002, Brownstone et al., 2000). Hence, enabling for taste variation between individuals and panel data structures. The only difference with the model in Eq. (3.13) is that now the coefficients are random and they are indexed by individual, i.e., $\boldsymbol{\beta}_{\mathbf{i}}$.

In path-based route choice, each OD pair can be considered to be an independent problem, as they do not share the alternatives between them: one route is only possible from one origin to one destination. This implies that there is one route choice model for each OD pair *od*. If the OD pairs are described by the same attributes, the coefficients can be assumed to be equal for all the OD pairs, resulting in one representation of the utility for each *od*, i.e.,

$$U_{ijs}^{od} = (\boldsymbol{x}_{\boldsymbol{ijs}}^{\boldsymbol{od}})^T \boldsymbol{\beta}_{\boldsymbol{i}}^{\boldsymbol{od}} + \varepsilon_{ijs}^{od} \,, \tag{3.14}$$

where the tastes vary across individuals and OD pairs. However, it makes sense to assume that the tastes of an individual i are the same regardless of the OD pair, i.e., $\boldsymbol{\beta}_{\boldsymbol{i}}^{\boldsymbol{od}} = \boldsymbol{\beta}_{\mathbf{i}}$. Moreover, if the data comes from the same experimentation technique, there is no reason to consider the scale parameter to be different between OD pairs. Therefore, the joint model in Eq. (3.14) can be treated as the panel data MXL model, given by

$$U_{ijs} = \mathbf{x_{ijs}}^T \boldsymbol{\beta}_\mathbf{i} + \varepsilon_{ijs} \,, \tag{3.15}$$

where each choice situation s represents a choice situation t in a particular OD pair $s = (od, t)$, and each alternative j belongs to the choice set of the OD pair od.

Model in Eq. (3.15) is the model used in this work. It is justified because participants in the route choice experiments make choices over different OD pairs. The assumptions made with this model are:

- tastes or preferences of individuals vary in the population;
- preferences towards the route attributes are the same, regardless of the OD pair and choice situation;
- the scale parameter is constant throughout the OD pairs.

3.2.4 Estimation of random utility models

The problem of estimating a random utility model consists in finding the values for the coefficients $\boldsymbol{\beta}$ that *better* explain the data. Given the observed data $\{y_{ij}, \mathbf{x_{ij}} \,|\, \forall j\}$ in a single choice situation and the unknown coefficients $\boldsymbol{\beta}$, the likelihood is given by

$$\mathcal{L}(\boldsymbol{\beta} \,|\, \mathcal{D}) = \prod_i \prod_j [Pr(y_{ij} = 1 | \mathbf{x_{ij}}; \boldsymbol{\beta})]^{y_{ij}} \,, \tag{3.16}$$

where $\mathcal{D} = \{y_{ij}, \mathbf{x_{ij}} \,|\, \forall i, j\}$ is the set of observed choices and attributes. The standard method to find an estimator $\hat{\boldsymbol{\beta}}$ is by maximum likelihood estimation, which can be done analytically when the likelihood has a closed form, such as in the MNL model. However, this is not the case for the MXL model, where the coefficients $\boldsymbol{\beta}$ are random, and thus need to be integrated out.

For the case of the MXL model the parameters that are estimated are $\boldsymbol{\theta}$, i.e., the parameters defining the shape of the distribution of the coefficients $\boldsymbol{\beta}$. The likelihood function for the panel data MXL model is given by

$$\mathcal{L}(\boldsymbol{\theta} \,|\, \mathcal{D}) = \prod_{\forall i} \int_{\Omega(\beta)} \prod_{s=1}^{S_i} \frac{e^{V(\boldsymbol{x_{ijs s}}; \boldsymbol{\beta}_\mathbf{i})}}{\sum_{k=1}^{J} e^{V(\boldsymbol{x_{iks}}; \boldsymbol{\beta}_\mathbf{i})}} f_\beta(\boldsymbol{\beta}_\mathbf{i}; \boldsymbol{\theta}) d\boldsymbol{\beta}_\mathbf{i} \,, \tag{3.17}$$

where now $\mathcal{D} = \{y_{ijs}, \mathbf{x_{ijs}} \,|\, \forall i, j, s\}$ is the set of observed choices and attributes. Obtaining the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ requires to solve the multiple integral in Eq. (3.17). Nevertheless, since the integral has no closed form, it needs to be numerically approximated, which could present convergence problems and be computationally expensive. An alternative approach is to regard the MXL model as a Bayesian hierarchical model, which has the advantage of avoiding the numerical multiple integration in Eq. (3.17) (Train, 2001, Huber and Train, 2001, Train, 2003, Regier et al., 2009, Balcombe et al., 2009).

Mixed logit as a hierarchical Bayesian model The mixed logit can be framed as a hierarchical Bayesian model (Train, 2001, 2003). Denote by $\boldsymbol{y_{is}}$ a vector representing the observed choice of individual i in choice situation s. The elements of $\boldsymbol{y_{is}}$ take the value $y_{isj} = 1$ when alternative j was chosen, and $y_{isj} = 0$ otherwise; as there is one and only one chosen alternative in each situation s, $\sum_j y_{isj} = 1$. For ease of exposition, the vectors $\boldsymbol{y_{is}}$ will be assumed to be of length J for all i and s, though it may not be necessarily the case. The vector $\boldsymbol{y_{is}}$ can be regarded as a realization of a categorical distribution (the multidimensional generalization of the Bernoulli distribution) with parameter $\boldsymbol{p_{is}}$, a vector which elements p_{isj}, $j = 1, \ldots, J$ are the probabilities of choosing each of the alternatives. Naturally, $\sum_j p_{isj} = 1$. The observed attributes of the individuals and the alternatives are represented in the vector, $\boldsymbol{x_{ijs}}$, and the probabilities of choosing the alternatives depend on these attributes and the coefficients $\boldsymbol{\beta}_\mathbf{i}$. The coefficients $\boldsymbol{\beta}_\mathbf{i}$ are random variables with probability density function $f_{\beta}(\cdot \,|\, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the parameters of the distribution. The hierarchical Bayesian model can be written, for each individual i, as

$$\begin{aligned} \boldsymbol{y_{is}}|\boldsymbol{\beta}_\mathbf{i} &\sim Cat(\boldsymbol{p_{is}}(\boldsymbol{\beta}_i)) \quad \forall s \\ p_{isj} &= p(y_{isj} = 1 \,|\, x_{isj}, \boldsymbol{\beta}_\mathbf{i}) \quad \forall j, s \\ \boldsymbol{\beta}_\mathbf{i} &\sim f_{\beta}(\cdot\,;\, \boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim f_{\boldsymbol{\theta}}(\cdot\,;\, \boldsymbol{\rho}). \end{aligned} \tag{3.18}$$

Note that the parameter $\boldsymbol{\theta}$ is treated as a random variable. In Bayesian procedures, the unknown parameters are treated as random variables to make explicit the uncertainty about its true value, its distribution, $f_{\boldsymbol{\theta}}(\cdot\,;\, \boldsymbol{\rho})$, is called the prior.

In words, the model in Eq. (3.18) is explained as follows. Suppose that the values of the parameters $\boldsymbol{\theta}$ are known. These parameters are referred to as population parameters, as they determine the shape of the distribution of the tastes of the individuals in a population, $f_{\beta}(\cdot\,;\, \boldsymbol{\theta})$. For each individual i, a coefficient $\boldsymbol{\beta}_\mathbf{i}$ is drawn from this distribution. Then, the probability of choice for each alternative j in each choice situation s is computed conditional on $\boldsymbol{\beta}_\mathbf{i}$. Observe that in the MXL model the conditional probability is given by the MNL model formula. Finally, for each situation s a vector of choices $\boldsymbol{y_{is}}$ is drawn from the categorical distribution with parameters given by the probabilities $\boldsymbol{p_{is}}$. This is a natural way of framing the MXL model, and it is given by the joint probability of the unknown parameters $\boldsymbol{\theta}$, the unknown coefficients $\boldsymbol{\beta}_\mathbf{i}$ for all i, and the observations $\mathcal{D} = \{\boldsymbol{y_{is}}, \boldsymbol{x_{ijs}} \,|\, \forall i, j, s\}$,

$$Pr(\mathcal{D}, \boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\}) = Pr(\mathcal{D} \,|\, \boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\}) Pr(\{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \boldsymbol{\theta}) Pr(\boldsymbol{\theta}) \,. \tag{3.19}$$

Bayesian estimation Inference in the Bayesian context refers to obtaining the joint distribution of the unobserved parameters that best fits the data (Barber, 2011). Since the coefficients $\boldsymbol{\beta}_\mathbf{i}$ in the case of the MXL model are not observed (they are latent variables), they can be treated as parameters of the full model in Eq. (3.19). Therefore, by the Bayes' theorem,

$$Pr(\boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \mathcal{D}) \propto Pr(\mathcal{D} \,|\, \boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\}) Pr(\{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \boldsymbol{\theta}) Pr(\boldsymbol{\theta}) \,. \tag{3.20}$$

To estimate the joint distribution of the parameters, first, the *prior* distribution, representing the researchers' beliefs over the values of the parameters $\boldsymbol{\theta}$, is defined. The prior is $Pr(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(\boldsymbol{\theta}; \boldsymbol{\rho})$, and when no information about prior is available, the values of the hyperparameters, $\boldsymbol{\rho}$, can be chosen to be weakly-informative (high variances) to reflect the uncertainty over the *real* value of $\boldsymbol{\theta}$. Then, the prior is updated through the likelihood function to obtain the *posterior* distribution of the unknown parameters $Pr(\boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \mathcal{D})$. The likelihood function is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \mathcal{D}) &= Pr(\mathcal{D} \,|\, \boldsymbol{\theta}, \{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\}) Pr(\{\boldsymbol{\beta}_\mathbf{i} \,|\, \forall i\} \,|\, \boldsymbol{\theta}) \\ &= \prod_{\forall i} \prod_{s=1}^{S_i} \frac{e^{V(\boldsymbol{x_{ij_ss}}; \boldsymbol{\beta}_\mathbf{i})}}{\sum_{k=1}^{J} e^{V(\boldsymbol{x_{iks}}; \boldsymbol{\beta}_\mathbf{i})}} f_{\beta}(\boldsymbol{\beta}_\mathbf{i} \,|\, \boldsymbol{\theta}) \,. \end{aligned} \tag{3.21}$$

Note that by considering the coefficients $\boldsymbol{\beta}_\mathbf{i}$ as parameters in a hierarchical Bayesian model, the model is simplified as there is no longer need to solve the multiple integral in Eq. (3.17). Still, the joint posterior distribution in Eq. (3.20) has no closed form. However, samples from the joint posterior distribution can be obtained using the Gibbs sampling method (Levin and Peres, 2017).

3.2.5 Choice models in this thesis

The MXL model is the model chosen in this thesis to analyse the responses of participants in the MDG experiments. Even though discrete choice models have been criticised due to their utility maximisation assumption (Simon, 1957, Kahneman and Tversky, 1979), which has been found to be violated in some situations, they provide a parsimonious yet effective framework to analyse and predict the choices of individuals. The usefulness of the discrete choice models comes from their simplicity and their ability to explain and predict the choices of decision-makers within an acceptable error. The use of the MXL model in this thesis is justified below.

Discrete choice models, including the MXL model, can easily incorporate the attributes as explanatory variables of the choices. This enables to efficiently estimate the coefficients related to these variables, and their interpretation is straightforward: the tastes or presences of individuals for the different attributes. In the specific case of the MXL model, the coefficients are considered as random variables, meaning that the individuals' tastes vary in the population. This assumption is realistic, since individuals are heterogeneous. In addition, random coefficients allow errors due to unobserved factors to be assigned in a finer way (to each variable), compared to the general error in the MNL model; an important property, considering that in this thesis no data were collected on the characteristics of individuals. Moreover, the MXL model is easily adapted to situations in which the individual make repeated choices, i.e., panel data, as is the case in the MDG experiments. Another reason for the preference of the MXL model in this work is its flexibility, in the sense that it can estimate the implicit correlation between the coefficients, and it allows to explicitly specify the correlation structure in the data. The former enables to capture and interpret certain aspects of the behaviour of individuals, the latter enables to respect the structure of a phenomena. In this thesis, a structure is specified in order to correlate the utilities in a joint departure time and route choice model. A Bayesian approach is used in this thesis to estimate the MXL models. The reason is because the MXL model can be regarded as a hierarchical Bayesian model. Therefore, complex models can be straightforward specified into a Markov chain Monte Carlo algorithm to obtain samples of the posterior distributions of the parameters. In other words, there is no need to code an optimiser for a specific problem.

4 Approximating travellers' route choices at full-scale urban network level

In a city-scale network, trips are made in thousands of origin-destination (OD) pairs connected by multiple routes, resulting in a large number of alternatives with diverse characteristics that influence the route choice behaviour of the travellers. As a consequence, to accurately predict user choices at full network scale, a route choice model should be scalable to suit all possible configurations that may be encountered. In this chapter, a new methodology to obtain such a model is proposed. The main idea is to use clustering analysis to obtain a small set of representative OD pairs and routes that can be investigated in detail through computer route choice experiments to collect observations on travellers behaviour. The results are then scaled-up to all other OD pairs in the network.

It was found that 9 OD pair configurations are sufficient to represent the network of Lyon, France, composed of 96,096 OD pairs and 559,423 routes. The observations, collected over these nine representative OD pair configurations, were used to estimate three mixed logit models. The predictive accuracy of the three models was tested against the predictive accuracy of the same models (with the same specification), but estimated over randomly selected OD pair configurations. The obtained results show that the models estimated with the representative OD pairs are superior in predictive accuracy, thus suggesting the scaling-up to the entire network of the choices of the participants over the representative OD pair configurations, and validating the methodology in this study.

This chapter is published as the journal article: González Ramírez, H., Leclercq, L., Chiabaut, N., Becarie, C., Krug J. (2019). Unravelling travellers' route choice behaviour at full-scale urban network by focusing on representative OD pairs in computer experiments. PLOS ONE 14(11): e0225069. https://doi.org/10.1371/journal.pone.0225069

4.1 Motivation

Urban congestion occurs when traffic demand locally exceeds the network capacity. The local demand is the combination of the global travel demand between the different origins and destinations and the travellers' route choices, which define how many trips are made at the same place in a given time window. Thus, at a city-scale level, i.e., considering all the OD pairs and links in the network, route choice is a key determinant of urban transportation network performance. Route choice behaviour has been extensively studied in the transportation literature from two main but different angles. The first, related to human factors and mainly founded in discrete choice models (Manski and McFadden, 1981, Train, 2003, Walker and Ben-Akiva, 2002), is focused on the identification of the determinants of travellers' individual choices. This line of research is based on investigating travellers' behaviour through experiments that consist in either observing their choices in the field (revealed preference) or asking them what would be their choices in hypothetical scenarios (stated preference). The second line of research, tackles the problem at full-scale and aims to solve the network loading problem to determine static or dynamic traffic states over all the network links. In this case, the interactions between the demand and the route choices on all the OD pairs in the network are considered altogether to define general principles that determine the network equilibrium. This is, for example, the case of the deterministic network equilibrium principle (Wardrop, 1952), that states that the travellers are selfish optimisers who only try to minimise their travel costs when choosing a route amongst all the alternatives; at the equilibrium, all the used routes that connect an OD pair have the same minimal cost.

Theoretically, the study of route choice from an individual and network level are consistent. Various approaches have been proposed to join the individual behaviour and network level. This is specially investigated using the RP method, since it allows to observe the behaviour of travellers over different network facilities and circumstances (Baillon and Cominetti, 2008, Fosgerau et al., 2013, Zimmermann and Frejinger, 2019). However, RP methods lack of precise information about the travel times (or the traffic conditions) on the network when choices are made (see Section 2.1). This difficulty is one of the main reasons why SP methods have been broadly used to study and estimate the route choices of travellers. When considering SP methods, there is a lack of connection between the study of route choice at an individual and network level (Yildirimoglu and Kahraman, 2018b). The reason is that, on the one hand, SP studies of route choice behaviour are focused in specific determinants of travellers' route choice and, therefore, are based on simple scenarios (two or three routes in few OD pair configurations) that do not cover the multiplicity of situations that are found in a city-scale transportation network. In these experiments, particular attention has been paid to the study of how travellers learn from experience (Iida et al., 1992, Bogers, 2005, Selten et al., 2007), the impact of advanced travel information systems (ATIS) (Adler and McNally, 1994, Lotan, 1997, Mahmassani and Liu, 1999, Ben-Elia and Shiftan, 2010, Ben-Elia and Avineri, 2015, Abdel-Aty et al., 1997, Srinivasan and Mahmassani, 2000), and the effect of travel time variability and risk attitudes in the travellers choices (De Moraes Ramos et al., 2013, Avineri and Prashker, 2005, de Palma and Picard, 2005). On the other hand, in the network loading problem, representations have been designed as simplified mathematical abstractions that permit to calculate the network loading under different behavioural principles, such as the deterministic user equilibrium (Wardrop, 1952), stochastic user equilibrium (Sheffi, 1985) or bounded rational user equilibrium (Mahmassani and Chang, 1987). These representations often assume that the only variable influencing travellers' route choice behaviour is the travel time, ignoring other local factors, related to the network OD configuration, that have been recognised to influence route choice behaviour (Bovy and Stern, 1990, Ramming, 2002, Bekhor et al., 2006, Papinski et al., 2009, Zhu and Levinson, 2015). One of the main reasons for the gap between research in SP studies of individual route choice behaviour and network loading is the lack of observations at large scale over a sufficient number of OD configurations, that would allow discrete choice models to scale-up at the network level and thus enable the design of network equilibrium founded in a more user-oriented approach. As mentioned above, this is not the case for RP studies, that have in turn other difficulties. The ambition of this study is to fulfil this gap by the selection of OD pairs that are representative of the OD configurations that are found in a transportation network, and then use these OD pairs in computer experiments to collect data on travellers' route choice behaviour.

In a city-scale network, trips are made in thousands of OD pairs connected by several routes (in the case concerning this study, the city of Lyon in France, the network has 96,096 OD pairs and 559,423 routes), resulting in a large number of diverse routes, consequence of the topology of the network. For example, the route alternatives connecting an OD pair located in the central part of a city are likely to have short length, a high number of intersections and turns, but are unlikely to include segments of freeways. In contrast, the routes connecting an OD pair that traverses the city are longer and are more likely to include routes with fewer number of intersections and segments of freeways. From the point of view of the design of experiments, this implies that the number of scenarios must be reduced to a small but representative set of scenarios, such that the choices of travellers in any scenario found in the network can be approximated by a choice model estimated with this small set. More specifically, a representative set of OD pairs and routes is such that, for any randomly sampled OD pair in the network it is possible to find an OD pair in the representative set with similar attributes. Thus, assuming that the choices of travellers are similar for similar situations, an estimated model on the representative OD pairs could adequately reproduce the choices in the rest of the OD pairs. The question addressed in this work is: how to find a set of OD pairs and routes, such that it is representative of the OD configurations and route attributes found in the network, while being small enough so that a sufficient number of observations on route choices can be collected through computer route choice experiments?

The solution proposed in this work is based on k-means clustering (Hastie et al., 2009) of the full set of OD pairs and routes in the network. In cluster analysis, the observations, in this case OD pairs and connecting routes routes, are grouped in clusters characterised for having elements that are similar among themselves, but dissimilar to the elements in the other clusters. In the problem pertaining this chapter, the elements in a cluster will show similar orientation, length, % of freeway, directness and number of turns, and thus a cluster C_i will be, for example, composed mainly of OD pairs of short length in the central part of the city, with direct routes and low % of freeway composition, whereas another cluster C_j will be composed of OD pairs representing long trips traversing the city, with some non-direct routes composed mainly of freeways. Assuming that there are k clusters, the elements of a cluster can be regarded as belonging to a same class of OD pair configurations, and the whole network as being composed of elements of k different classes. Therefore, the OD pairs and routes in the network can be represented by elements in the k clusters. A natural choice to represent the elements in a cluster is the mean element in the cluster (cluster centroid), as it is the point with minimum Euclidean distance to all the elements in the cluster. Thus, the cluster centroids are chosen as representative of the clusters' elements, and the k clusters' centroids as representative of the OD pairs and routes in the whole network. These OD pairs and routes are then used in computer experiments to collect data on travellers route choice behaviour. Note that the set of representative OD pairs and routes found with k-means is a sample of the attributes of the network, so the question that arises here is if a model estimated over this representative set can adequately reproduce the choices in the rest of the OD pairs in the network. To answer this question, three discrete choice models are estimated with the observations over the representative set. The discrete choice model used in this work is a joint mixed logit model (MXL), which under certain conditions, as is the case in this study, is equivalent to the panel data formulation of MXL models (Train, 2003, McFadden and Train, 2000, Bhat and Castelar, 2002, Brownstone et al., 2000). The predictive accuracy of these models is compared with the predictive accuracy of the same models, but estimated with randomly chosen sets of OD pair configurations in a sort of cross validation procedure.

The results of the above methodology are that the models estimated with the observations over the representative OD pair configurations are better in predicting the route choices on *unseen* OD pairs, i.e., on OD pairs not used for the estimation process. On the one hand, these results demonstrate how a careful selection of OD pairs for experiments on route choice behaviour can improve the results of a choice model in a broader set of OD pairs and, on the other hand, that cluster analysis can be used to find these OD pair configurations. These findings have direct implications for urban traffic simulators, which solve the network loading problem to determine the time-evolving traffic states in the network. The scalable route choice model proposed in this paper can be implemented in such simulators without adding significantly computational complexity, compared to the usual simple equilibrium rule, e.g., user equilibrium. Furthermore, the use of clustering techniques to find the most relevant OD pairs and routes in the network, provides an efficient method to calibrate route choice models that can be easily replicated in any urban transportation network.

4.2 Materials and Methods

4.2.1 Obtaining representative OD pairs and routes

The road network used in this study is the Lyon-full network, described in Section 2.2.1. The network has 285 zones, 29 entry points and 28 exit points, giving a total number of 96,096 OD pairs (see Fig 2.2). The total number of routes in the network is 559,423, with an average number of 5.82 routes per OD pair. The selected route features in this study are the informed travel time, the length, directness, number of turns per kilometre and the percentage of freeway in the route composition. These features were selected as they are variables relevant in travellers' route choice behaviour (Bovy and Stern, 1990, Ramming, 2002, Bekhor et al., 2006, Papinski et al., 2009, Zhu and Levinson, 2015), and because they are the attributes that participants can observe in the computer route choice experiments. An OD pair and three routes connecting the origin and destination, defined as *OD-routes*, are characterised by the variables describing the origin and the destination (latitude, longitude and the Euclidean distance between them), and the variables describing the three routes connecting them (the length of the route, the number of turns per kilometre, the directness of the route, and the percentage of freeway in the route). An OD-routes is then defined by 17 variables: 5 OD pair specific and 12 describing the routes (4 for each route). An OD-routes is represented as a vector in which the attributes of the three routes appear ordered by length, from shortest to longest. A depiction of the OD-routes objects is shown in Fig 4.1.

For the clustering of the OD pairs and routes, and thus the route choice experiments, the short routes (less than 1.5 km) and highly overlapping routes belonging to the same OD pair (sharing more than 70% of their links) were not considered. The reason is that very short trips lack of real alternatives: usually there is an unique route to travel from origin to destination. The highly overlapping routes are removed from the analysis because, from a route choice experiment perspective, the similarity between the routes may cause participants not to consider some routes as real alternatives and, furthermore, highly overlapping routes lack of the variability required for a choice model to capture the impact of each route attribute in the

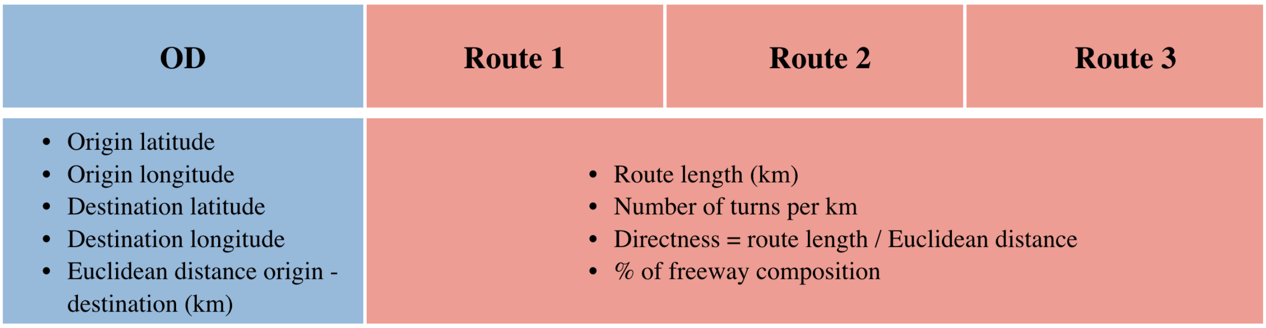

Figure 4.1: OD-routes vector. The vector is composed of the attributes of the OD pair and the three routes connecting the origin and destination, with $length(Route\_1) \leq length(Route\_2) \leq length(Route\_3)$.

choices. After removal of the very short trips and the high overlapping routes, the OD-routes are obtained by considering all the possible combinations of three routes from the set of routes joining that particular OD pair. For example, if there are 5 routes joining an OD pair, then the total number of OD-routes that are obtained is $\binom{5}{3} = 10$. The total number of OD-routes in the network is 624,490.

Before clustering, the data was normalised so that all the variables describing the OD-routes have the same weight in determining the dissimilarity between observations; this step is necessary when the range of the variables are not comparable, as is the case in the OD-routes where the directness of the routes takes values in the interval $(0, 1)$, but the length of the routes takes values in the interval $(0, 35)$. The OD-routes are clustered using k-means with Euclidean distance, determining the optimal number of clusters, k^*, using the *elbow method* (Hastie et al., 2009). The idea behind the elbow method is to select the optimal number of clusters k^*, so that the mean dissimilarity of the elements in the clusters does not decrease significantly with the $k^* + 1$ clustering. The measure of dissimilarity of the elements in a cluster is the within-cluster sum of squares (WCSS), i.e., the sum of the square distance between the elements in a cluster. One of the OD-routes among the 1% nearest to the theoretical cluster centroid is selected as the cluster centroid. This is done because the theoretical centroid, i.e., the mean of the variables of the elements in the clusters, may not be part of the data.

4.2.2 Route choice experiments

The route choice experiments for this study were carried out using the mobility decision game, described at full extent in Chapter 2. The data on route choice behaviour in this chapter comes from six route choice experiments carried out between February 2018 and February 2019 over the Lyon-full network. The participants in the experiments were students at the University of Lyon taking part to the courses of traffic theory (66%), staff from the IFSTTAR (French Institute of Science and Technology for Transport, Development and Networks) and other universities, who received an invitation by e-mail to remotely join the experiments via a web browser (34%). The participants were instructed to choose the route that they consider the *best* to complete a trip on time. Three of the six experiments were specifically implemented for the purpose of this work, so they were configured to obtain observations in the 9 representative OD pairs. The rest of the experiments were implemented for previous studies, so they were configured over 21 OD pairs different from the 9 representative set; data coming from these experiments was used to validate the methodology in this work. Throughout the six experiments, 3,334 choices of 483 participants were recorded, from which 802 choices of 73 participants were made over the nine representative OD pairs. In the experiments, the participants were confronted to several route choice problems in the different OD pairs, the task of the participants was to choose one of the three alternative routes to complete the trip before a given time.

4.2.3 Route choice model

Joint random utility maximisation (RUM) models arise in situations in which decisions of the same individuals are observed in several related choice problems, and correlation among their decisions is suspected. A special case in joint RUM models is when the choice problems share part of their variables. In this situation, the coefficients of the shared variables in the model can be assumed to be equal across the choice problems. When the choice problems share all of their variables, then the joint RUM model consists of an unique representation of the utility, given by the variables and their respective coefficients, which is equivalent to a panel data RUM (see Section 3.2.3 for more details). In this study, several models are estimated, one for each OD pair. However, as the OD pairs are described by the same variables, the coefficients can be assumed to be equal for all the OD pairs. Therefore, the utility of the joint model is reduced to a single representation, and the model can be estimated as a panel data model. The joint model for route choice, used in this study, is based on the mixed multinomial logit model (MXL) for panel data, introduced in Section 3.2.2.

Three panel data mixed logit models are estimated using the observations collected in the route choice experiments. Five variables are used in the specification of the models. Four of these variables correspond to the variables used in the selection of the representative OD-routes for the route choice experiments, which will help to test if the choices in the representative OD-routes (cluster centroids) can approximate the choices in other OD-routes. The fifth variable is the estimated travel time that the participants received during the experiments. These variables are known to influence the route choice behaviour of travellers and that can be observed by the participants in the computer route choice experiments.

Let the individuals and alternatives be indexed by i and j, respectively. Since participants were allowed to repeat decisions in the same OD pair, the choice situation, indexed by s, represents the pair (od, t), where od is the OD pair in which the decision was made and t indexes the moment of the choice. The explanatory variables considered in the model are

- LEN_j, the length (in km) of the route j;
- DIR_j, the directness of the route j, defined as the length of j divided by the Euclidean distance between origin and destination;
- TNR_j, the number of turns per kilometre in the route j;
- FRW_j, the percentage of freeway that composes the route j;
- INF_i, binary variable indicating if participant i received information; and
- ITT_{js}, the informed travel time in the route j in OD pair and moment s, the variable is normalised by OD pair by dividing the informed travel time by the free flow travel time in the fastest of the three routes. The normalization is done in order to make travel times amongst the different OD pairs comparable.

The specifications of the three models M1, M2 and M3 are

$$U_{ijs} = \beta_{i1} FRW_j + \beta_{i2} DIR_j + \beta_{i3} TNR_j + \beta_{i4} LEN_j + \beta_{i5} ITT_{js} INF_i + \varepsilon_{ijs}\,, \tag{M1}$$

$$\begin{aligned} U_{ijs} = {} & \beta_{i1} FRW_j + \beta_{i2} DIR_j + \beta_{i3} TNR_j + \beta_{i4} LEN_j + \beta_{i5} ITT_{js} INF_i \\ & + \beta_{i6} ITT_{js} LEN_j INF_i + \varepsilon_{ijs} \end{aligned} \tag{M2}$$

$$\begin{aligned} U_{ijs} = {} & \beta_{i1} FRW_j + \beta_{i2} DIR_j + \beta_{i3} TNR_j + \beta_{i4} LEN_j + \beta_{i5} ITT_{js} INF_i \\ & + \beta_{6} ITT_{js} LEN_j INF_i + \varepsilon_{ijs}\,. \end{aligned} \tag{M3}$$

In models M1 and M2, the coefficients β_{ip} for $p = 1, .., 6$ are independent and normally distributed, i.e., $\beta_{ip} \sim N(b_p, \sigma_p^2)$. In model M3, the coefficient β_6 is fixed for all individuals (not random), and the coefficients β_{ip} are correlated for $p = 1, .., 4$, i.e., $\boldsymbol{\beta}_{\mathbf{i}} \sim N_4(b, \Sigma)$, but independent from $\beta_{i5} \sim N(b_5, \sigma_5^2)$. Model M1 is the simplest MXL model considering the five variables. In models M2 and M3 the interactions between the route length and the travel time information are taken into account, allowing for the preference towards the length of the route to change depending on the informed travel time. In model M3 the correlations between the coefficients β_{ip} for $p = 1, .., 4$ are also estimated. In MXL models, the parameters that are estimated are the means and variances (covariances) of the coefficients' distributions, $\hat{b}_p$, $\hat{\sigma}$ and $\hat{\Sigma}$.

In this work, Bayesian inference is used to estimate the choice models (see Section 3.2.4). In Bayesian methods, the parameters of the model (b and Σ) are assumed to be random variables rather than fixed values. Inference, in this context, refers to obtaining the joint distribution of the parameters that best fits the data. To estimate the joint distribution of the parameters, first, a *prior* distribution, h, representing the researchers' beliefs over the values of the parameter, is defined. Then, when data becomes available, the prior is updated through the likelihood function to obtain the *posterior* distribution, H. As a result of the Bayes' theorem, the posterior distribution is proportional to the prior multiplied by the likelihood. In the general case, the posterior distribution H of the parameters is

$$H(b, \Sigma, \beta_i, \forall i|Y, X) \propto \left[\prod_{i=1}^{N}\prod_{s=1}^{S_i}\prod_{j\in\mathcal{C}(s)} Pr(y_{ijs}=1|x_{ijs};\beta_i)\phi_n(\beta_i|b,\Sigma)\right]h(b,\Sigma), \tag{4.1}$$

where X represents the alternative and individuals' attributes; Y the observed choices and ϕ_n is the multivariate normal density function of the random coefficients parametrised by b (mean) and Σ (covariance matrix). The expression in brackets is the likelihood of the observed choices and h is the joint prior distribution of the model's parameters. The joint priors, h, for the three MXL models estimated in this work, M1, M2 and M3 (see the Results section for the specification of the models), are, respectively,

$$\begin{aligned}
h(b_p, \sigma_p^2, p=1,...,5) &= \prod_{i=1}^{5}\phi(b_i|\mu_0,\sigma_0^2)f_{IG}(\sigma_i^2|r_0,\lambda_0)\\
h(b_p, \sigma_p^2, p=1,...,6) &= \prod_{i=1}^{6}\phi(b_i|\mu_0,\sigma_0^2)f_{IG}(\sigma_i^2|r_0,\lambda_0)\\
h(b_p, \Sigma, \sigma_p^2, p=5,...,6) &= \phi_4(b_{i=1,...,4}|\mu_0,\Sigma_0)f_{IW}(\Sigma|I_0,k_0)\\
&\qquad * \phi(b_5|\mu_0,\sigma_0^2)f_{IG}(\sigma_5^2|r_0,\lambda_0)\phi(\beta_6|\mu_0,\sigma_0^2)
\end{aligned}$$

where ϕ is the density function of the normal distribution, ϕ_n of the n-variate normal distribution, f_{IG} the density of the inverse-Gamma distribution and f_{IW} of the inverse-Wishart distribution. The inverse-Gamma is the conjugate prior for the variance of the normal distribution, and the inverse-Wishart its generalisation for the multivariate case.

The right hand side in Eq. (4.1) has no closed form, however samples from the joint posterior distribution H can be obtained using the Gibbs sampling method (Levin and Peres, 2017). In this study, the Gibbs sampler software JAGS(Plummer, 2003) and the R(R Core Team, 2018) package *rjags* were used to obtain 10000 samples of the posterior distribution H after a burn-in period of 20000 samples. The values of the hyperparameters μ_0, σ_0^2, r_0, λ_0, Σ_0, I_0 and k_0, which define the priors, were chosen to be weakly-informative (very high variances). In other words, it is assumed high uncertainty on the real values of the parameters that are being estimated. They are shown in Table 4.1.

Table 4.1: Hyperparameters of the prior distribution h.

Hyperparameter	Description
$\mu_0 = 0$; $\sigma_0 = 10,000$	Prior guesses of the mean and variance of the b parameter
$r_0 = 0.001$; $\lambda = 0.001$	Prior guesses of the shape and rate of the σ parameter
$\Sigma_{0ii} = 10,000$, $\Sigma_{0ij} = 0$ for $i \neq j$	Prior guess of the covariance of the b parameter
$I_{0ii} = 4$, $I_{0ij} = 0$, $i \neq j$; $k_0 = 4$	$(1/k_0)I_0$ is the prior guess of the covariance Σ

4.3 Results

4.3.1 Clustering

To determine the optimal number of clusters, k-means algorithm, with $k = 1, ..., 30$, was performed over the 624,490 OD-routes in the network. The mean within-cluster sum of squares (WCSS) is plotted against the number of clusters k in Fig 4.2. In the results, the optimal number of clusters is not clear, according to the elbow method: big improvements happen for the first values of k ($k \leq 4$); for values $5 \leq k \leq 9$ the improvement is mediocre; and for $k \geq 10$ the improvements are rather small. In terms of the purpose of this chapter, choosing a small number of k has the risk of sub-representing the OD-routes in the network and, more important, a small number of OD-routes in the route choice experiments implies that the variability in the route attributes is also small, posing a problem in estimating a route choice model (overfitting). In this sense, choosing high values of k is preferable, even if some of the clusters are similar. However, the needed number of observations in the route choice experiment increases with the number of OD-routes, implying higher costs in the organisation of the experiments, not to mention the difficulties to recruit participants. In view of these limitations, the number of clusters is set to $k = 9$. The clustering results are presented in Table 4.2 and the centroids are depicted in the map shown in Fig 4.3.

Table 4.2: Summary of the cluster analysis results.

Cluster	1	2	3	4	5	6	7	8	9	Total
No. obs	62,479	86,004	60,063	63,101	44,130	49,956	53,036	119,557	86,164	624,490
WCSS	564,188	578,003	569,158	527,322	464,730	491,656	535,509	759,205	663,009	5,152,782
Variance	9.03	6.72	9.48	8.36	10.5	9.84	10.1	6.35	7.69	15.38

With $k = 9$, the variability of the full set of OD-routes is reduced in 46.4%. If well, this reduction may not be big in terms of clustering analysis, it can be seen (Fig 4.4) that the road attributes of the cluster centroids cover likely values to be observed in the network. To be more specific, 83% of the values of the attributes of the OD-routes in the network lie in the range of the centroids: 83% for the Euclidean distance and the directness, 89% for the freeway composition, 85% for the number of turns per kilometre and 90% for the route length.

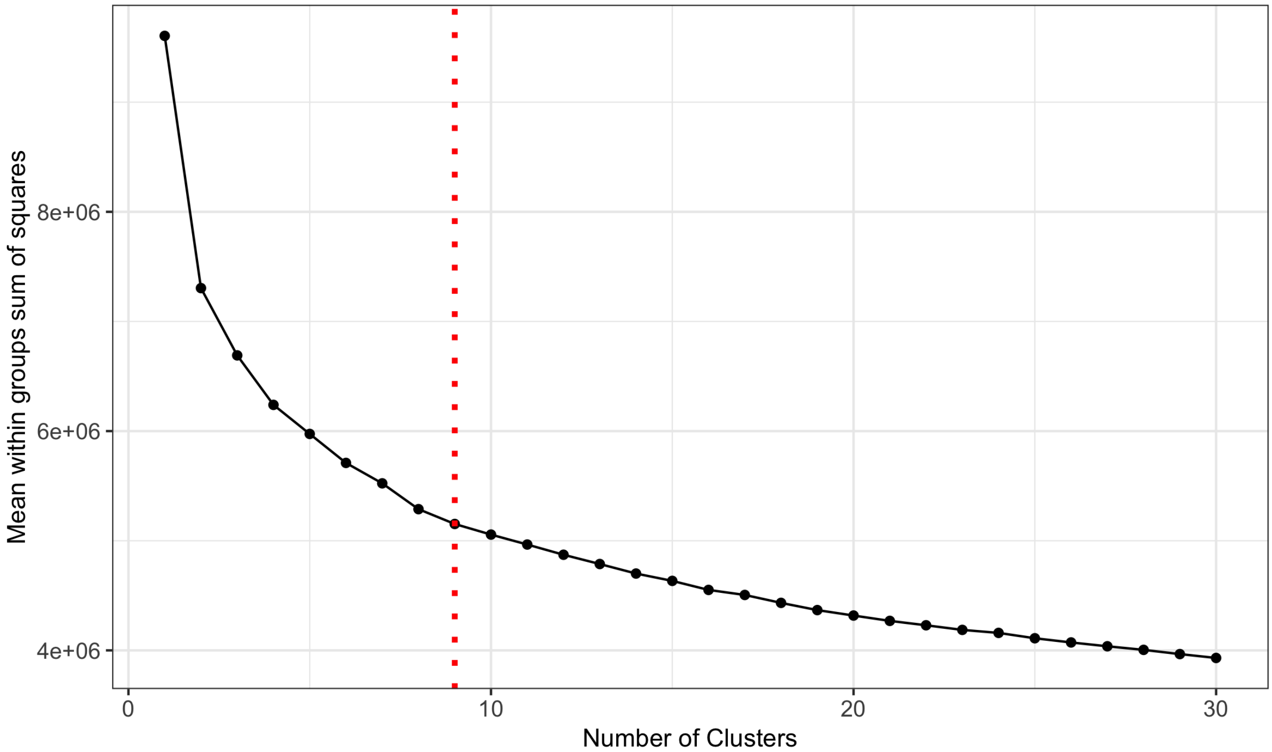

Figure 4.2: Determining the number of clusters k. The sum of squared errors for k-means clustering of the 624,490 OD-routes with $k = 1, \ldots, 30$. After $k = 9$ the decrease in the mean WCSS is marginal.

Furthermore, the resulting p-values of the two-sample Kolmogorov-Smirnov test (Fig 4.4) are high: $p - value > 0.1$ for the five variables, suggesting that there is not enough statistical evidence (with a significance level of $\alpha = 0.1$) to reject the null hypothesis that the values of the attributes of the centroids and the full network come from the same distribution. This implies that a random selected OD pair or route in the network is likely to have attributes similar to one of the nine OD-routes used in the route choice experiments. In this sense, the nine cluster centroids can be regarded as representative of the network.

A further characterisation of the nine clusters, based on their elements' attributes (see appendix 4.A), is proposed as follows:

Clust. C_1**:** Medium-range direct trips going from south to north, with routes having small number of turns per kilometre and some freeway segments.

Clust. C_2**:** Short non direct trips mainly in the central part of the network, with routes having a lot of turns per kilometre and no freeway segments.

Clust. C_3**:** Medium-range direct trips going from north to south, with routes having small number of turns per kilometre and some freeway segments.

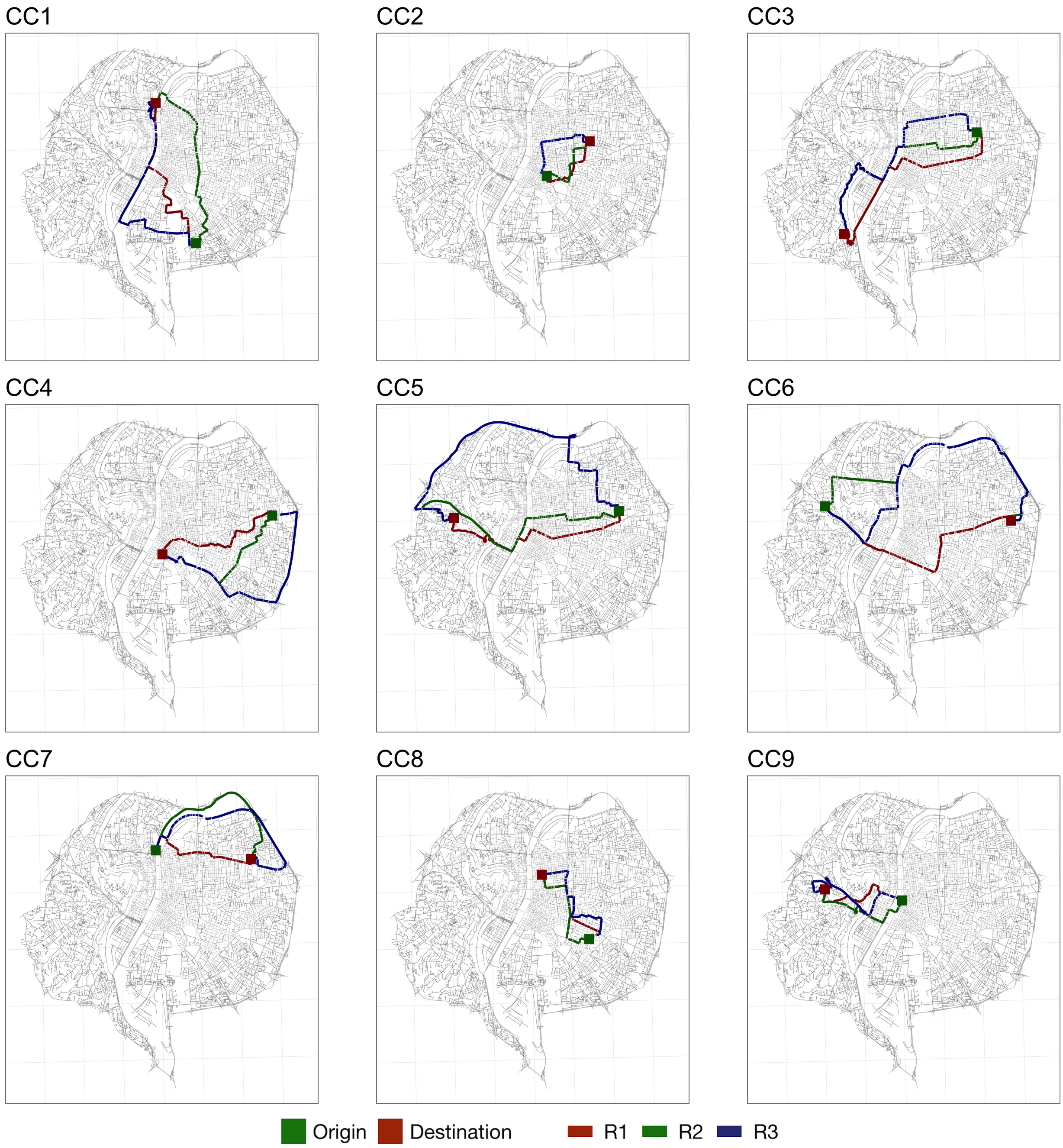

Figure 4.3: Representative OD-routes. Maps of the nine cluster centroids used in the MDG experiments.

Clust. C_4**:** Medium-range direct trips mostly in the central part of the network, with routes having average number of turns per kilometre and with the longest route highly composed of freeway segments.

Clust. C_5**:** Long trips going from east to west, with routes having a small number of turns and with large portions of freeway.

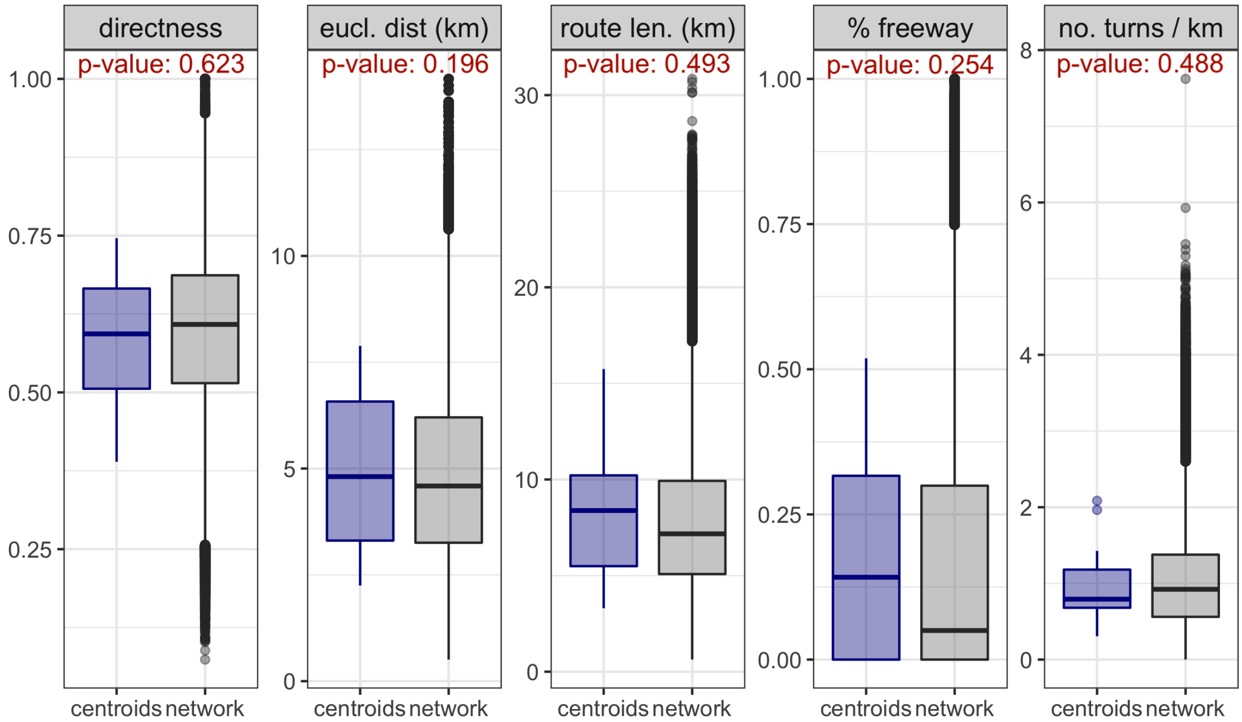

Figure 4.4: Cluster centroids as representative OD-routes. The distribution of the attributes of the selected OD pairs are similar to that of the whole network. The p-values of the Kolmogorov-Smirnov, presented in red in the top of each panel, indicate the lack of statistical evidence (with a confidence level of 0.90) to reject the hypothesis that the two distributions are the same.

Clust. C_6**:** Long trips going from west to east, with routes having a small number of turns and with large portions of freeway.

Clust. C_7**:** Medium-range non direct trips in the central part of the network, with routes with average number of turns per kilometre and high portions of freeway.

Clust. C_8**:** Short direct trips mainly in the central part of the network with routes with low number of turns per kilometre (among short trips) and no freeway segments.

Clust. C_9**:** Short non direct trips mainly in the central part of the network, with routes with low number of turns per kilometre (among short trips) and some freeway segments.

4.3.2 Experiment results

Three route choice experiment sessions were carried out using the nine OD pairs and routes obtained from the clustering analysis of the network. In total, 73 individuals participated in the three sessions, from these participants, 56 (77%) received estimates of the travel times in each route. Participants recorded a total number of 802 choices in the nine defined OD-routes, with an average number of 11 choices per participant, and an average number of 89 choices in each OD-routes. The choices of the participants are presented in Fig 4.5, where it can be immediately noticed that travel time information changes the behaviour of the participants.

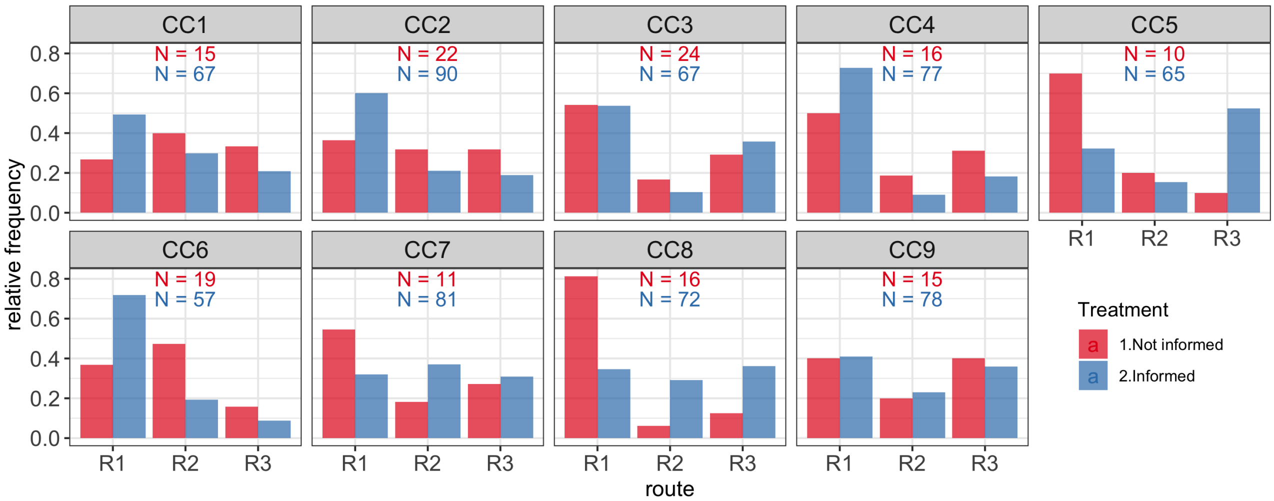

Figure 4.5: Route choice distribution in the nine cluster centroids. The choices of the informed participants are different from those of the uninformed participants.

4.3.3 Route choice model estimation

The estimated parameters for the three models are shown in Table 4.3; more detailed result of the posterior distribution of the parameters can be found in appendix 4.B, and the details of the computational effort for the estimation process in appendix 4.C.

The estimated parameters $\hat{b}_p$ represent the mean preferences in the population. The positive sign of the estimates $\hat{b}_1$ and $\hat{b}_2$ in the three models is interpreted as *the average traveller prefers routes with high composition of freeways, and direct routes*. On the contrary, the negative signs of $\hat{b}_3$ and $\hat{b}_5$ mean that *the average traveller avoids routes with many turns and higher travel times*. These results are in line with the findings in (Papinski et al., 2009), and provide more evidence in favour of travel time as the most important variable in route choice. Note that in the three models $\hat{b}_3 \approx 0$, but with large standard deviations $\hat{\sigma}_3$, meaning that (i) the sign is positive for a large number of participants (near half), and that (ii) even when the mean of the coefficient is close to zero, this variable is still important for a large percentage of the participants, specially in model M3, where $Pr(|\beta_{i3}| > 1) = 0.42$. The case for the length of the route is different. The standard deviation, $\hat{\sigma}_4$ is small for models M1 and M2, implying that the length of the route is not important for the majority of the participants, $Pr(|\beta_{i4}| < 0.2) = 0.93$ and $Pr(|\beta_{i4}| < 0.2) = 0.81$, respectively. However, for model M3 the length of the route becomes more important, $Pr(|\beta_{i4}| < 0.2) = 0.29$, suggesting correlation between these two

Table 4.3: Estimates of the MXL models' coefficients for the nine representative OD pairs. Mean (standard error) of the sampled posterior distributions of the parameters of the MXL models.

Coefficient	M1	M2	M3
$\hat{b}_1$ (FRW_j)	1.96 (0.85)	2.01 (0.74)	2.11 (0.80)
$\hat{b}_2$ (DIR_j)	4.61 (1.65)	4.00 (2.10)	4.56 (1.87)
$\hat{b}_3$ (TNR_j)	-0.15 (0.26)	-0.14 (0.26)	-0.20 (0.30)
$\hat{b}_4$ (LEN_j)	0.01 (0.12)	-0.11 (0.13)	-0.14 (0.16)
$\hat{b}_5$ (ITT_{js})	-3.86 (0.85)	-4.58 (1.01)	-5.28 (1.24)
$\hat{b}_6$ $(ITT_{js} * LEN_j)$	-	0.08 (0.06)	0.13 (0.10)
$\hat{\sigma}_1$ (FRW_j)	0.61 (0.61)	0.58 (0.62)	2.33 (1.15)
$\hat{\sigma}_2$ (DIR_j)	1.10 (1.00)	0.88 (0.98)	3.00 (2.66)
$\hat{\sigma}_3$ (TNR_j)	0.72 (0.39)	0.76 (0.40)	1.24 (0.29)
$\hat{\sigma}_4$ (LEN_{js})	0.11 (0.05)	0.10 (0.06)	0.51 (0.13)
$\hat{\sigma}_5$ (ITT_{js})	4.62 (0.87)	4.64 (0.90)	4.72 (0.91)
$\hat{\sigma}_6$ $(ITT_{js} * LEN_j)$	-	0.07 (0.04)	-
$\hat{\sigma}_{12}$ $(FRW_j\text{-}DIR_j)$	-	-	1.96 (8.73)
$\hat{\sigma}_{13}$ $(FRW_j\text{-}TNR_j)$	-	-	1.21 (1.60)
$\hat{\sigma}_{14}$ $(FRW_j\text{-}LEN_{js})$	-	-	-0.36 (0.65)
$\hat{\sigma}_{23}$ $(DIR_j\text{-}TNR_j)$	-	-	0.63 (2.28)
$\hat{\sigma}_{24}$ $(DIR_j\text{-}LEN_{js})$	-	-	0.83 (2.06)
$\hat{\sigma}_{34}$ $(TNR_j\text{-}LEN_{js})$	-	-	0.00 (0.19)

variables. Finally, note that in models M2 and M3 the mean preference for the length of the routes can be written as $(b_4 + b_6 ITT_{js} INF_i)$, with $b_4 < 0$ and $b_6 > 0$, meaning that the informed travel time diminishes the preference for shorter routes.

4.3.4 Choices on representative OD-routes

Until now, the discussion on the representativeness of the nine selected OD-routes (the cluster centroids) has been in terms of the route attributes. In this section, the representativeness of the OD-routes is assessed in terms of how well a choice model, estimated using the cluster centroids, can be generalised to the entire road network or, in other words, how well it scales-up the travellers' choices to other OD pairs in the network. The hypothesis is that if the choices in the nine cluster centroids are representative of the choices in the entire network, then the predictive accuracy of a model, estimated with observations in the nine cluster centroids, should be higher than the predictive accuracy of models (with the same specification) estimated with observations in random sets of OD-routes. To this end, data collected in other route choice experiments carried out with the MDG platform is used. The data consists of route choice observations in 21 OD-routes, defined with a different methodology for previous experiments, and not comprising the representative OD pairs.

The methodology to validate the representative OD-routes is based on bootstrapping for model validation: at each step, a random part of the data is left-out of the estimation process, and then used to measure the predictive accuracy of the model. However, in this case, the predictive accuracy of the models obtained at each iteration is compared to the predictive accuracy of the model estimated with the nine cluster centroids. The predictive accuracy of model M for a given OD-routes od is defined as the discrepancy between the observed and the predicted choice distributions, given by

$$\mathrm{err}(M, od) = \sum_{j=1}^{3} \max(0, obs_j - pred_j)\,, \tag{4.2}$$

where obs_j and $pred_j$ are the observed and predicted choice proportions (probabilities) for route j. Note that the error measure, $\mathrm{err}(M, od)$ has a direct interpretation in terms of traffic assignment: the percentage of trips that are wrongly distributed amongst the three alternative routes. Let C be the set of choice observations in the 9 cluster centroids and T the set of observations in the 21 test OD-routes. Denote by M^* the model (it can be either M1, M2 or M3) estimated with observations on the nine cluster centroids, C. Then, at iteration r ($r = 1, ..., 40$),

1. sample all observations from 9 randomly selected OD-routes in $(C \cup T)$, define the sampled observations as T_r;

2. estimate the model M_r (M1, M2 or M3) with the observations in T_r;

3. obtain the mean prediction errors of models M^* and M_r for all out-of-sample OD-routes, i.e., for all $od \in (C \cup T) - T_r$,

$$\mathrm{MPE}_r(M^*) = \sum_{od \in (C \cup T) - T_r} w_{od} \cdot \mathrm{err}(M^*, od)$$
$$\mathrm{MPE}_r(M_r) = \sum_{od \in (C \cup T) - T_r} w_{od} \cdot \mathrm{err}(M_r, od)\,,$$

   where the weight w_{od} is the percentage of OD-routes in the cluster to which od belongs, multiplied by the inverse of the number of OD-routes in $(C \cup T)$ that belong to that cluster. The weighting is done to adjust for the probability of observing an OD-routes in the network like od. This follows since some clusters are over-represented in T, as the OD-routes in T were not randomly selected from the network, but they were selected following a different methodology in previous studies.

The $\mathrm{MPE}_r(M^*)$ is compared to the $\mathrm{MPE}_r(M_r)$, $r = 1, \ldots, 40$ for the three model specifications. In Fig 4.6, $\mathrm{MPE}_r(M^*)$ is plotted against $\mathrm{MPE}_r(M_r)$, with blue dots when $\mathrm{MPE}_r(M^*) \leq$

$\text{MPE}_r(M_r)$, and red otherwise. The models estimated with the clusters' centroids performed better in predicting the choices of travellers than most of the models estimated with randomly selected OD-routes. To be more specific, $\text{MPE}_r(M^*) \leq \text{MPE}_r(M_r)$ in 35 out of 40 cases (87.5%) for models M1 and M2, and in 31 cases (77.5%) for model M3. Furthermore, in the cases when the model estimated with the centroids performed worst, i.e., $\text{MPE}_r(M^*) \leq \text{MPE}_r(M_r)$, the errors were close to those of the models estimated with randomly selected OD-routes. Define the improvement of M^* with respect to M_r as $\alpha_r = (\text{MPE}_r(M_r) - \text{MPE}_r(M^*))/\text{MPE}_r(M_r)$. Then, the mean improvements, $\bar{\alpha}$, are 14% for model M1, 14.5% for model M2 and 9.9% for model M3. In 20% of the test cases, α_r is at least 26%, 25% and 22% for models M1, M2 and M3, respectively; and α_r reaches 52%, 48% and 43% in the worst case scenarios. This result highlights the importance of a careful selection of OD pairs in route choice model estimation. As the MPE represents the percentage of trips that are not assigned to the right route, and since the total number of trips at a city level can be very high, about 1 million in the Lyon Metropolis during one day, even low α_r values may have an impact on how the traffic is distributed on the network.

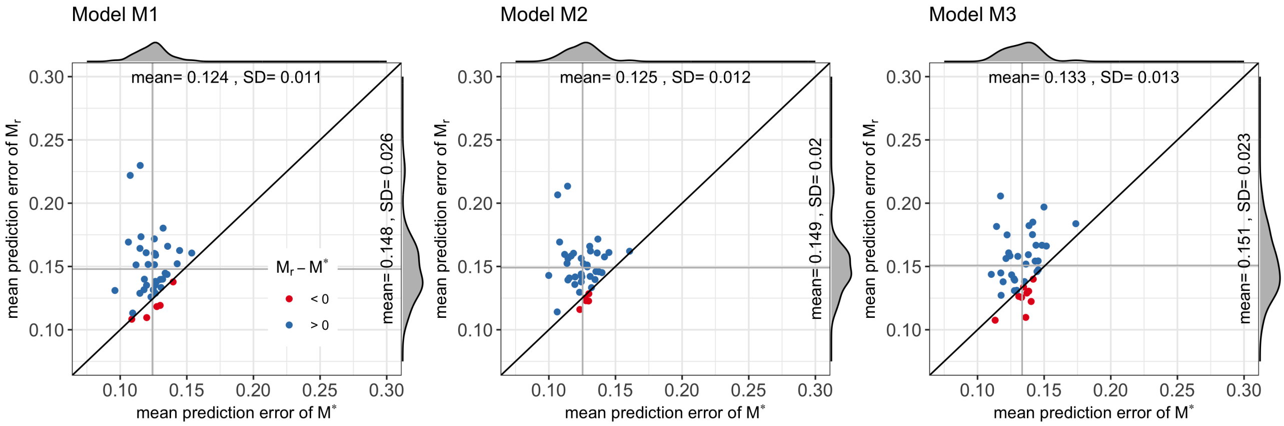

Figure 4.6: Mean predictive errors. The MPEs of model M^* are smaller than the MPEs of models M_r in the majority of the cases (blue dots). Furthermore, in the cases where the MPEs of models M^* are bigger (red dots), the differences are small (close to identity line).

If the MPE is analysed by whether or not the participants received travel time information (Fig 4.7), it can be seen that the models M^* are better than the models M_r for the uninformed participants than for the informed ones. For models M1 and M3, $\text{MPE}_r(M^*) \leq \text{MPE}_r(M_r)$ in 97.5% of the cases, and for model M2 in 95%; and when the participants were informed, $\text{MPE}_r(M^*) \leq \text{MPE}_r(M_r)$ in 65% for model M1, 77.5% for model M2 and 67.5% for M3. In the case of the uninformed participants the values of $\bar{\alpha}$ are 20%, 14% and 15%, respectively for models M1, M2 and M3. The high performance of the models M^* for the uninformed participants implies that the models are capable of approximating the choices of this group in a variety of scenarios, i.e., the models estimated with the nine centroids generalise well to other OD-routes for this group. Moreover, considering that the informed travel time was not part of the variables used in the clustering of the OD-routes (% of freeway, directness, no. of turn per kilometre, distance), this result suggests that the choices in the cluster centroids are representative of the choices in the entire network, thus validating the methodology proposed in this chapter.

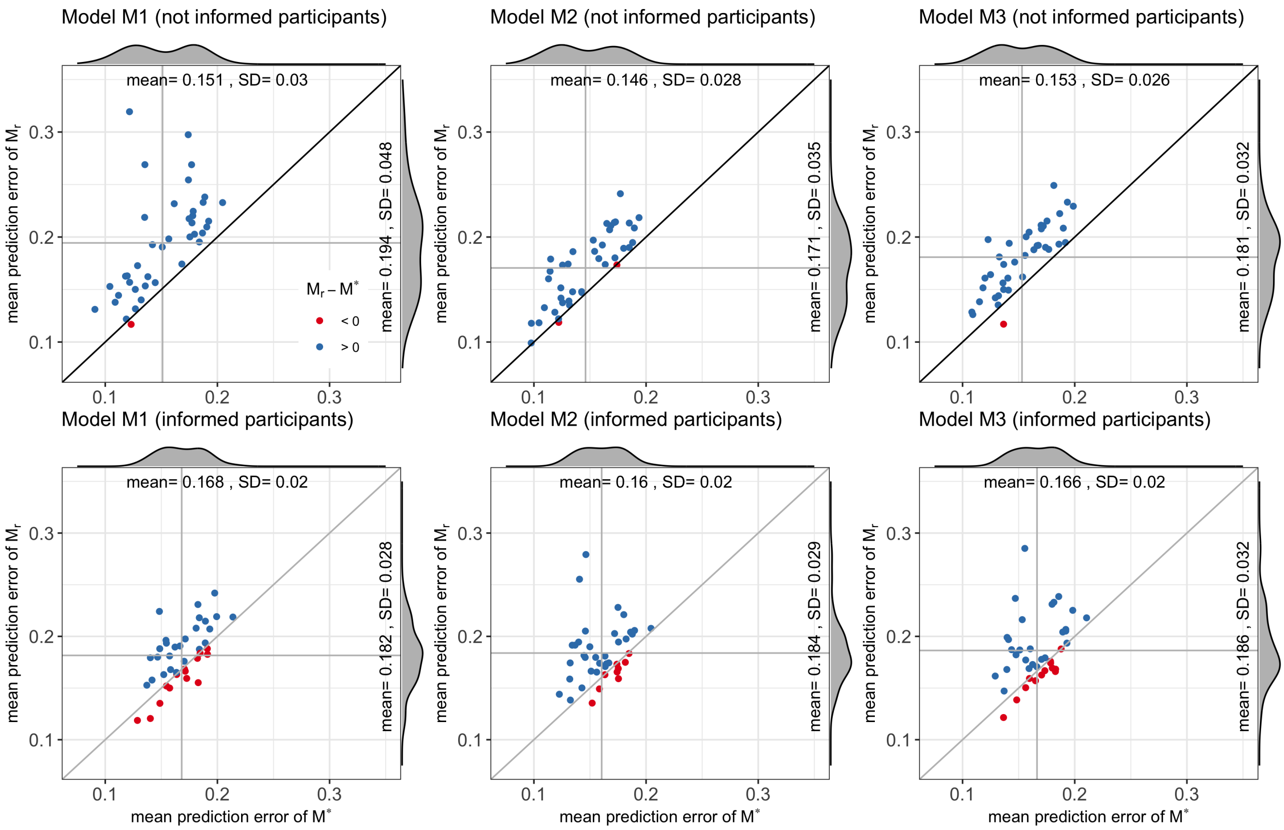

Figure 4.7: Mean predictive errors by information group. The models estimated with the cluster centroids are clearly better in predicting the choices for the uninformed participants.

The predictive errors of the representative models, M^*, and the test models, M_r, can be disaggregated by OD-routes. In the results, shown in Fig 4.8, it is clear that the magnitude and the variance of the predictive errors depend on the OD-routes where the choices are being predicted. The choices in some OD-routes are difficult to predict, regardless of the training set used to estimate the models. Furthermore, there is no clear pattern indicating that these errors are associated with the road characteristics of the OD-routes: two OD-routes belonging to the same cluster, i.e., having similar route attributes, may have a low and a high prediction error. Such is the case of OD-routes c3_od2 and c3_od3, both belonging to cluster C_3, but with errors below 0.1 for the former and above 0.2 for the later. Similar cases can be found in cluster C_5 and C_8. An important observation is that the models estimated with the cluster centroids M^* are not as accurate in predicting the choices in individual OD-routes as the models M_r, for some values of r. In fact, their prediction errors are amongst the lowest 25% in only 8 out of 21 test OD-routes for models M1 and M3, and in 6 for model M2. However, at the same time, the individual errors are almost never amongst the highest 75%: in 0 OD-routes for model M1, in 2 for model M2, and in 1 for M3. Moreover, when the individual errors are averaged to obtain the MPE (as in the previous analysis), the models M^* outperform the models M_r for the majority of values of r. This result implies that a model M_{r_0} having low prediction errors for some OD-routes has also high prediction errors in other OD-routes, and therefore its mean predictive accuracy is reduced. In this sense, the models estimated with the cluster centroids, M^*, are preferred, as they will show a relative better global prediction accuracy without incurring in large errors in individual OD-routes.

The models estimated with the representative OD-routes, M^*, are compared in terms of their prediction errors over the 21 validation OD-routes. The error distributions of the three models, depicted in Fig 4.9, show that, practically, there is no difference in the predictive accuracy. This means that the interaction between the informed travel time and length of the route in models M2 and M3 does not improve the predictive accuracy; nor considering the correlations in model M3 does.

4.4 Conclusions and discussion

In this study, it was demonstrated that the choices of participants in a route choice experiment over a small but representative set of OD configurations can be scaled-up to the entire network. To obtain the set of representative OD configurations, a new methodology based on k-means cluster analysis is proposed. First, the OD configurations in the network, i.e., the OD pairs and three connecting routes, are represented in vector form according to the attributes of the OD pairs and routes. Then, these OD configurations are clustered in order to obtain a partition of the road network and the cluster centroids selected as representative of the entire network. The main hypothesis is that the choices of travellers over the entire network can be approximated with route choice models estimated using data collected for the representative set. The obtained results point in this direction. These results were obtained estimating the models over choice sets with three alternative routes and it is left as future work to see if the results generalise to larges choice sets. This may pose two difficulties. The first related to the definition of the OD-routes vector, as the number of alternative routes vary across OD pairs. The second difficulty is related to experimentation and the high cognitive burden of presenting participants with a large number of alternatives.

In the current study, for the city of Lyon in France, 9 OD pairs and their connecting routes were used as representative of 624,490 OD configurations (OD pairs connected by three alternative routes). These nine representative OD configurations cover around 83% of the values

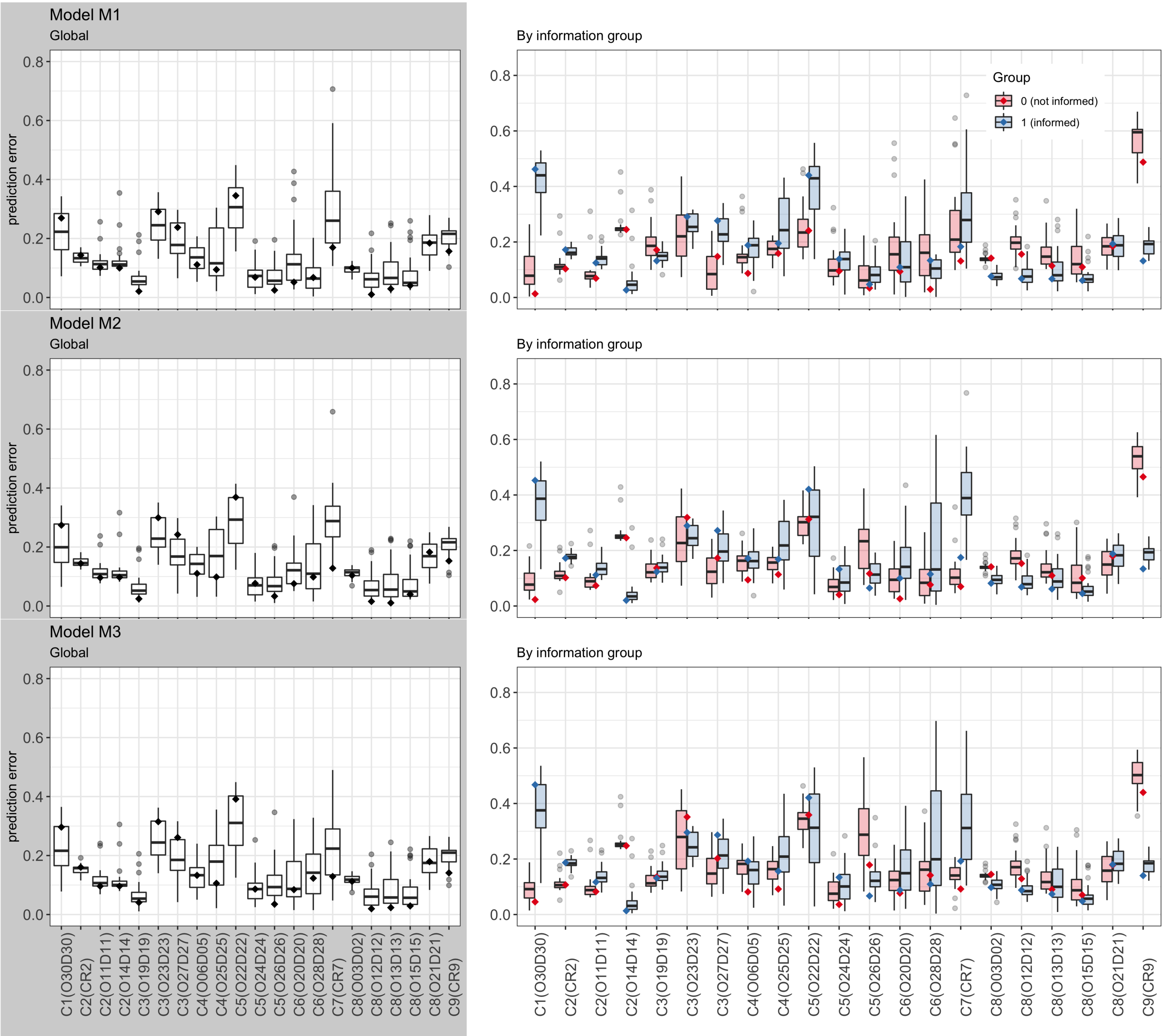

Figure 4.8: Distributions of the predictive errors the 21 validation OD-routes. The level and the variability of the errors amongst the different OD-routes imply that the choices in some OD-routes are difficult to predict, regardless of the training set used to estimate the models.

of the attributes of the OD-routes in the network. The predictions of the models estimated with the representative set were superior in most of the test cases (87.5% and 77.5% in the general case). For the uninformed participants, whose decisions were based on the same attributes used in the clustering, the predictions are better in at least 95% of the test cases. By estimating the route choice model with the cluster centroids, the mean prediction errors are reduced by up to 14.5% for model M1 (similar results are observed for models M2 and M3). The reduction of the prediction error is more than 22% for the 20% of the test cases, and it goes up to 51% in the worst case. This demonstrates that a careful selection of the OD configurations significantly improves the prediction accuracy. Another significant finding, is that the models estimated with

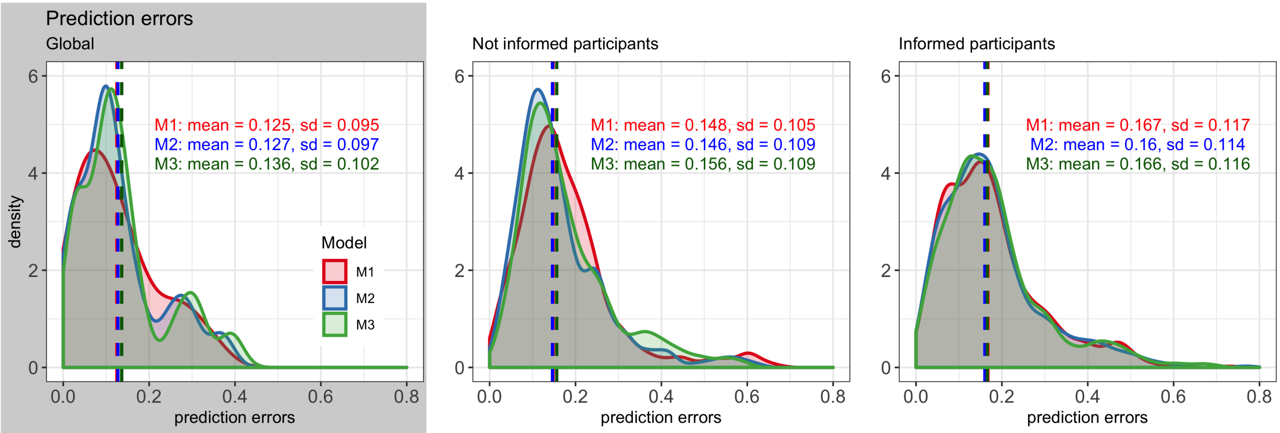

Figure 4.9: Distributions of the prediction errors of the models M^* on the 21 validation OD-routes. There are no significant differences between the error distributions.

the representative OD configurations are more robust than the ones obtained from the models with random OD configurations. The models estimated with the representative set never show extreme errors for individual OD pairs, contrary to the models estimated with random sets of OD configurations. This implies that the models estimated with the representative set will show a relative better global prediction accuracy without incurring in large errors on individual OD-routes. This result is important when predicting the trip distribution over the network, as high errors in individual OD pairs may have significant impact in local traffic conditions, causing spreading. The last finding is that estimating the models with the representative OD pairs leads to an average prediction error of 12.7%. This value can be considered quite low when considering the scale of the city, the heterogeneity of OD configurations, and the actual performance of user equilibrium approaches.

From the clustering analysis in this study, it is clear that there are OD pairs in the network that are not well represented by the representative set of nine OD configurations. Therefore, it cannot be claimed that the choices in these non-represented OD pairs can be well approximated by the set of nine OD pairs found in this study. However, these non-represented OD pairs are those with attributes not covered by the representative set, which are no more than 17% of the OD configurations in the network. Note that this result does not hinder the usefulness of the proposed methodology, as it can be extended by either using other clustering techniques that allow taking into account for these *atypical* OD configurations or by including more clusters in the representative set.

4.A Cluster characterisation

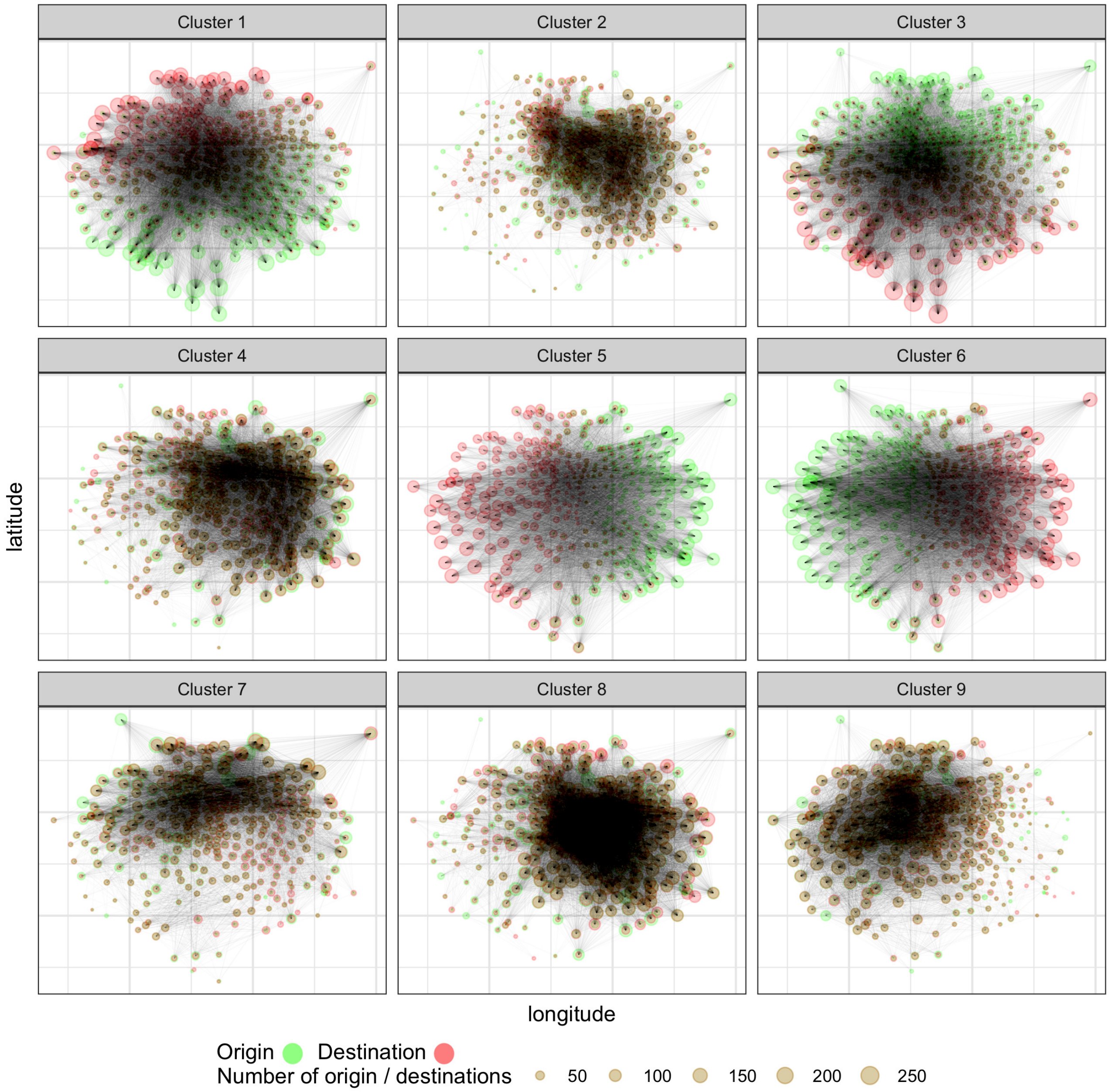

Figure 4.10: Origin-destination zones in each cluster. The clusters show differentiated geographical patterns. For example, clusters 1,3,5 and 6 represent trips that originate and terminate in the limits of the network and, thus, represent longer trips.

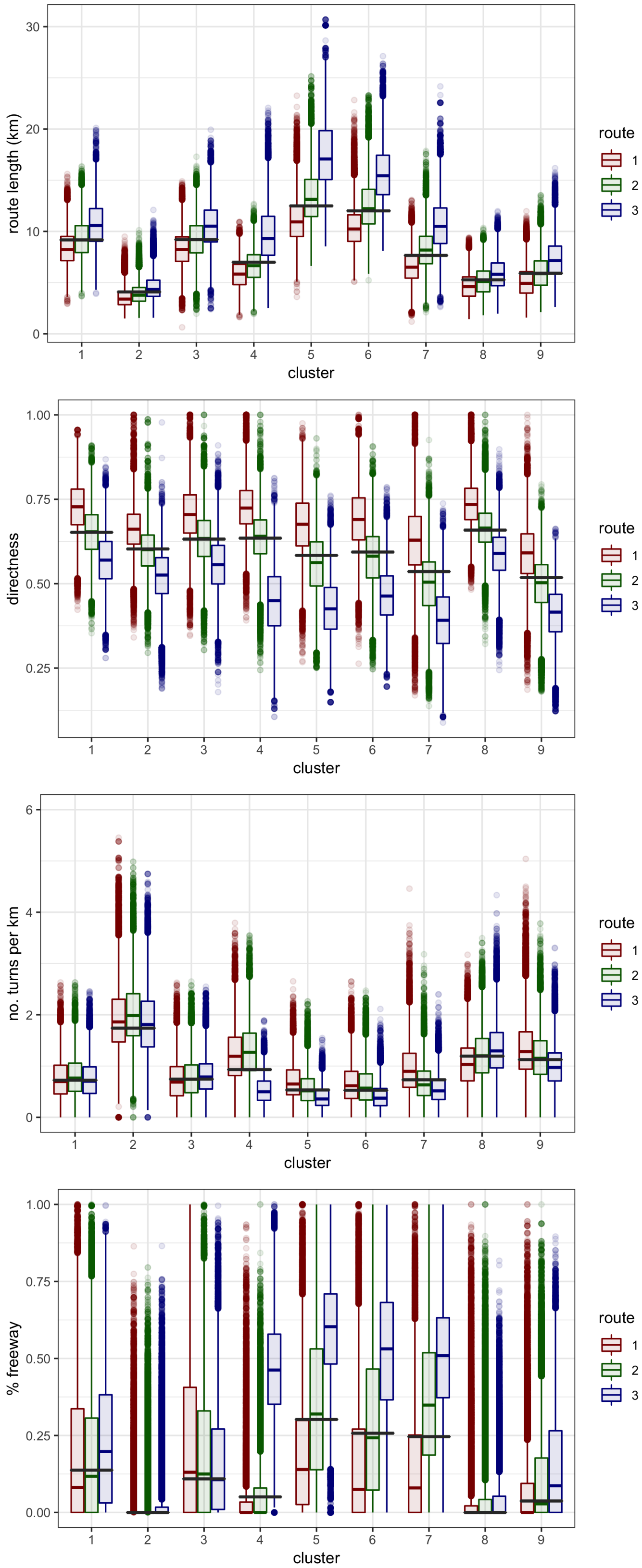

Figure 4.11: Attributes of the routes in the clusters. The attributes of the routes show differentiated attributes between the clusters.

4.B MXL models' estimates

Table 4.4: Complete estimates of the MXL models' coefficients. Mean, standard deviation and some quantiles of the sampled posterior distributions of the parameters of the MXL models.

			M1				
Coefficient	mean	s.d.	q. 2.5%	q. 25%	q. 50%	q. 75%	q. 97.5%
$\hat{b_1}$ (FRW_j)	1.96	0.85	0.44	1.31	1.92	2.67	3.47
$\hat{b_2}$ (DIR_j)	4.61	1.65	1.22	3.56	4.64	5.73	7.55
$\hat{b_3}$ (TNR_j)	-0.15	0.26	-0.65	-0.33	-0.15	0.02	0.36
$\hat{b_4}$ (LEN_j)	0.01	0.12	-0.26	-0.07	0.02	0.09	0.22
$\hat{b_5}$ (ITT_{js})	-3.86	0.85	-5.61	-4.42	-3.85	-3.29	-2.23
$\hat{\sigma_1}$ (FRW_j)	0.61	0.61	0.04	0.15	0.38	0.88	2.19
$\hat{\sigma_2}$ (DIR_j)	1.10	1.00	0.04	0.20	0.85	1.75	3.43
$\hat{\sigma_3}$ (TNR_j)	0.72	0.39	0.05	0.43	0.74	1.00	1.46
$\hat{\sigma_4}$ (ITT_{js})	0.11	0.05	0.03	0.07	0.10	0.14	0.23
$\hat{\sigma_5}$ (ITT_{js})	4.62	0.87	3.12	4.00	4.55	5.15	6.52
			M2				
Coefficient	mean	s.d.	q. 2.5%	q. 25%	q. 50%	q. 75%	q. 97.5%
$\hat{b_1}$ (FRW_j)	2.01	0.74	0.69	1.47	1.96	2.57	3.41
$\hat{b_2}$ (DIR_j)	4.00	2.10	0.74	2.34	3.85	5.80	7.81
$\hat{b_3}$ (TNR_j)	-0.14	0.26	-0.64	-0.32	-0.15	0.03	0.35
$\hat{b_4}$ (LEN_j)	-0.11	0.13	-0.32	-0.21	-0.12	-0.01	0.13
$\hat{b_5}$ (ITT_{js})	-4.58	1.01	-6.60	-5.25	-4.57	-3.91	-2.64
$\hat{b_6}$ $(ITT_{js} * LEN_j)$	0.08	0.06	-0.04	0.04	0.08	0.12	0.21
$\hat{\sigma_1}$ (FRW_j)	0.58	0.62	0.03	0.11	0.31	0.91	2.15
$\hat{\sigma_2}$ (DIR_j)	0.88	0.98	0.03	0.11	0.44	1.41	3.41
$\hat{\sigma_3}$ (TNR_j)	0.76	0.40	0.05	0.48	0.79	1.05	1.50
$\hat{\sigma_4}$ (ITT_{js})	0.10	0.06	0.03	0.06	0.09	0.13	0.23
$\hat{\sigma_5}$ (ITT_{js})	4.64	0.90	3.11	4.01	4.57	5.19	6.62
$\hat{\sigma_6}$ $(ITT_{js} * LEN_j)$	0.07	0.04	0.02	0.04	0.07	0.10	0.17
			M3				
Coefficient	mean	s.d.	q. 2.5%	q. 25%	q. 50%	q. 75%	q. 97.5%
$\hat{b_1}$ (FRW_j)	2.11	0.80	0.61	1.56	2.10	2.63	3.75
$\hat{b_2}$ (DIR_j)	4.56	1.87	0.92	3.31	4.58	5.88	8.19
$\hat{b_3}$ (TNR_j)	-0.20	0.30	-0.81	-0.40	-0.20	0.01	0.38
$\hat{b_4}$ (LEN_j)	-0.14	0.16	-0.46	-0.25	-0.14	-0.03	0.17
$\hat{b_5}$ (ITT_{js})	-5.28	1.24	-7.76	-6.10	-5.26	-4.43	-2.87
$\hat{b_6}$ $(ITT_{js} * LEN_j)$	0.13	0.10	-0.06	0.06	0.12	0.20	0.33
$\hat{\sigma_1}$ (FRW_j)	2.33	1.15	0.83	1.39	2.04	3.10	4.91
$\hat{\sigma_2}$ (DIR_j)	3.00	2.66	0.79	1.25	1.95	3.43	10.57
$\hat{\sigma_3}$ (TNR_j)	1.24	0.29	0.76	1.04	1.22	1.43	1.91
$\hat{\sigma_4}$ (ITT_{js})	0.51	0.13	0.37	0.44	0.48	0.55	0.85
$\hat{\sigma_5}$ (ITT_{js})	4.72	0.91	3.11	4.08	4.66	5.29	6.66
$\hat{\sigma_{12}}$ $(FRW_j\text{-}DIR_j)$	1.96	8.73	-14.89	-0.84	0.51	3.68	25.0
$\hat{\sigma_{13}}$ $(FRW_j\text{-}TNR_j)$	1.21	1.60	-0.85	0.15	0.79	1.85	5.48
$\hat{\sigma_{14}}$ $(FRW_j\text{-}ITT_{js})$	-0.36	0.65	-2.21	-0.53	-0.19	-0.02	0.58
$\hat{\sigma_{23}}$ $(DIR_j\text{-}TNR_j)$	0.63	2.28	-2.96	-0.39	0.30	1.22	6.41
$\hat{\sigma_{24}}$ $(DIR_j\text{-}ITT_{js})$	0.83	2.06	-0.35	-0.01	0.15	0.58	7.09
$\hat{\sigma_{34}}$ $(TNR_j\text{-}ITT_{js})$	0.00	0.19	-0.45	-0.10	0.01	0.11	0.38

4.C Computational effort

Table 4.5: Execution time (in seconds) to draw 30,000 samples of the posterior distribution of the parameters. Estimation for the 9 cluster centroids which are composed of 802 route choices.

Model	user	system	elapsed
M1	347.983	0.549	350.600
M2	452.597	0.959	458.878
M3	372.546	1.014	376.076

Table 4.6: Hardware and software specifications.

Hardware	
Model Name	iMac
Processor Name	Intel Core i5
Processor Speed	3.2 GHz
Number of Processors	1
Total Number of Cores	4
L2 Cache (per Core)	256 KB
L3 Cache	6 MB
Memory	16 GB
Software	
Operating System	macOS 10.14.2 (18C54)
R version	R version 3.5.1 (2018-07-02)
JAGS version	4.3.0

5 Travel time and bounded rationality in travellers' route choice behaviour

Recent empirical studies have found that travellers route choices deviate from perfect rationality, by showing that urban trips do not necessarily follow the shortest-time routes (Papinski et al., 2009, Thomas and Tutert, 2010, Zhu and Levinson, 2015, Hadjidimitriou et al., 2015, Yildirimoglu and Kahraman, 2018b). However, there is no consensus on how much the travellers' route choice behaviour deviates from the perfect rational assumption. The objective of this study is to contribute to the understanding on how travellers process travel time when making route choices, and to quantify to what extent users are strict travel time minimisers or if bounded rationality is observed. The question of whether travellers evaluate travel time differences in absolute or relative terms is also addressed, and the heterogeneity in the route choice behaviour of travellers investigated. The results of route choice computer experiments, focused on the route choices in diverse OD pairs and traffic conditions, are analysed. In total, 496 participants recorded 5,535 route choices over 41 OD pairs. It was found that travellers evaluate relative rather than absolute differences in travel time. In 60.5% of the trips participants chose the fastest route, but this percentage is 80% when the travel time of the alternatives is at least 30% higher than the fastest route. Only 10% of the individuals chose the fastest route in all trips, confirming the hypothesis of bounded rationality. The participants exhibited heterogeneous travel time indifference bands: at least 70% of them would not consider routes with travel times 1.5 times slower than the fastest alternative; the average participant was indifferent to relative travel time differences of less than 31%.

This chapter is a journal article currently under review: González Ramírez, H., Leclercq, L., Chiabaut, N., Becarie, C., Krug J. (2019). Travel time and bounded rationality in travellers' route choice behaviour: a computer route choice experiment. Manuscript submitted for publication to Travel Behaviour and Society.

5.1 Motivation

Travel time is often considered the most important variable in explaining the route choice behaviour of travellers (Bovy and Stern, 1990). From an individual point of view, routes with longer travel times result in higher opportunity costs, i.e., less time that the traveller could allocate into other activities (value of time), thus, decreasing the likelihood of being chosen. When studying traffic assignment, it is traditionally assumed that travellers are perfectly rational, in the sense that they know the travel times in all the alternative routes (perfect information) and they will always choose the one with the minimum travel time. As a consequence of this hypothesis, the traffic states in a transportation network must fulfil the User Equilibrium (UE) condition, originally stated by Wardrop (1952): "the journey times in all routes actually used are equal and less than those that would be experienced by a single vehicle on any unused route". By relaxing the perfect information assumption (but not the rationality of users), Sheffi (1985) defined the Stochastic User Equilibrium (SUE). At the individual level, SUE means that users are still strict optimisers, i.e., they choose the minimum travel time alternative, but they have no perfect information from the travel times in the system. However, recent empirical studies have shown that travellers do not necessarily choose the minimum travel time route. In the study of Zhu and Levinson (2015), GPS itineraries were collected from 143 residents of the Minneapolis-St. Paul metropolitan area (United States) during a period of 13 weeks, finding that 40% of the trips followed the strict shortest-time path. More important, in almost 90% of the trips travellers chose routes no more than 5 minutes longer than the shortest time route, meaning that users may not be strict optimisers, but consider travel time as a key decision input. In a related study, Yildirimoglu and Kahraman (2018a,b) use GPS trajectories of taxi trips in the city of Shenzhen, China, to compare the actual paths followed to those implied by UE. The results show that 38.2% of the taxi trips followed the shortest-time path. Similar results can also be found in the work of Bekhor et al. (2006), who by analysing data of 188 participants in a survey consisting of the description of their habitual route to work, found that 37% chose a route that overlaps in 90% with the shortest-time alternative, or in the work of Papinski et al. (2009), who examined the GPS traces and survey answers of 31 individuals residing in Ontario, Canada. In the survey, approximately 50% of the individuals stated that minimising travel is the most important factor in their route choices. These values, however, differ considerably from the results reported in other studies. Hadjidimitriou et al. (2015) analyse the GPS coordinates of 89 travellers in the province of Reggio Emilia, Italy, concluding that only 25% of the trips matched the shortest path route (considering a match to be the routes that overlap at least in 80% with the shortest route). The authors found that travellers selected routes on average 1.3 longer than the shortest path. By analysing the data from an experiment in Virginia, United States, involving 20 participants who completed trips on 5 OD pairs over a period of 20 days, Vreeswijk et al. (2014) found that in 74% of the cases the average shortest time route was chosen. However, this percentage varies from 63% to 90% depending on the OD pair. Thomas and Tutert (2010) used license plate observations in the Dutch city of Enschede to conclude that 75% of the trips followed the shortest time paths.

The fact that travellers do not necessarily choose the fastest route is explained, on the one hand, by the presence of other route attributes that make some alternatives more desirable than others. For example, distance, the number of intersections, traffic lights, complexity of the paths, the percentage of freeway, aesthetics (Bovy and Stern, 1990, Ramming, 2002, Bekhor et al., 2006, Papinski et al., 2009) and travel time reliability (Avineri and Prashker, 2005, Abdel-Aty et al., 1997, Mahmassani and Liu, 1999). On the other hand, sub-optimal choices of travellers are explained by limitations in travel time perception and cognitive biases that cause deviations from perfect rationality (see Di and Liu (2016) for a review on cognitive biases in route choice behaviour). The cognitive limitations of human reasoning are the cornerstone of bounded rationality. Under bounded rationality, decision-makers search a solution until a satisfactory (not necessarily optimal) alternative is found, thus departing from perfect rationality. This idea was introduced by Simon (1957) as an alternative model of decision-making process that departs from the classical utility maximisation assumption of expected utility or random utility models (Manski and McFadden, 1981, Train, 2003, Walker and Ben-Akiva, 2002). In the context of traffic, bounded rationality was first discussed in Mahmassani and Chang (1987) who introduced the notion of "indifference band" and studied network equilibrium under bounded rationality assumption (BRUE). The idea of indifference band is that travellers are only willing to switch their usual route when time savings are above a threshold. Or, to put it another way, a decision-maker is indifferent to the travel time of the alternatives when their difference is under a threshold (indifference band). The set of alternatives under this condition are called *satisficing*, a term coined by Simon (1957) to refer to alternatives that both satisfy and suffice. By modifying a random utility model to include this threshold, Watling et al. (2018) propose a bounded choice model and formulate the bounded stochastic equilibrium (Bounded SUE). Bounded rationality could therefore explain why travellers do not necessarily choose the shortest-time routes, but close alternatives that may have other appealing features while being considered equivalent from strict travel time point of view.

In the above-cited studies, there is no consensus on the number of travellers that follow the shortest time route, nor the size of the indifference band: the percentage of travellers that chose the fastest route ranges from 25% to 75%. Moreover, five of the six studies are revealed preference (RP), i.e., those based in GPS traces and license plate observations. While RP methods are not affected by validity issues, they have the disadvantage of low control of the experimental environment, meaning that the diversity of explored situations may be limited. In the context of route choice, the list of alternatives and, more important, the related travel times are not known, making it necessary to infer them. This could introduce some errors in the estimates of the proportion of route choices for the shortest time route. Additionally, in some of the experiments the number of participants is small (20 and 31) and thus the estimates may not generalise to the segment of the population under study. The objective of this study is to contribute to understanding how travellers process travel time when making route choices and to quantify to what extent users are strict travel time minimisers or if bounded rationality explains better the observed choices. Also, a Mixed Logit Model (MXL) (McFadden, 1984, McFadden and Train, 2000, Walker and Ben-Akiva, 2002), estimated considering only satisficing alternatives, is compared to the estimates of an unrestricted MXL model to assess the impact that the indifference bands may have on the route choice probabilities. The question of whether travellers evaluate travel time differences in absolute or relative terms is addressed. Does a difference of 5 minutes weigh equally in a 10-minute trip as in a 30-minute trip? The answer to this question is necessary in determining the indifference band. To this purpose, the results of several stated preference route choice computer experiments, carried out using a dedicated simulation game platform, are analysed.

Computer-based experiments have been largely used to study the route choices of travellers, with particular attention to the study of how travellers learn from experience (Bogers, 2005, Selten et al., 2007), the impact of advanced travel information systems (ATIS) (Adler and McNally, 1994, Mahmassani and Liu, 1999, Ben-Elia and Shiftan, 2010, Abdel-Aty et al., 1997, Srinivasan and Mahmassani, 2000, Bifulco et al., 2014), the effect of travel time variability and risk attitudes on the travellers choices (De Moraes Ramos et al., 2013, Avineri and Prashker, 2005, Bogers et al., 2006), and the impact of human choices on network performance (Iida et al., 1992, Tawfik et al., 2010), to mention some. In this chapter, experiments focus on travellers' route choices considering travel time information. Participants made choices over 41 OD pairs in the network of the city of Lyon, France, joined by three alternative routes and presented over a map representation of the city. The OD pairs and routes were selected such that the values of their physical attributes (length, directness, number of intersections, number of turns and freeway composition) show a significant variation, while the routes remain plausible alternatives. Furthermore, the traffic conditions in the network, and thus the travel times in the routes, varied between and within the different experiments. The variability of the route attributes and travel times make it possible to study their joint effect on participants' choices. In total, 496 participants recorded 5,535 route choices. From the total number of participants, 71% received travel time estimates in the three alternative routes, eliminating the travel time perception bias from the analysis, and providing a common ground to study the (bounded) rationality in route choice behaviour in the presence of travel time information.

The rest of the chapter is organised as follows. In Section 5.2, the route choice experimental tool, the Mobility Decision Game (MDG) is described. The methodology to estimate the perfect and bounded rationality, as well as the size of the indifference band is introduced in Section 5.3. The specification of a Mixed Logit Model (MXL) that only accounts for satisficing routes is also presented in this section. In Section 5.4, the results are discussed. First, a global analysis on the travel time minimisation behaviour of travellers is done, including the heterogeneity of the travel time minimisation behaviour by OD pair and by participant. Second, the effect of the absolute (time) and relative (percentage) travel time differences in the travel time minimisation behaviour of travellers is studied. Third, the bounded rationality in the route choice behaviour is analysed, and the heterogeneity of the indifference band is addressed: the distribution of the indifference bands in the population is estimated. Finally, the results of the MXL model are analysed to study the effect of the indifference bands on the probability of route choice. This also permits to test the influence of the route attributes on the choices of the participants. A summary of the important results and the main conclusions are presented in Section 5.5.

5.2 Computer route choice experiments

The data on route choice behaviour in this chapter comes from 6 route choice experiments carried out between February 2018 and February 2019. In total, 496 individuals participated in the experiments. The participants were students from the University of Lyon taking part in the courses of traffic theory (66%), staff from the IFSTTAR (French Institute of Science and Technology for Transport, Development and Networks) and other universities, who received an invitation by e-mail to remotely join the experiments via a web browser (34%). The great majority of the participants, 80%, are from the city of Lyon, 10% from other cities in France, and 10% from other countries. All participants have signed, before the experiments begin, an informed consent form describing the objectives of the study, the data collection and processing, and the confidentiality rules. Participants could opt out of the experiment at any time. No personal data were mandatory to participate to the experiments as people had the opportunity to identify themselves by a login of their choice. Finally, all data were fully anonymised and processed as such. At the beginning of the experiments, the participants were briefed about the objective of the experiment and the interface of the experimental platform; for the participants that joined the experiments via web, a document with the instructions was shared. The participants were instructed to choose the route that they consider to be the *best* to complete a trip on time.

From the 496 participants, 353 (71%) received traffic information as estimates of the travel time in the alternative routes. The participants recorded a total number of 5,535 choices amongst three alternative routes in the 41 OD pairs (Fig 5.1). It is important to mention that not all of the 41 OD pairs were available in each experiment in order to guarantee a sufficient number of observations in each one: the maximum number of OD pairs in a single experiment was 15. The distribution of the number of choices per participant is presented in Fig. 5.2(a), where it can be seen that participants recorded a different number of choices; the average is 11.2. These choices are distributed over an average of 5.41 OD pairs, meaning that participants repeated, on average, 2 choices in the same OD pair (Fig. 5.2(b)). The variation on the number of choices per participant is explained by the duration of the experiments and the availability of the players: some experiments were carried out in sessions of 30 minutes while others in sessions of 1 hour and participants could opt out of the experiment at any moment.

Figure 5.1: Choice distributions for the 41 OD pairs in the MDG experiments.

It is worth mentioning that, even when participants made repeated choices, learning is not observed. The learning process was limited by the design of the experiments, where participants make several simultaneous choices (up to 10), i.e., they do not have to wait until a trip is completed to make the next choice. Furthermore, the OD pairs in the MDG are not presented in any particular order, so participants make choices in other OD pairs before encountering a repetition. As a result, participants might have trouble memorising the travel time information provided in their past choices. This, along with the low number of repetitions in the same OD pair (2 on average) prevented participants from learning. To see this in a quantitative manner, the trend of the percentage of times that the fastest route is chosen is analysed against the

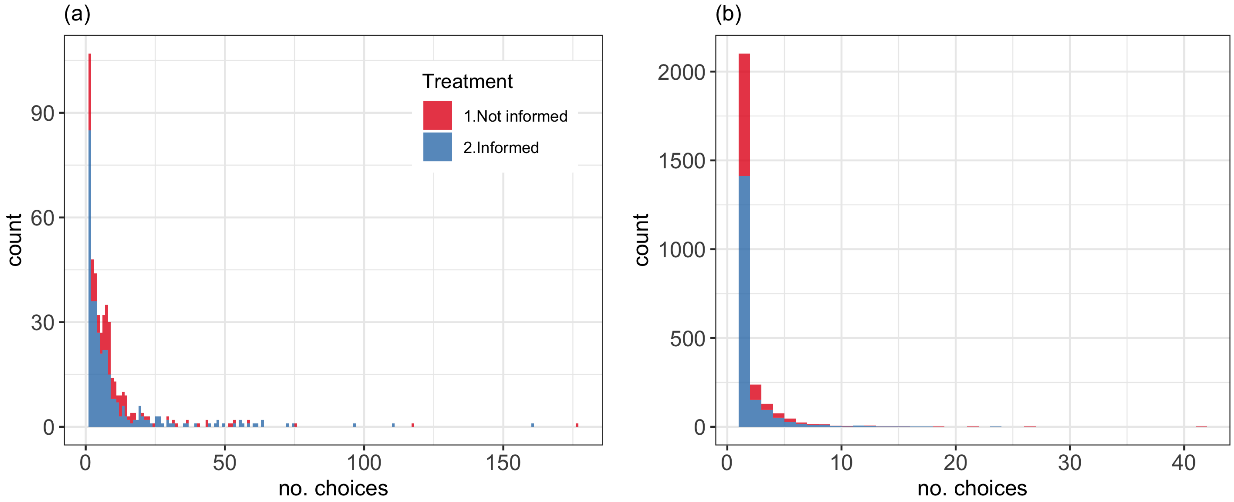

Figure 5.2: The distributions of (a) the number of choices per participant and (b) the number of choices per participant per OD pair. This later plot shows that participants barely made more than 2 choices in the same OD pair.

ordered choice number. Let $F(t)$ be the percentage of times that the fastest route is chosen in choice number t, then if there is a learning process, one would expect that $F(t+1) - F(t) > 0$, i.e., a positive trend in the series. The regression $F(t+1) - F(t) = \phi + \epsilon$ is estimated, and the hypothesis test $H_0 : \phi \leq 0$ is performed. The analysis is done for the participants that received travel time information, as the rest of the chapter concerns mainly this group. Note that since the number of choices per participant vary (Fig. 5.2(a)), the values $F(t)$ are obtained with different number of observations: $F(1)$ is estimated with the first choice of participants and thus all participants contribute to its computation; $F(20)$ is estimated with the 20-th choice of participants, but there are only around 10% of participants that made at least 20 choices. Therefore the observations need to be weighted in the regression. The result of the regression is $\hat{\phi} = 0.0066$ with a standard error of 0.0136. The test for $H_0 : \phi \leq 0$ (p-value = 0.3146) suggests that there is not enough evidence to reject the null hypothesis with a high significance level (significance 0.1). Hence, no learning process is suspected. The differences $F(t+1) - F(t)$ are presented in Fig. 5.3 along with $\hat{\phi}$.

5.3 Methodology

5.3.1 Travel time minimisation behaviour

To study to what extent the participants in the experiments are travel time minimisers, the *minimisation rate*, defined as the proportion of times that the fastest route informed to participants was chosen is computed. When travel time is the only variable that travellers take

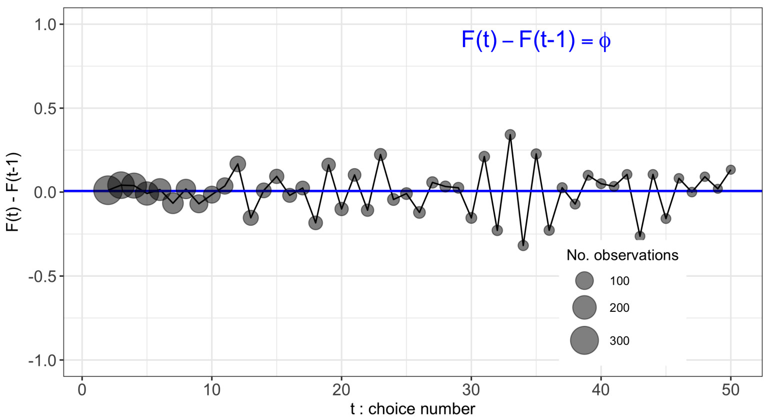

Figure 5.3: Differences in the proportion of choices for the fastest route $F(t+1) - F(t)$. If $F(t+1) - F(t) = 0$, then there is no clear trend in the data.

into account when making a route choice, then the minimisation rate can be interpreted as the proportion of perfect rational choices, assuming a perfect information scenario. This is the case in this study, where the participants received travel time information on each of the alternative routes. Denote as $F_{(k)}$ the proportion of times that the k-th fastest route informed to participants route was chosen, then $F_{(1)}$ is the minimisation rate. Although the proportions $F_{(k)}$ are defined for the general case, i.e., for choice problems with more than 3 alternative routes, in this study $k = 1, 2, 3$. The proportions $F_{(k)}$, can be computed globally, at OD pair level, at route level, and at participant level. Computed at global level, $F_{(k)}$ allows to make general conclusions about the travel time minimisation behaviour and how this relates to the differences in travel time between the alternative routes. At OD pair and route level, $F_{(k)}$ allows to investigate if the minimisation behaviour is influenced by the characteristics of the routes, other than travel time. Finally, heterogeneity in participants choices can be observed by computing $F_{(k)}$ at participant level. The quantities $F_{(k)}$ can be formally defined in terms of probability. In this chapter, the terms probability and proportion are used interchangeably.

Let R_{od} be the set of alternative routes belonging to the OD pair od with J alternative routes. Define C as a random variable taking the value $C = j$ $(j = 1, 2, \ldots, J)$, if the route $r_j \in R_{od}$ is chosen, and $I_{(k)}$ as the random variable taking the value $I_{(k)} = j$ $(j = 1, 2, \ldots, J)$, when route $r_j \in R_{od}$ is informed to participants to be the k-th fastest route. Then, the probability that the route r_j is chosen, given that it was the k-th fastest amongst the J alternatives in the OD pair od, is given by

$$F_{(k)}^{j,od} = Pr(C = j \mid I_{(k)} = j, OD = od). \tag{5.1}$$

At OD pair level, the proportion of times that the k-th fastest route was chosen, $F_{(k)}^{od}$, is obtained by integrating the expression in Eq (5.1) over all the routes $r_j \in R_{od}$, i.e.,

$$\begin{aligned} F^{od}_{(k)} &= \sum_{j \,|\, r_j \in R_{od}} Pr(C = j \,|\, I_{(k)} = j, OD = od) \times Pr(I_{(k)} = j \,|\, OD = od) \\ &= \sum_{j \,|\, r_j \in R_{od}} Pr(C = j, I_{(k)} = j \,|\, OD = od) \\ &= Pr(C = I_{(k)} \,|\, OD = od). \end{aligned} \tag{5.2}$$

Likewise, the global proportion of times that the k-th fastest route was chosen, $F_{(k)}$, is obtained by integrating the expression in Eq (5.2) over all OD pairs, this is

$$\begin{aligned} F_{(k)} &= \sum_{od} Pr(C = I_{(k)} \,|\, OD = od) \times Pr(OD = od) \\ &= Pr(C = I_{(k)}). \end{aligned} \tag{5.3}$$

The proportions $F_{(k)}$ by individual i are obtained in a similar fashion by conditioning by individual instead of the OD pairs, i.e., $F^i_{(k)} = Pr(C = I_{(k)} \,|\, i)$.

5.3.2 Travel time boundedly rational behaviour

Remember that under boundedly rational behaviour, a traveller is indifferent to the travel time of the alternatives when their difference is under a threshold (indifference band). The set of alternatives under this condition are called *satisficing*. Analogous to the perfect rational behaviour, the proportion of times that participants chose a satisficing route is computed. Note that the above definition does not mean that the traveller is indifferent to the satisficing routes, in the sense that she or he will choose any of them with the same probability. Rather, the definition means that the effect of travel time is negligible among the satisficing routes. This last interpretation allows for other attributes to play a role in the choices of travellers. Thus, under boundedly rational behaviour, travellers do not necessarily choose the shortest travel time route, but a satisficing route. The indifference band in this study is defined relative to the fastest route.

Let $ITT_{(j)}$ and $ITT_{(k)}$ be the travel time information in the j-th and k-th fastest routes in a choice problem, such that $ITT_{(j)} \leq ITT_{(k)}$. The difference in travel time information can be computed in absolute (time) or relative (percentage) terms as $\Delta ITT_{j,k} = ITT_{(k)} - ITT_{(j)}$, and $\%\Delta ITT_{j,k} = (ITT_{(k)} - ITT_{(j)})/ITT_{(j)}$, respectively. For ease of exposition, in the rest of this section, the differences in the travel time information will be denoted as $\Delta ITT_{j,k}$ to refer to either the absolute or the relative difference. Contrary to the minimisation rate, where each choice problem has a minimum travel time route, in bounded rationality a choice problem may have one, two or more satisficing routes. This implies that for some choice problems the probability of choosing the fastest route needs to be estimated, for other choice problems the probability of choosing the fastest or second fastest, and so on. These probabilities can be written as the conditional probability of choosing a satisficing route, given that there are n satisficing routes. Formally, for a given indifference band $IB(\alpha) = [0, \alpha]$, define the set $S_n(\alpha)$ as the set of choice problems with exactly n satisficing alternative routes. If there are N alternative routes in the choice problems, these sets are

$$\begin{aligned}
S_1(\alpha) &= \{\text{choice problem} \,|\, \Delta ITT_{1,2} > \alpha\} \\
S_2(\alpha) &= \{\text{choice problem} \,|\, \Delta ITT_{1,2} \leq \alpha \wedge \Delta ITT_{1,j} > \alpha \,\forall\, j > 2\} \\
&\dots \\
S_n(\alpha) &= \{\text{choice problem} \,|\, \Delta ITT_{1,n} \leq \alpha \wedge \Delta ITT_{1,j} > \alpha \,\forall\, j > n\} \\
&\dots \\
S_N(\alpha) &= \{\text{choice problem} \,|\, \Delta ITT_{1,N} \leq \alpha\}\,.
\end{aligned} \tag{5.4}$$

Using the same notation as in the previous section, the conditional probabilities of choosing a satisficing route, given that there are n satisficing routes, are then

$$\begin{aligned}
Pr(\text{satisficing}|S_n(\alpha)) &= Pr(C \in \bigcup_{k=1}^{n} \{I_{(k)}\} \,|\, S_n(\alpha)) \\
&= \sum_{k=1}^{n} Pr(C = I_{(k)} \,|\, S_n(\alpha))\,,
\end{aligned} \tag{5.5}$$

where the last equality is because the events $C = I_{(j)}$ and $C = I_{(k)}$ are disjoint for $j \neq k$. Finally, the total probability of choosing a satisficing route can be obtained as

$$Pr(\text{satisficing}|\alpha) = \sum_{n=1}^{N} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)).$$

The probabilities $Pr(\text{satisficing}|S_n(\alpha))$ and $Pr(\text{satisficing}|\alpha)$ are also estimated for subsamples of the data to create some variation. In total, 141 subsamples are obtained: 41 by removing one OD pair at a time, and 100 by randomly selecting 70% of the observations with repetition. This sampling strategy (bootstrap) allows to observe the effect that heterogeneous participants and route attributes may have on the estimates of the probabilities.

Note that $Pr(\text{satisficing}|S_1(\alpha))$ is *the proportion of times that the fastest route is chosen, given that the difference in the travel time information between the fastest and the rest of the alternatives is more than* α. Since $S_1(\alpha)$ is equivalent to the case when only one route is satisficing (the fastest route), $Pr(\text{satisficing}|S_1(\alpha))$ can also be interpreted as the proportion of perfect rational choices. The analysis in Section 5.4.2 is based on the probabilities $Pr(\text{satisficing}|S_1(\alpha))$ for different values of α, with special interest in comparing the results between the subsets S_1 when defined in terms of absolute or relative time differences. In Section 5.4.3, the perfect rationality behaviour is relaxed by considering the cases in which more than one route is satisficing.

5.3.3 Estimation of indifference band by participant

Until now, the bounded rationality has been studied for hypothetical values of α. Moreover, the values of α have been considered to be equal for all the travellers. However, this assumption does not hold (see Section 5.4.4), meaning that travellers are heterogeneous with respect to their indifference bands. To estimate the indifference band of individual i, the travel time differences of the routes chosen by i are considered. Formally, let $C_{i,m}$ represent the chosen route by individual in choice problem m, then by assuming that all the participants always choose a satisficing route, the indifference band of each individual i can be estimated as

$$\hat{\alpha}_i{}^{max} = \max\{\Delta ITT_{1,k} \,|\, C_{i,m} = k,\, \forall m\}\,.$$

Nevertheless, this definition is restrictive, in the sense that not all information on the travel time differences of the chosen routes is used. For example, a participant that chose a route k with $\Delta ITT_{1,k} = 2$ and the fastest route in the choice problems $m = 2, \ldots, M_i$ will have $\hat{\alpha}_i{}^{max} = 2$, without considering that $\Delta ITT_{1,k'} = 0$ in for all $k' \neq k$. Therefore, two other estimators for α_i are considered in this study: the 95 percentile and the median of the distribution of $\Delta ITT_{1,k} \,|\, C_{i,m} = k$, with $m = 1, \ldots, M_i$. Respectively,

$$\begin{aligned} Pr(\Delta ITT_{1,k} < \hat{\alpha}_i{}^{95} \,|\, C_{i,m} = k,\, \forall m) &= 0.95 \\ Pr(\Delta ITT_{1,k} < \hat{\alpha}_i{}^{50} \,|\, C_{i,m} = k,\, \forall m) &= 0.50\,. \end{aligned}$$

5.3.4 A MXL model for route choice conditioned on the indifference band

The previous sections introduced a methodology to compute the probability of choosing a satisficing route as a function of the indifference band. That methodology allows to make general conclusions about the perfect and bounded rational behaviour of travellers, and assumes no route choice model. Nevertheless, the probability of choosing a specific route is not known. To fill-in this gap, a discrete choice model, specifically, a Mixed Logit Model (MXL) for panel data is estimated. The model is specified considering the indifference band α_i as an input. α_i is exogenous to the model, and it determines which routes are part of the satisficing set. The routes that do not belong to the satisficing set have a probability equal to zero of being chosen.

Let $y_{ij} = 1$ be the event of of individual i choosing route j ($y_{ij} = 0$ otherwise) and α_i be its exogenous determined indifference band. For ease of exposition, the subscript for the repeated choices of individuals is eliminated from the notation. Then, the conditional probability of choosing the alternative j can be written as

$$Pr(y_{ij} = 1 \mid \beta_i, \alpha_i, x_{i\cdot}) = \begin{cases} \frac{exp(x_{ij}^T \beta_i)}{\sum_{k \in S(\alpha_i)} exp(x_{ik}^T \beta_i)}, & \% \Delta ITT_{1,j} \leq \alpha_i \\ 0, \quad \text{otherwise}, \end{cases} \tag{5.6}$$

where $x_{i\cdot}$ is the vector of observed attributes of the individual i and routes in the choice problem, β_i is the vector of coefficients specific to the individual, and $S_n(\alpha_i)$ is the set of satisficing routes for the indifference band α_i. The probability in Eq. (5.6) is conditioned by the vector of coefficients β_i, which represents the preferences or tastes of individual i for the the different attributes x_{ij}. MXL models assume that these preferences are drawn from a probability distribution representing the taste heterogeneity between the individuals. Here, it is assumed that $\beta_i \sim \mathcal{N}(\bar{\beta}, \Sigma)$, allowing correlation between preferences. The unconditional choice probability is given by the multiple integral

$$Pr(y_{ij} = 1 \mid \alpha_i, x_{i\cdot}; \bar{\beta}, \Sigma) = \int_{\Omega_\beta} Pr(C_i = j \mid \beta, \alpha_i, x_{ij}) \times Pr(\beta \mid \bar{\beta}, \Sigma) d\beta \,, \tag{5.7}$$

where $\bar{\beta}$ and Σ are the parameters defining the distribution of individuals' preferences which need to be estimated.

The above model is a two step process: (1) the individual i conforms the set $S_n(\alpha_i)$ by discarding the not satisficing alternatives from a larger set and, then, (ii) he/she chooses an alternative from $S_n(\alpha_i)$. In the interpretation given here, the first step regards a boundedly rational process, and the second a rational process since the probability in Eq. (5.6) is given by a MXL model. Note that for a perfect rational traveller $\alpha_i = 0$, which implies that the fastest route is always chosen. It is important to mention that, contrary to random utility models where all the alternatives have a nonzero probability, in Eq. (5.6) it is possible to assign a zero probability to an alternative in the data that was actually chosen. In other words, if the chosen alternative results to be outside of the indifference band, then it will not be considered in the satisficing set, which violates one of the assumptions of RUMs: the chosen alternative must be part of the chosen set. To avoid this issue, the estimator of the indifference band used as an input is $\hat{\alpha}_i^{max}$, which guarantees that all the chosen routes in the data belong to the satisficing set. The variables that enter the model are

- FRW_j: the % of freeway that composes the routes;
- DIR_j: the directness of the trip, defined as the euclidean distance divided by the length of the route;
- TRN_j: the number of turns per kilometre;

- INT_j: the number of intersections per kilometre;
- $\%\Delta ITT_{i,(1,j)}$: relative travel time difference between route j and the fastest route; the subindex i indicates that is participant i who received the information.

To estimate the MXL model in this study a Bayesian approach was adopted. In Bayesian inference, a posterior distribution for the parameters $\bar{\beta}$ and Σ is obtained after updating the prior distribution through the likelihood function using the Bayes theorem (see Gelman et al. (2014) for details on Bayesian methods or Train (2001) for Bayesian methods applied to choice modelling). This contrasts with maximum likelihood estimation methods, where point estimates for the parameters are found. The *posterior probability distribution*, denoted as $Pr_{post}(\bar{\beta}, \Sigma)$, has no closed form. However, the Bayesian methods, such as Markov Chain Monte Carlo and Gibbs samplers (Levin and Peres, 2017), allow to obtain samples from this distribution. Estimating the model in equation (5.7) with Bayesian methods has the advantage of providing the means to predict the choices of individuals for which the indifference band has not been observed. To see this, let $\tilde{x_{i\cdot}}$ be the measurable attributes of a new individual and the alternatives in a future choice problem. The *posterior predictive choice probability* for the model in expression (5.7) is given by

$$\begin{aligned}
Pr_{pred}(\tilde{y_{ij}} = 1 \,|\, \tilde{x_{i\cdot}}) &= \int Pr(y_{ij} = 1 \,|\, \alpha_i, x_{i\cdot}; \bar{\beta}, \Sigma) \times Pr_{post}(\bar{\beta}, \Sigma) d(\bar{\beta}, \Sigma) \\
&\approx \frac{1}{S} \sum_{s=1}^{S} Pr(y_{ij} = 1 \,|\, \alpha_i, x_{i\cdot}; (\bar{\beta}, \Sigma)_s) \\
&= \frac{1}{S} \sum_{s=1}^{S} \int_{\Omega_\beta} Pr(y_{ij} = 1 \,|\, \beta, \alpha_i, x_{ij}) \times Pr(\beta \,|\, (\bar{\beta}, \Sigma)_s) d\beta \\
&\approx \frac{1}{S} \sum_{s=1}^{S} \frac{1}{I} \sum_{i=1}^{I} Pr(y_{ij} = 1 \,|\, \beta_i^s, \alpha_i, x_{ij}) \,,
\end{aligned} \tag{5.8}$$

where $(\bar{\beta}, \Sigma)_s$ is a sample from the posterior $Pr_{post}(\bar{\beta}, \Sigma)$, and β_i^s a sample from $Pr(\beta \,|\, (\bar{\beta}, \Sigma)_s)$. The samples β_i^s, for all s, are samples of the posterior distribution of the parameters estimated for the individual i. These samples are readily available after model estimation, the reason is that in Bayesian inference for MXL models the individual coefficients β_i are considered as parameters of the model in order to avoid the integral in expression (5.7), which may cause numerical instabilities and an increase in computational effort (Train, 2001). Note that in the last equality in expression 5.8 the indifference bands α_i are treated as individual-specific parameters, and that they are being integrated (along with the coefficients β_i) across individuals. The posterior predictive, $Pr_{pred}(\tilde{y_{ij}} = 1 \,|\, \tilde{x_{i\cdot}})$, is therefore the average of the predicted choices of the individuals in the training set, where each individual has its own indifference band.

The *log pointwise predictive density* (*lppd*), a measure of goodness-of-fit of a model, is the Bayesian analogous of the log-likelihood, and it is obtained by considering the joint posterior predictive probability. Let $\mathcal{D}$ be a set of observations (they can be future or observed choices). Then, assuming independent observations, the *lppd* is given by

$$\begin{aligned} lppd &= logPr^{pred}(\mathcal{D}) \\ &= log \prod_{i} \prod_{k} \left[Pr_{pred}(y_{ik} = 1 \mid x_{i\cdot})\right]^{y_{ik}} \\ &= \sum_{i} \sum_{k} \left[y_{ik} \times logPr_{pred}(y_{ik} = 1 \mid x_{i\cdot})\right] \;, \end{aligned} \tag{5.9}$$

where $(y_{ik}, x_{i\cdot})$ are observed and they can be in or out-of-sample. When the elements in $\mathcal{D}$ are the same used to fit the model, then *lppd* is a measure of goodness-of-fit. When the elements in $\mathcal{D}$ are out-of-sample observations, then *lppd* is a measure of predictive error. Although the *lppd* does not give an absolute scale to evaluate a model, it can be compared between different models, and therefore used for model selection. Higher values of *lppd* are desirable. Observe that only the actual chosen alternatives ($y_{ik} = 1$) contribute in the computation of the *lppd*. In some applications, however, it is of interest to evaluate the aggregated predicted probability for each alternative, i.e.,

$$\bar{Pr}_{pred}(j) = \frac{1}{I} \sum_{i=1}^{I} Pr_{pred}(y_{ij} = 1 \mid x_{i\cdot}) \,. \tag{5.10}$$

This is the case in route choice, in which the aggregated choice probability represents the predicted usage of the routes. Therefore, second way to test the model's performance is to measure the discrepancy between the observed and predicted choice distributions. The following measure is proposed to measure this discrepancy

$$err(Pr_{obs}, \bar{Pr}_{pred}) = \sum_{j=1}^{J} max(0, Pr_{obs}(j) - \bar{Pr}_{pred}(j)) \,, \tag{5.11}$$

where J is the number of alternatives. Fig. 5.4 shows an example of how *err* is computed. An advantage of this definition is that the error is in probability units, for example, $err = 0.1$ means that 10% of the choices will be erroneous on average. Note that in this study there are 41 OD pairs, therefore *err* needs to be computed separately for each OD pair and then averaged to obtain the global error.

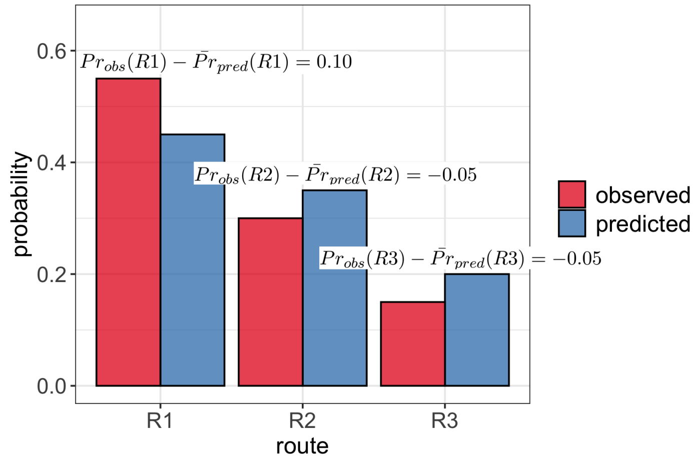

Figure 5.4: Example of the computation of err for a given OD pair. In this case, $err(Pr_{obs}, \bar{P}r_{pred}) = 0.10$.

5.4 Results

5.4.1 Perfect rationality

The distribution of the choices of the participants amongst the fastest, second fastest and slow routes is presented in Table 5.1. In the results of the route choice experiments, there is a clear difference between the choices of participants who receive travel time information (informed participants) and those who did not. The difference between the informed participants and those who did not received information is confirmed by a χ^2 test, rejecting the null hypothesis (with a significance level of α =0.001) that the observed distributions are the same. It is worth mentioning that participants were told that the travel time information are estimates, and that they do not have any other means of learning the travel time on the alternative routes. Therefore, it can be concluded that information has an impact on the choices of the participants, and that the most preferred routes (in the case of no information) differ from those with information, meaning that travel time information is not necessarily aligned to the preferred route attributes. The most notable difference between these distributions is for $F_{(1)}$, thus, it can be concluded that travel time information causes a minimisation behaviour in participants. It is interesting to note that $F_{(1)} > F_{(2)} > F_{(3)}$ in the not informed case, suggesting that participants are somehow minimising the travel time by choosing the routes with the characteristics that they believe lead to smaller travel times. A second observation is that informed participants preferred slower routes in nearly 40% of the cases, meaning that they are not necessarily strict travel time minimisers. Could this behaviour be explained by bounded rationality? Are there factors other than travel time influencing the choices of the participants? These questions will be further investigated.

Table 5.1: Global percentage of times that the fastest, second fastest and third fastest routes were chosen.

Travel time info.	$F_{(1)}$	$F_{(2)}$	$F_{(3)}$
No	0.460	0.328	0.212
Yes	0.605	0.236	0.159
$\chi^2 = 106$, df = 2, p-value < 2.2e-16			

The minimisation rate in each of the OD pairs, $F^{od}_{(1)}$, is presented in Fig 5.5(a), where it can be seen that $F^{od}_{(1)}$ shows a high variability, with values of $F^{od}_{(1)}$ ranging between 0.27 and 0.92. In the case of the three OD pairs with the largest minimisation rates, the alternative with the high composition of freeway was almost always the fastest according to the information given to participants. On the contrary, in the OD pair with the lowest minimisation rate, either the alternative with high composition of freeway was not the fastest, or the three alternative routes were similar in their attributes. This can be observed in 2.A and 2.B; a formal analysis is presented in Section 5.4.5. The (weighted) mean of $F^{od}_{(1)}$ is equal to the global minimisation rate, i.e., $F_{(1)} = 0.605$, with a standard deviation of 0.16. The distribution of the minimisation rate at participant level is included in Fig 5.5(b), where it can be seen that the travel time given to participants does not have the same effect on all individuals. The group of perfect rational participants, who chose the fastest route in all of the choice problems, is relatively small, representing only 9.5% of the total number of participants. Moreover, the minimisation rate of participants is highly heterogeneous, showing a more or less even distribution between minimisation rates of 0.20 and 1.0. The mean of the minimisation rate by participant is 0.58, with a standard deviation of 0.24. This clearly highlights that the great majority of travellers do not make perfect rational decisions in all the choice problems they face, even when travel time estimates are available, thus suggesting a boundedly rational behaviour in route choice. As it will be shown later in this chapter, the variability in $F_{(1)}$ comes, primarily, from the travel time information in the alternative routes, and secondly, from the route attributes that make a route more attractive to the travellers.

5.4.2 Perfect rationality and differences in the travel time information

Travellers are not necessarily travel time minimisers, but how does the difference in travel time information in competing routes influence the behaviour of travellers? Do travellers value absolute or relative differences in travel time? The distributions of the absolute differences in travel time information between the two fastest routes, ΔITT, and the relative differences $\%\Delta ITT$ are shown in Fig 5.6.

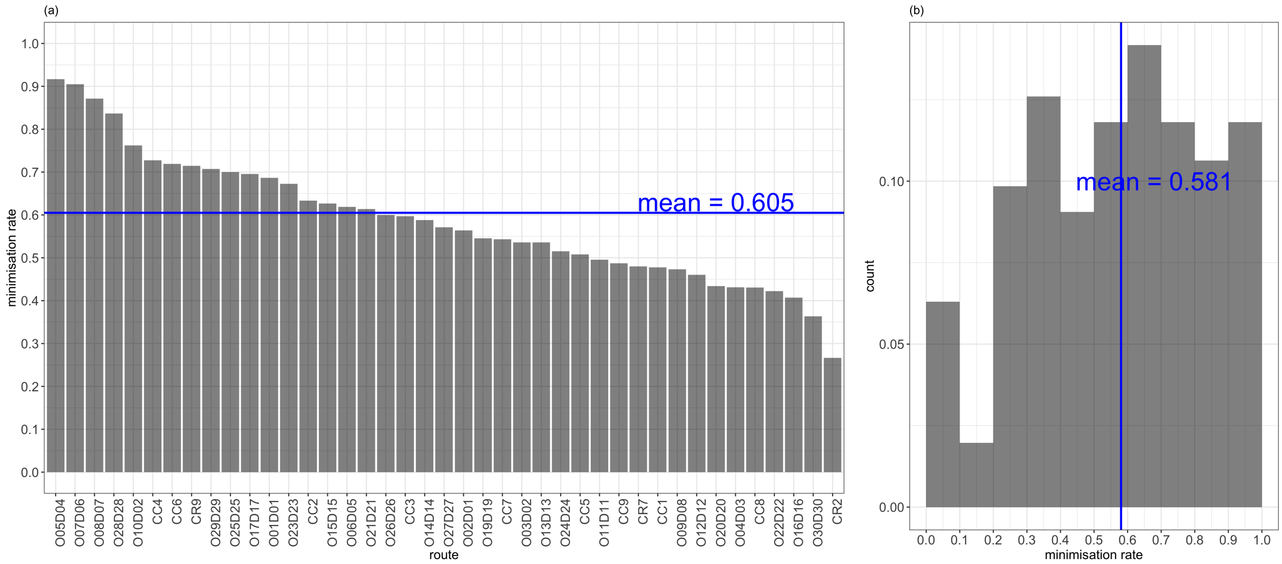

Figure 5.5: (a) minimisation rates per OD pair, and (b) distribution of the minimisation rates of participants.

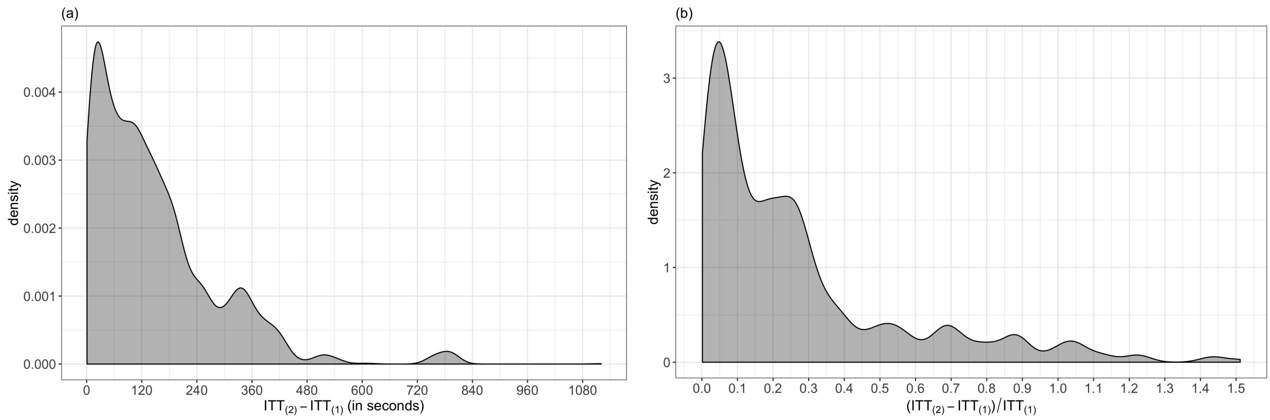

Figure 5.6: Kernel density estimation of the distribution of (a) the absolute (seconds) and (b) the relative differences in travel time.

The proportion of times that the k-th fastest route was chosen in each route choice problem, $F_{(k)}$, is obtained for the subsets $S_1(\alpha)$, defined in Eq. (5.4), with α being the 20-quantiles of the distributions. Remember that the sets $S_1(\alpha)$ are those in which the difference between the fastest and second fastest route is at least α. The minimisation rate, $F_{(1)}$, and the proportions $F_{(2)}$ and $F_{(3)}$, estimated for the different subsets $S_1(\alpha)$, are shown in Fig 5.7, where it can be noticed that, in general, the larger the difference in travel time information, the larger the minimisation rate for both $\Delta ITT_{1,2}$ and $\%\Delta ITT_{1,2}$. This result is not surprising; travel time is recognised as the most important variable in route choice. However, there are important differences between the minimisation rate when considering $\Delta ITT_{1,2}$ or $\%\Delta ITT_{1,2}$. The first difference is that in the case of $\Delta ITT_{1,2}$, the maximum minimisation rate is barely above 0.75, whereas for $\%\Delta ITT_{1,2}$ it can reach values little higher than 0.90. The second difference is that for $\Delta ITT_{1,2}$ the minimisation rate does not show an increasing trend, having a high decrease in $F_{(1)}$ for large values of $\Delta ITT_{1,2}$. This behaviour can hardly be explained, as one would expect that larger differences in travel time information would induce larger minimisation behaviour: why would I choose the fastest route when it is 2 minutes faster, but not when it is 8 minutes faster? This is not the case $\%\Delta ITT_{1,2}$, where $F_{(1)}$ shows an increasing trend.

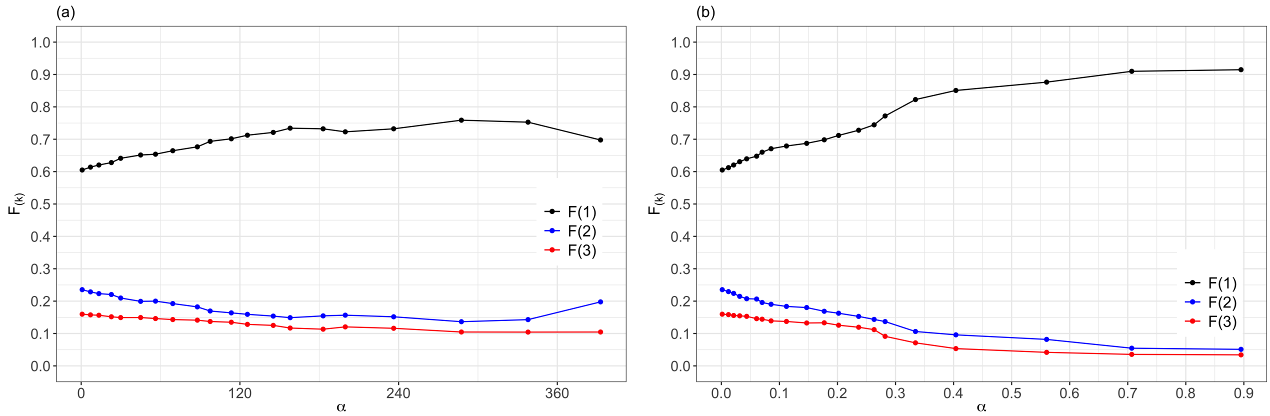

Figure 5.7: Proportion of times the fastest, $F_{(1)}$, second fastest, $F_{(2)}$ and slowest, $F_{(3)}$ routes are chosen, computed for the subsets $S(p)$ for (a) the absolute difference in travel time information, and (b) the relative difference in travel time information. Note that the number of observations to compute the values $F_{(1)}$ decreases with increasing α. In order to maintain the comparability between the proportions between the two definitions, the cut points were determined by the 20-quantiles of each distributions (Figure 5.6), that is why they are not equally distanced in the plots. This assures that the number of observations to compute the proportions are the same for the absolute and relative cases.

To formalise the above findings, two logistic regressions are fitted to the data. In both regressions, the response variable Y is binary, taking the value $Y_i = 1$ if the participant chose the fastest route according to the given information; the regressors are $\Delta ITT_{1,2}$ or $\%\Delta ITT_{1,2}$. The results of the logistic regressions are presented in Table 5.2, and their predictions for $F_{(1)}$ for the subsets $S_1(\alpha)$ are plotted along the observed values of $F_{(1)}$ in Fig 5.8. In both cases, the regressors are different from zero, with a significance level of 0.001, however, the Akaike information criterion (AIC) is smaller when using $\%\Delta ITT_{1,2}$ as a regressor, meaning a better fit with this explanatory variable. This result is confirmed by the Hosmer and Lemeshow goodness of fit test (Hosmer and Lemesbow, 1980). The null hypothesis H_0 of this test is that the observed proportions are similar to the predicted proportions in different subsets of the data. H_0 is rejected for $\Delta ITT_{1,2}$, but not for $\%\Delta ITT_{1,2}$, suggesting a *good* fit of the regression with $\%\Delta ITT_{1,2}$.

The above results show that the relative differences in travel time information, $\%\Delta ITT_{1,2}$, explain better the minimisation behaviour of the participants, and therefore their route choice

Table 5.2: Summary of the logistic regressions with dependent variable $Y = 1$ when the fastest route was chosen, and regressors $\Delta ITT_{1,2}$ or $\%\Delta ITT_{1,2}$. The Hosmer and Lemeshow (H&L) goodness of fit test is included in the table.

	Absolute difference			Relative difference		
Coefficient	Estimate (s.e.)	z statistic	$Pr(>\|z\|)$	Estimate (s.e.)	z statistic	$Pr(>\|z\|)$
intercept	-0.0305 (0.0517)	-0.59	0.555	-0.1556 (0.0484)	-3.218	0.0013
$\Delta ITT_{1,2}$	0.0033 (0.0003)	11.02	$< 2e-16$			
$\%\Delta ITT_{1,2}$				2.4528 (0.1629)	15.058	$< 2e-16$
	Deviance=4,774.9 AIC=4,778.9			Deviance=4,614.395 AIC=4,618.4		
H &L GOF	χ^2=79.802, df=18, p-value=9.282e-10			χ^2=25.444, df=18, p-value=0.1132		

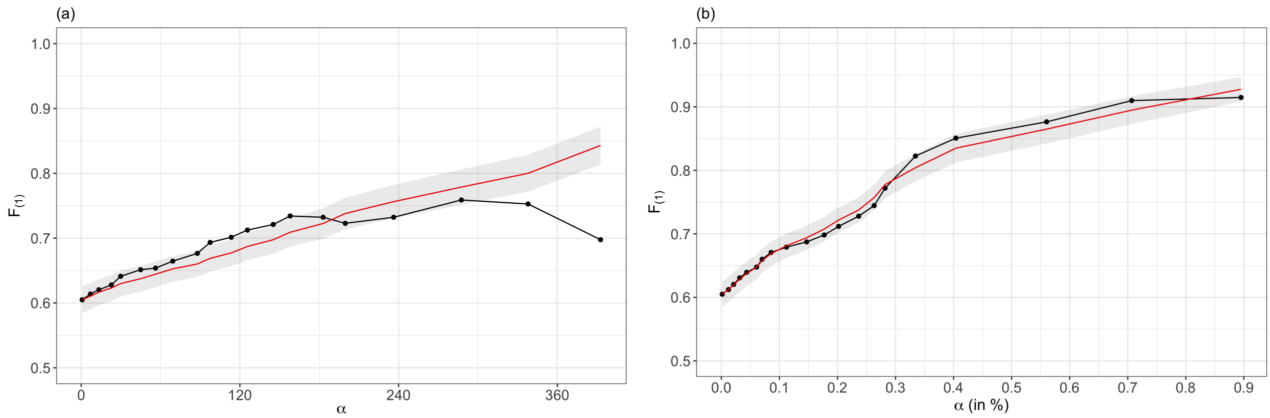

Figure 5.8: Minimisation rate, $F_{(1)}$, computed for the subsets $S_1(\alpha)$ for (a) the absolute difference in travel time information, and (b) the relative difference in travel time information. The red lines are the predictions of the logistic models.

behaviour. This suggests that travellers evaluate the travel time in a problem-wise manner, i.e., relative to the travel time in the competing alternatives, and not as an absolute difference in time units.

5.4.3 Bounded rationality in route choice

In the previous section, the analysis was based on subsets $S_1(\alpha)$, i.e., the perfect rational behaviour. In this section, the bounded rationality of travellers is studied, so the cases when two or more routes are satisficing are analysed. First, the probability of choosing a satisficing route is estimated for different values of α for the case of choice problems with three alternative routes. Then, lower and upper bounds are derived for the general case when there are more than three alternative routes. In view of the above results, the analyses in this section are restricted to the relative differences $\%\Delta ITT$. As in the previous section, α is given by the 20-quantiles of the distribution of the travel time differences $\%\Delta ITT_{1,2}$.

The probability of choosing a satisficing route for the different values of α, $Pr(\text{satisficing}|\alpha)$, is shown in Fig 5.9(a), along with the conditional probabilities $Pr(\text{satisficing}|S_n(\alpha))$, $n = 1, 2, 3$. The fraction of the data that each set $S_n(\alpha)$ represents, i.e., $Pr(S_n(\alpha))$, is presented in Fig 5.9(b). In this last figure, it can be seen that the fraction of the choice problems in which there is only one satisficing route, $Pr(S_1(\alpha))$, decreases with α, while the fraction of problems with three satisficing routes, $Pr(S_3(\alpha))$, increases. This behaviour is expected, as larger indifference bands imply more satisficing alternatives. In Fig 5.9(a), it can be observed that, in general, the probability of choosing a satisficing route $Pr(\text{satisficing}|S_n(\alpha))$ increases with α, and that $Pr(\text{satisficing}|S_2(\alpha)) \geq Pr(\text{satisficing}|S_1(\alpha))$. These trends have a different cause. In the first case, the satisficing alternatives become more desirable as a consequence of a larger difference in the travel time information between the fastest route (which is always satisficing) and the not satisficing routes, i.e, larger values of α. In the second case, the trend is explained because, for the same value of α, the number of satisficing routes is larger in $S_2(\alpha)$ than in $S_1(\alpha)$, thus making it more likely to choose one. These two observations can be generalised to the case of choice problems with more than three alternative routes. In the first case, by arguing that the travel time has a negative effect on the choices of travellers, i.e., the larger the travel time differences between the routes the more likely the faster routes are chosen. In the second case, by arguing that more satisficing alternatives imply necessarily less non-satisficing routes, so the probability of choosing a satisficing route is higher for larger values of n. Furthermore, more satisficing routes means a greater diversity of the route attributes, giving travellers more options from where to choose.

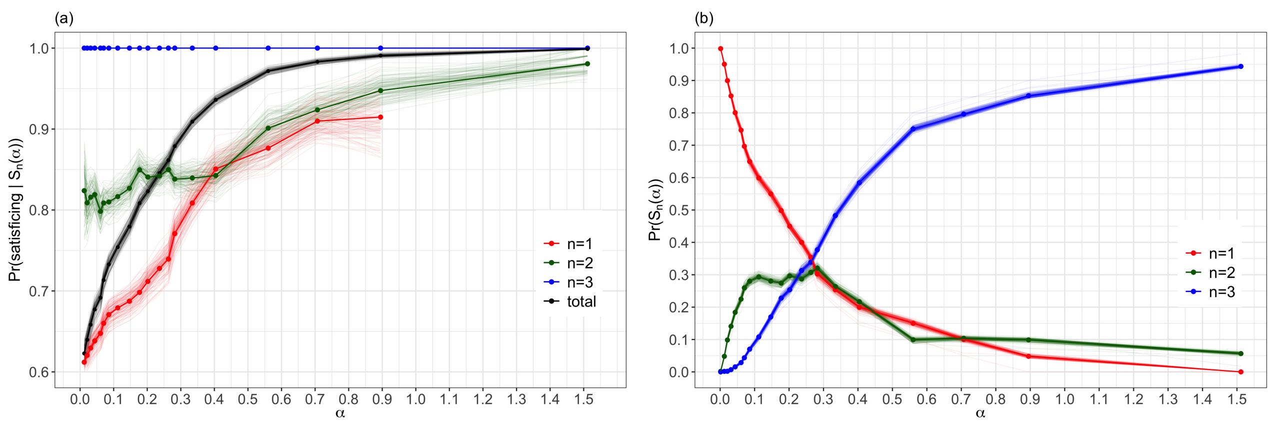

Figure 5.9: (a) Conditional, $Pr(\text{satisficing}|S_n(\alpha))$, and unconditional, $Pr(\text{satisficing}\,|\,\alpha)$, probability of choosing a satisficing route as a function of α. (b) Probability of observing $S_n(\alpha)$ in the data, i.e, the fraction of observations with $n = 1, 2, 3$ satisficing routes for different values of α. These probabilities, computed for the bootstrap subsamples, are also included in the figures with lighter colours. The values α where the probabilities are computed correspond to the 20-quantiles of the relative travel time difference distribution (Figure 5.6(b)).

The probability $Pr(\text{satisficing}|\alpha)$ shown in Fig 5.9(a), was estimated for choice problems with three alternative routes, but how would it look in the general case, i.e., for more than three alternatives? To answer this question, first note that since there are only three alternative routes in the choice problems, the probability $Pr(\text{satisficing}|S_3(\alpha)) = 1$ for all α. Assuming that the probabilities $Pr(\text{satisficing}|S_n(\alpha))$ and their weights $Pr(S_n(\alpha))$ ($n = 1$ and $n = 2$) in the general case can be estimated from the case of three alternatives, then the total probability estimated here, $\hat{Pr}(satisficing|\alpha)$, overestimates the *real* total probability that would be observed in the presence of more than three alternative routes. Therefore, $\hat{Pr}(satisficing|\alpha)$ can be considered as an upper bound for this *real* probability. To see this,

$$\begin{aligned}
Pr(\text{satisficing}|\alpha) &= \sum_{n=1}^{N} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)) \\
&\leq \sum_{n=1}^{2} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)) + \sum_{n=3}^{N} Pr(S_n(\alpha)) \\
&= \sum_{n=1}^{2} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)) + (1 - \sum_{n=1}^{2} Pr(S_n(\alpha))) \\
&\approx \sum_{n=1}^{2} \hat{Pr}(satisficing|S_n(\alpha)) \times \hat{Pr}(S_n(\alpha)) + \hat{Pr}(S_3(\alpha))\,,
\end{aligned} \tag{5.12}$$

where the inequality is obtained by making $Pr(\text{satisficing}|S_n(\alpha)) = 1$ for all $n \geq 3$, and the equality since $\sum_{n=1}^{N} Pr(S_n(\alpha)) = 1$ (they are disjoint events). To obtain a lower bound, recall from the previous analysis that it can be assumed that $Pr(\text{satisficing}|S_{n+1}(\alpha)) \geq Pr(\text{satisficing}|S_n(\alpha))$ for all n. Thus,

$$\begin{aligned}
Pr(\text{satisficing}|\alpha) &= \sum_{n=1}^{N} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)) \\
&\geq \sum_{n=1}^{2} Pr(\text{satisficing}|S_n(\alpha)) \times Pr(S_n(\alpha)) \\
&\qquad + Pr(\text{satisficing}|S_2(\alpha)) \sum_{n=3}^{N} Pr(S_n(\alpha)) \\
&\approx \sum_{n=1}^{2} \hat{Pr}(satisficing|S_n(\alpha)) \times \hat{Pr}(S_n(\alpha)) \\
&\qquad + \hat{Pr}(satisficing|S_2(\alpha))\hat{Pr}(S_3(\alpha)).
\end{aligned} \tag{5.13}$$

The lower and upper bounds for the probability in the general case are shown in Fig 5.10, along with the estimated conditional probabilities $\hat{Pr}(satisficing|S_n(\alpha))$. The results show that the estimates of the proportion of boundedly rational choices are higher than the estimates for the perfect rational choices, and that the difference is higher for $\alpha < 0.35$. For $\alpha = 0.35$, the estimated proportion of rational choices is 82%, whereas for boundedly rational choices it is between 84% and 92%.

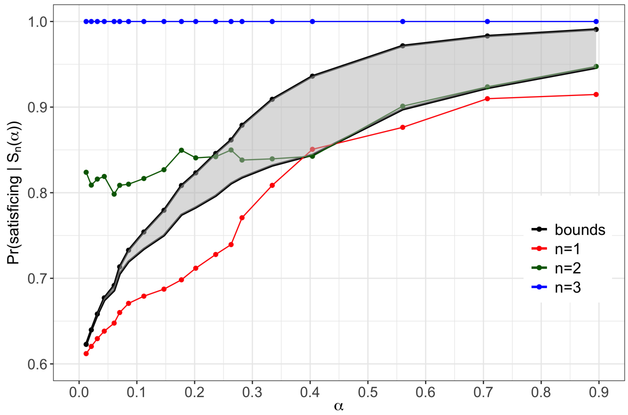

Figure 5.10: Total probability of choosing a satisficing route for the general case (grey area). Perfect rationality is represented by the red line.

In Fig. 5.9(a), it can be seen that the estimates for the bootstrap subsamples do not differ considerably from the estimate considering all data, specially for the unconditional probability $Pr(\text{satisficing}|\alpha)$. This implies that, at aggregated level, the heterogeneity of participants and route attributes have little impact on the probability of choosing a satisficing route. As in the previous section, a logistic regression is fitted to the data to obtain a mathematical expression for the upper and lower bounds in the general case. The regression is fitted to the bootstrap subsamples to produce some variation. The results of the models are summarised in Table 5.3, where it can be seen that, for both cases, the regressor α is statistically significant (significance level 0.001), and that the Hosmer and Lemeshow goodness of fit test do not reject the null hypothesis that the observed and predicted probabilities are the same. Therefore, the upper and lower bounds for the probability of choosing a satisficing route, given the size of the indifference band α can be approximated by the logistic functions

$$Pr(\text{satisficing}\,|\,\alpha)^{upper} = \frac{e^{0.49+5.23\alpha}}{1+e^{0.49+5.23\alpha}}$$
$$Pr(\text{satisficing}\,|\,\alpha)^{lower} = \frac{e^{0.61+2.85\alpha}}{1+e^{0.61+2.85\alpha}}\,.$$

These bounds are shown in Fig. 5.11 along with the observed values of the bootstrap subsamples.

The conditional probabilities $Pr(\text{satisficing}|S_n(\alpha))$ can be decomposed as the sum of the more simple probabilities $Pr(C = I_{(k)}\,|\,S_n(\alpha))$, i.e., the sum of the probabilities of choosing the k-th fastest route, given that there are n satisficing routes (see Eq. (5.5)). This decomposition is shown in Fig. 5.12, where it can be seen that the probability of choosing the fastest route is

Table 5.3: Summary of the logistic regressions to approximate the upper and lower bands for $Pr(\text{satisficing}|S_n(\alpha))$ in the general case. The Hosmer and Lemeshow (H&L) goodness of fit test is included in the table.

	Upper bound			Lower bound		
Coefficient	Estimate (s.e.)	z statistic	$Pr(>\|z\|)$	Estimate (s.e.)	z statistic	$Pr(>\|z\|)$
intercept	0.4943 (0.0676)	7.313	2.62e-13	0.6122 (0.0620)	9.872	< 2e-16
α	5.2326 (0.3962)	13.206	< 2e-16	2.8531 (0.2690)	10.606	< 2e-16
	Deviance=2.0312 AIC=1367.2			Deviance=6.4038 AIC=1578.5		
H &L GOF	χ^2=1.1096, df = 8, p-value = 0.9975			χ^2=4.6106, df = 8, p-value = 0.7983		

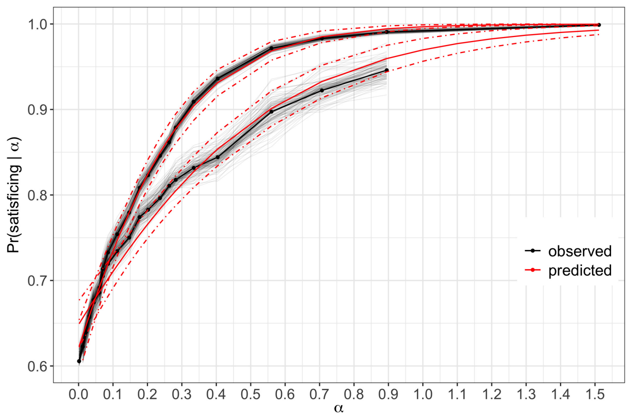

Figure 5.11: Predicted upper and lower bounds for the probability of choosing a satisficing route. The 95% prediction error interval is represented with a dashed line.

higher, no matter the value of α, i.e., $\hat{Pr}(C = I_{(1)} \,|\, S_n(\alpha)) > \hat{Pr}(C = I_{(2)} \,|\, S_n(\alpha))$ for all values of α. As expected, the preference for the fastest route amongst the satisficing routes increases with increasing values of α, however, it is interesting to note that the preference for the fastest route is much higher even for small values of α. This means that informing a route to be the fastest has already an effect on the preferences of the participants, regardless of the difference in the travel time with the rest of the alternatives. This effect is specially important in the case of the perfect rational travellers (9.5% in this study), who will always choose the fastest route. The probability of choosing the fastest route is approximately 29% higher in the case of $S_2(\alpha)$ and 116% higher in the case $S_3(\alpha)$.

5.4.4 Heterogeneity of the indifference band

Participants are heterogeneous in their indifference bands. This can be observed in Fig. 5.10, where $\hat{Pr}(satisficing|\alpha) < 1$, contradicting the boundedly rational hypothesis that travellers choose satisficing routes. To put it another way, if participants were all boundedly rational with the same indifference band given by α^*, then $\hat{Pr}(satisficing|\alpha) = 1$ for all the values of

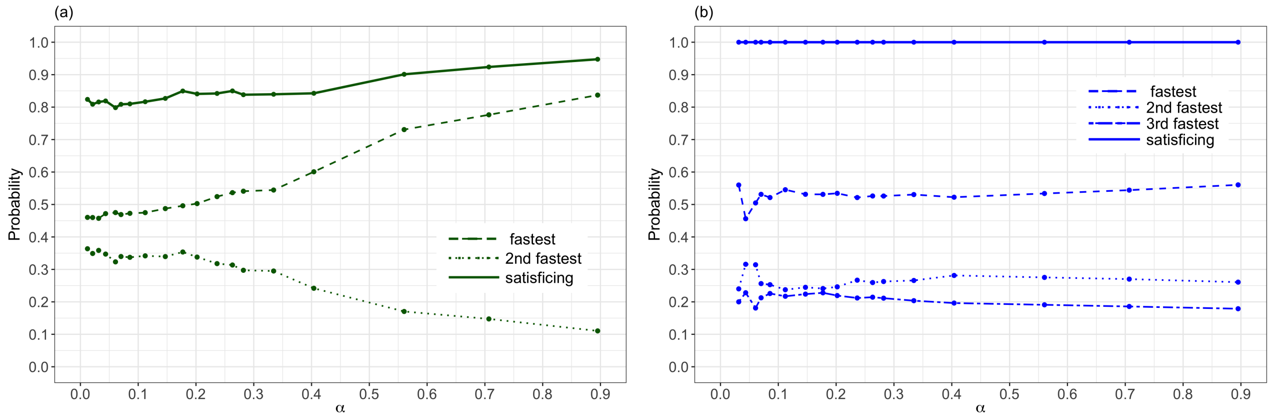

Figure 5.12: Probabilities of choosing the k-th fastest route amongst the n satisficing routes for (a) $n = 2$ and (b) $n = 3$.

$\alpha \geq \alpha^*$. This is clearly not the case, unless a large (and therefore not meaningful) value of α^* is considered. In this section, the heterogeneity of the indifference bands is analysed. To this purpose, the estimators for the indifference band at individual level, α_i, defined in Section 5.3.3, are computed.

The distribution of the estimators $\hat{\alpha}_i{}^{max}$, $\hat{\alpha}_i{}^{95}$ and $\hat{\alpha}_i{}^{50}$ are presented in Fig. 5.13. It can be observed that for the estimators $\hat{\alpha}_i{}^{max}$ and $\hat{\alpha}_i{}^{95}$ the proportion of perfectly rational participants, as it was found in Section 5.4.1, is 9.5% ($\hat{\alpha}_i = 0$). Furthermore, there is a group of participants with large indifference band, $\alpha_i > 1$, meaning that they will still consider routes twice slower than the fastest route. The percentage of participants with these large indifference bands is 5% and 2% for $\hat{\alpha}_i{}^{max}$ and $\hat{\alpha}_i{}^{95}$, respectively. It is likely that these participants were not engaged in the experiments, as it is difficult to believe that a traveller is willing to choose a route twice as slow as the fastest alternative. For these two estimators, a large heterogeneity is observed, with values more or less uniformly distributed in the interval $(0.15, 0.5]$; in both cases, around half of the observations lie in this interval. However, the distribution of $\hat{\alpha}_i{}^{95}$ accumulates around 30% of the observations in the interval $[0, 0.15]$, whereas the distribution of $\hat{\alpha}_i{}^{95}$ accumulates around 20% in the same interval. This explains the difference of 0.08 percentage points in the means of the distributions. In contrast, the distribution of the estimators $\hat{\alpha}_i{}^{50}$ tells a completely different story, showing low heterogeneity with around 80% of the observations having a very small indifference band: $\hat{\alpha}_i{}^{50} \leq 0.10$. Moreover, with this definition, the proportion of perfect rational participants would be 55%, which is high compared to the observed proportion of perfectly rational participants (9.5%). Considering that $\hat{\alpha}_i{}^{max}$ is a restrictive estimator, very sensitive to outliers, and that $\hat{\alpha}_i{}^{50}$ overestimates the perfect rationality, the estimator $\hat{\alpha}_i{}^{95}$ may be a good selection. This latter estimator is conservative, in the sense that it will include the great majority of the routes choices within the indifference band, but being less sensitive to outliers.

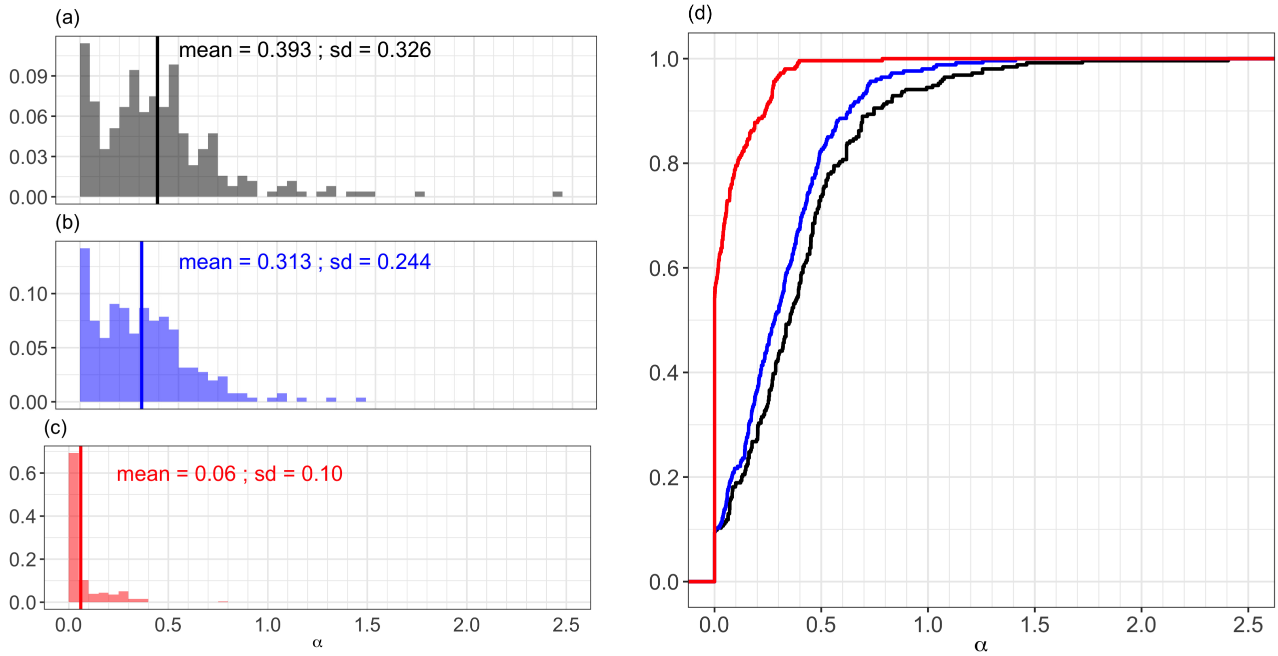

Figure 5.13: Distributions of the indifference band by participant estimated using (a) the maximum of the travel time differences of the routes chosen by the participants, (b) the 95 percentile, and (c) the 50 median. The cumulative distributions are included in (d).

To see the consequences of using the distinct estimators of α_i, the proportion of satisficing choices in the data set are computed assuming that the participants are heterogeneous and that their indifference bands are given by the three estimators. These proportions are presented in Table 5.4. As expected, when the estimators are defined as $\hat{\alpha}_i{}^{max}$, the probability of choosing a satisficing route is 100%, since $\hat{\alpha}_i{}^{max}$ was defined so all the observed choices are satisficing. By relaxing this condition, and defining the estimator as $\hat{\alpha}_i{}^{95}$, the observed probability of choosing a satisficing route is 89.9%, and 66.8% for the estimator $\hat{\alpha}_i{}^{50}$. These proportions are similar to those obtained by assuming that participants are homogeneous with α equal to the means of the distributions. Note that the homogeneous case is equivalent to evaluating $\bar{\alpha}_i$ in Fig. 5.10. For $\hat{\alpha}_i{}^{max}$ and $\hat{\alpha}_i{}^{95}$, this implies that by assuming homogeneity and an indifference band equal to the mean, it is possible to know with high probability (92.7 and 88.9%) which routes are travel time satisficing. However, a smaller indifference band is preferred, as it may reduce the number of alternative routes that are considered.

Table 5.4: Observed proportion of satisficing choices, (i) given different estimators for the individual indifference bands, α_i, and (ii) assuming homogeneity for the indifference band, $\bar{\alpha}_i$.

	$\hat{\alpha}_i{}^{max}$	$\hat{\alpha}_i{}^{95}$	$\hat{\alpha}_i{}^{50}$
heterogeneous α_i	100%	89.9%	66.8%
homogeneous $\alpha = \bar{\alpha}$	92.7%	88.9%	69.1 %

5.4.5 Estimating the MXL model for route conditioned on the indifference band

In this section, the route choice probability is obtained by estimating the route choice model presented in section 5.3.4. Name this model *Model_1*. The results are presented alongside the estimates of a second unrestricted MXL model (*Model_0*) that considers no indifference band. That is, in *Model_0* the three alternative routes are always considered by the decision maker and have a probability of being chosen greater than zero. The purpose of including *Model_0* in the analysis is to investigate how considering the indifference bands change user route choice behaviour. The models are compared in terms of their goodness-of-fit and predictive accuracy for out-of-sample observations at the end of this section. Both models were estimated using only the participants that received travel time information (353 participants and 3,664 choices). It is worth mentioning that, in order to facilitate the interpretation and comparison between the two models, the informed travel time variable is considered in the specification of *Model_1* for the alternatives inside the indifference band. The Gibbs sampler software JAGS (Plummer, 2003) and the R (R Core Team, 2018) package *rjags* were used to obtain samples from the posterior distribution of the parameters β_i, $\bar{\beta}$ and Σ. The values of the hyperparameters, which define the priors of $\bar{\beta}$ and Σ, were chosen to be weakly-informative (very high variances). In other words, it is assumed high uncertainty on the real values of the parameters that are being estimated. The estimates for models *Model_0* and *Model_1* are presented in Table 5.5; the complete summary of the estimates is included in 5.A.

Comparing the two models, it can be seen that there is a large difference in the distribution of the coefficients $\beta_{\%\Delta ITT}$, and that this difference is explained by a change in their mean values $\hat{\bar{\beta}}_{\%\Delta ITT}$ rather than a change in their variance: *Model_1* exhibits a mean closer to zero. Note that the distributions of the rest of the attributes do not vary considerably (Fig. 5.14). *Model_0* has a negative mean preference for travel time information $\hat{\bar{\beta}}^{M0}_{\%\Delta ITT} < 0$, meaning that the average traveller finds longer travel times undesirable. At individual level i, the preferences for $\%\Delta ITT$ show a high heterogeneity, as it can be deduced from the estimated standard deviation ($\hat{\sigma}_{\%\Delta ITT} = 4.138$). The proportion of participants with a negative preference for $\%\Delta ITT$ is $Pr(\beta^{M0}_{\%\Delta ITT} < 0) = 0.21$, i.e., four in five participants prefer shorter time routes. Moving on to *Model_1*. The estimates show a positive mean preference for the travel time information, $\hat{\bar{\beta}}^{M1}_{\%\Delta ITT} > 0$, result that may appear counter intuitive, as it is interpreted as the mean participant choosing longer routes. However, contrary to *Model_0* where 4/5 of participants show a preference for shorter routes, in *Model_1* $Pr(\beta^{M1}_{\%\Delta ITT} < 0) = 0.46$, i.e., there is no clear trend in the preferences for the travel time information. In words, it is equally likely to find an individual preferring shorter time routes than longer ones within the indifference band. This finding is in accordance with the bounded rational model assumption in

Table 5.5: Estimates for the mean and the covariance (standard deviation and correlation) of the parameters of the two MXL models *Model_0* and *Model_1* estimated for the participants that received travel time information. *lppd* is the log pointwise predictive density, an estimate of the predictive accuracy of the model: a higher value (compared to another model) means a better fit. WAIC (Watanabe-Akaike Information Criterion) penalises the *lppd* with the model complexity: a smaller value (compared to another model) means that the model represents a better alternative balancing goodness-of-fit and complexity. *err* is the discrepancy between the observed and predicted choice distributions.

	Model_0		*Model_1*	
Parameter	mean	s. error	mean	s. error
$\hat{\bar{\beta}}_{FRW}$	0.862	0.313	0.622	0.340
$\hat{\bar{\beta}}_{DIR}$	1.377	0.640	1.863	0.753
$\hat{\bar{\beta}}_{TRN}$	0.012	0.108	0.061	0.132
$\hat{\bar{\beta}}_{INT}$	-0.044	0.029	-0.076	0.032
$\hat{\bar{\beta}}_{\%\Delta ITT}$	-3.285	0.356	0.366	0.373
$\hat{\sigma}^2_{FRW}$	2.840 (1.685)	2.253	3.596 (1.896)	1.842
$\hat{\sigma}^2_{DIR}$	17.889 (4.230)	11.229	25.970 (5.096)	9.383
$\hat{\sigma}^2_{TRN}$	0.505 (0.711)	0.185	0.464 (0.681)	0.178
$\hat{\sigma}^2_{INT}$	0.052 (0.228)	0.011	0.058 (0.240)	0.014
$\hat{\sigma}^2_{\%\Delta ITT}$	17.121 (4.138)	3.230	10.867 (3.297)	2.575
$\hat{\sigma}_{FRW,DIR}$	4.660 (0.654)	4.628	7.470 (0.773)	3.564
$\hat{\sigma}_{FRW,TRN}$	0.389 (0.324)	0.478	0.543 (0.420)	0.451
$\hat{\sigma}_{FRW,INT}$	0.011 (0.028)	0.092	-0.031 (-0.069)	0.093
$\hat{\sigma}_{FRW,\%\Delta ITT}$	0.116 (0.017)	1.843	0.882 (0.141)	1.596
$\hat{\sigma}_{DIR,TRN}$	1.175 (0.391)	1.025	1.622 (0.467)	1.057
$\hat{\sigma}_{DIR,INT}$	-0.340 (-0.352)	0.227	-0.486 (-0.397)	0.274
$\hat{\sigma}_{DIR,\%\Delta ITT}$	9.290 (0.531)	3.853	8.525 (0.507)	3.696
$\hat{\sigma}_{TRN,INT}$	-0.029 (-0.180)	0.032	-0.025 (-0.151)	0.035
$\hat{\sigma}_{TRN,\%\Delta ITT}$	-0.530 (-0.180)	0.530	0.021 (0.009)	0.543
$\hat{\sigma}_{INT,\%\Delta ITT}$	-0.336 (-0.356)	0.160	-0.382 (-0.482)	0.150
lppd	-3307.05		-3334.77	
$WAIC = -2lppd + 2p_{waic}$	6,632.27 (p_{waic}=9.08)		6,688.20 (p_{waic}=9.33)	
err	0.110%		0.104	

this chapter: travellers are indifferent to travel time when choosing a route from the satisficing set. Furthermore, the physical attributes play, on average, a larger role in the choices of travellers in *Model_1* compare to *Model_0*. To see this, observe that there is no meaningful change in the distributions of the physical attributes between the models. Then, since the values of $\beta_{\%\Delta ITT}$ are closer to zero in *Model_1*, the importance of the travel time information relative to the rest of the attributes x, measured as $|\hat{\bar{\beta}}_{\%\Delta ITT}/\hat{\bar{\beta}}_x|$, decreases significantly. As in *Model_0*, the distribution of these coefficients show a high variance and, as a consequence, the coefficients $\beta_{\%\Delta ITT}$ may take large values. Nonetheless, the impact of this coefficient on the choice probabilities is lower than in the unrestricted model, as it multiplies smaller values of $\%\Delta ITT_{1,j}$ in the utility; routes with large differences would not be satisficing.

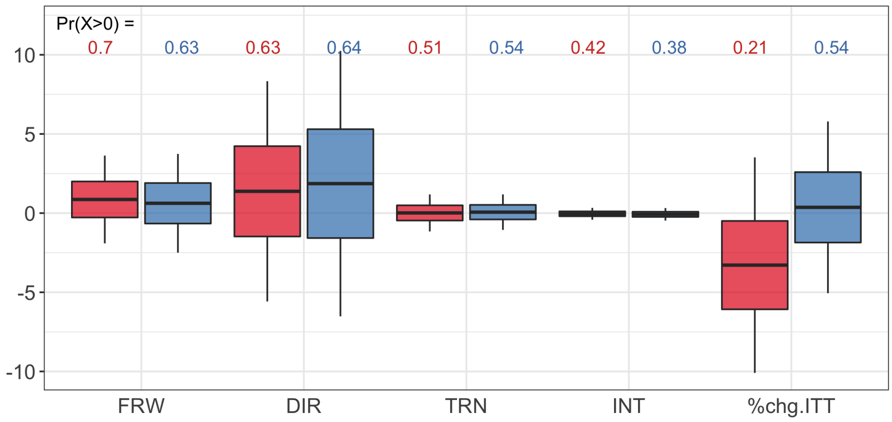

Figure 5.14: Distributions of the random coefficients $\beta \sim \mathcal{N}(\hat{\bar{\beta}}, \hat{\Sigma})$ for models *Model_0* and *Model_1*. The probability of a coefficient being greater than zero in shown at the top of the plot.

Continuing with the interpretation of the rest of the variables. A decrease in the mean preferences for FRW and INT can be noticed. In the case of FRW, the decrease places the average closer to zero, meaning that it becomes less important in the route choice. In the case of INT, the decrease makes it more important, meaning that within the indifference band a participant is less willing to choose a route with more intersections. It is important to mention that even though the coefficient seems small, this variable usually takes values $INT > 5$, making it an important variable defining the choice probabilities (see Fig. 2.9). This last variable is even more important than FRW. This is not the case for TRN, with coefficients near zero and taking values usually $TRN < 2$. The directness, DIR, is the most important attribute influencing the decisions of travellers in *Model_1*, but not in *Model_0* where $\%\Delta ITT$ dominates. It is interesting to note that DIR and $\%\Delta ITT$ are highly correlated (corr $= 0.5$) and that the correlation is positive. This implies that travellers who prefer direct routes are likely to prefer longer routes. This is true for both estimated models, suggesting that there may be two groups of travellers: one taking decisions mainly based on the travel time, and the other based on the directness or length of the trip.

Model_0 and *Model_1* are now evaluated in terms of their goodness-of-fit and their predictive accuracy. The goodness-of-fit is assessed by computing the *lppd* (expression (5.9)) and the discrepancy between the observed and predicted choice distributions *err* (expression (5.11)) using all the available observations. The estimation results in Table 5.5 show that the MXL model, *Model_0*, has a higher *lppd* than the boundedly rational model *Model_1* ($lppd^{M0}$ = -3,307 vs $lppd^{M1}$ = -3,335), meaning that, under *lppd*, the former model fits better the observations.

The Watanabe-Akaike Information Criterion (WAIC), the Bayesian analogous of the Akaike Information Criterion (AIC) that penalises the goodness-of-fit by the complexity of the models, is smaller for *Model_0*. As it can be seen in Table 5.5, the factor that penalises for the complexity of the models, p_{waic}, is similar in both cases, meaning that the difference in WAIC between the two models is only explained by the *lppd*. This is not surprising, since both models estimate the same number of parameters: α_i^{max} in the case of *Model_1* enters as a variable, it is not a parameter estimated by the model. If model selection were based on the WAIC, then *Model_0* should be selected. However, in terms of the discrepancy between the observed and predicted choice distributions measured by the error err, the results are the opposite. The error err is computed OD pair wise and then averaged considering the weight of each OD pair in the observations. The results, respectively for *Model_0* and *Model_1* are 11.0% and 10.4%, meaning that in this case *Model_1* fits better the observed route choice distributions. For completeness of the results, both *lppd* and err, aggregated per OD pair and weighted by the number of observations in each OD pair are presented in Fig. 5.15, where it can be seen that models' performance is OD pair dependent and that no model is systematically superior to the other.

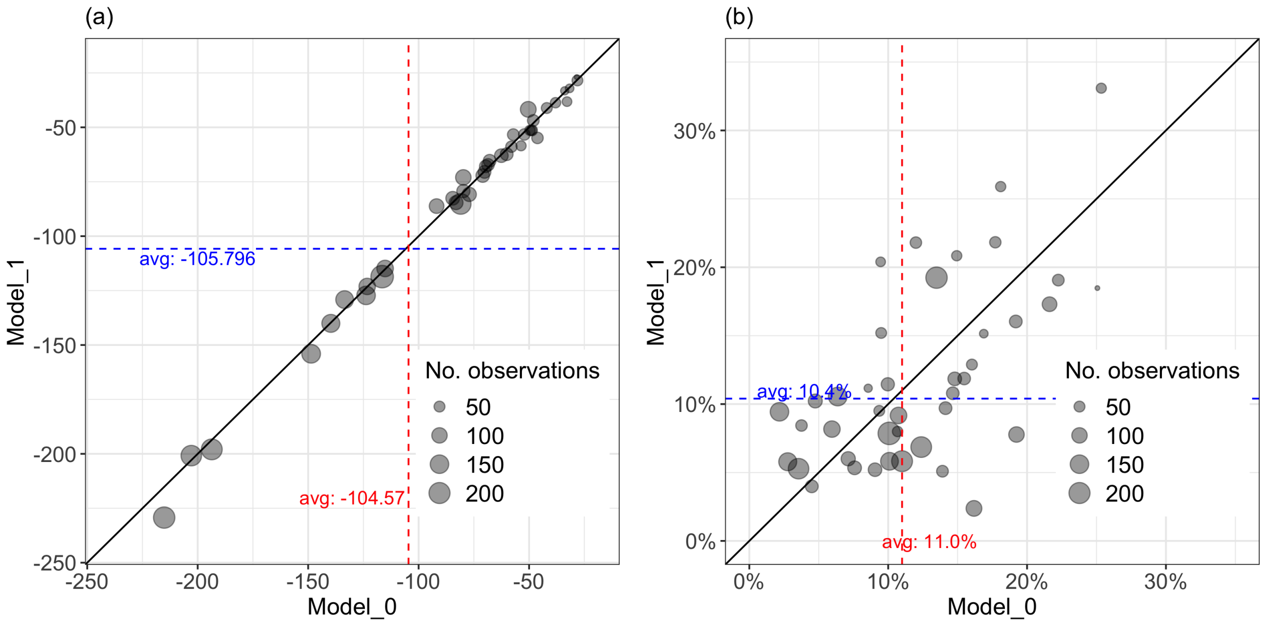

Figure 5.15: Goodness-of-fit of models *Model_0* and *Model_1* evaluated by OD pair considering (a) the *lppd* and (b) err measures. The average is represented by the dashed lines, it is obtained considering the weight of each OD pair in the data set.

To complete this analysis, the predictive accuracy of the models is obtained for out-of-sample observations to assess how the models generalise to unobserved choices. For this purpose, bootstrapping is performed with 10 iterations. At each iteration, 1/3 of the observations are removed from the training set, the models are estimated with the training set and the *lppd* and err are computed for the out-of-sample observations. Bootstrap validation is used (instead of cross-validation) to leave out a sufficient number of observations that permit to compute the choice distributions of the 41 OD pairs. The results are summarised in Fig. 5.16. The results throw the same conclusions as in the above analysis: in terms of the *lppd*, *Model_0* performs slightly better in predicting new choices: the average *lppd* values across the ten iterations are $\bar{lppd}^{M0}$ = -1,101.65 and $\bar{lppd}^{M1}$ = -1,116.49, but in terms of *err* the results are the opposite with err^{M0} = 12.65% and err^{M1} = 12.54%. The *lppd* is a measure related to the probability of observing the data, whereas *err* is a measure of discrepancy between the overall observed and predicted choice distributions. The opposite conclusions are explained because only the posterior predictive probability of the actual chosen alternatives contribute to the calculation of the *lppd*, while in the *err* the posterior predictive probabilities of the forgone alternatives are also taken into account. The error *err* is interpreted as the percentage of trips that are erroneously assigned (on average) on a given OD pair. Thus, from a route choice point of view, *err* is more informative as it is related to the collective behaviour of travellers (distribution of choices over the OD pair alternative routes). This is of crucial importance in estimating the network loading. As a final conclusion, it can be said that both models have a similar predictive accuracy. However, *Model_1* is more in accordance than *Model_0* with the findings in the descriptive analysis in this chapter, where bounded rational behaviour is observed. Moreover, the difference in predictive accuracy between the models could be larger in favour on *Model_1* in cases with many alternatives per OD pair. In this case, some of the alternatives may be not satisficing for all individuals, and thus *Model_1* would assign a probability of being chosen equal to zero, unlike *Model_0* which assigns always positive probability to all alternatives.

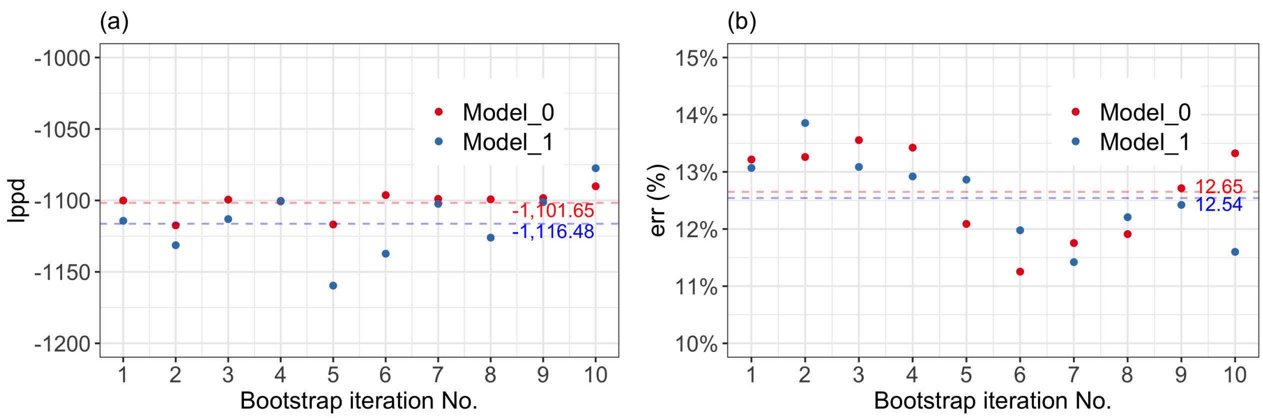

Figure 5.16: Out-of-sample (a) log pointwise predictive density (*lppd*) and (b) discrepancy between the observed and predicted choice distributions (*err*) for each iteration of the bootstrap validation.

5.5 Conclusions and discussion

In this work, the travel time minimisation behaviour and bounded rationality of travellers in route choice was studied through computer route choice experiments. In the experiments, participants made several route choices on 41 OD pairs presented over the road map of the city of Lyon. The choices of the participants were solely based on the travel time estimates (in minutes) and the map representation of the routes. It was found that, although participants received travel time estimates in the alternative routes, in 60.5% of the route choices participants chose the minimum travel time information route. This result lies within the range of those found in other studies (between 25% and 75%). However, it is important to take into consideration that the analysis presented here is based on route choices where participants received travel time information. Therefore, suggesting that in real-world situations, where travellers may not have travel time estimates on the forgone alternatives, the choices for the fastest route cannot be more than 60.5%. The percentage of choices for the fastest route was found to be OD pair and player dependant. According to the estimates of the MXL models, this dependency is explained by the heterogeneity of the preference of participants for the different route attributes, together with the variation of attributes between OD pairs. Apart from the travel time information, the directness of the routes resulted to be an important factor influencing the route choice of travellers.

The first main finding in this study is that travellers evaluate relative rather than absolute differences in travel time, at least for the ranges in travel time in the experiments. This means that a 5 minute difference in travel time weights different for trips of 10 and 30 minutes. In the first case, the difference represents an increment of travel time of 50% with respect to the alternative, whilst in the second case the difference is of 15%. Therefore, the 5 minute difference in the first case weighs more in favour of the fastest alternative. This implies that travellers minimise their travel time with respect to a reference point, given in this case by the travel time in the fastest route, and that the reference point is context-dependent, since it is evaluated in each route choice problem. This result has practical implications for the estimation of route choice models, and thus in traffic assignment, where expressing the travel time in relative terms could increase the realism of the predictions. For example, the travel times of the routes in each OD pair could be expressed as the percentage increase in travel time with respect to the minimum free flow travel time in that OD pair, or they can be transformed with the natural logarithm, as in the case of the Path Size Logit model (Ben-Akiva and Bierlaire, 1999), which also accounts for route overlapping. At individual level, a small percentage of the participants (10%) chose always the fastest route, these participants can be considered as perfect rational according to the definition given here. The behaviour of the rest of the participants can be better explained by bounded rationality. In this regard, it was found that the participants are heterogeneous with respect to their indifference band, and that at least 70% of them would not consider routes with travel time differences 1.5 times slower than the fastest alternative. The mean indifference band can be estimated as 31.3%, meaning that the average participant did not consider routes with travel time differences 1.3 times slower than the fastest alternative. This value coincides with the average additional travel time in the choices observed in Hadjidimitriou et al. (2015). If travellers are assumed to be homogeneous with an indifference band equal to the mean, it is possible to know with high probability (88.9%) which routes are travel time satisficing. An interesting finding is that amongst the satisficing routes, the minimum travel time route was always preferred, even for small relative differences in travel time. This suggests that just the fact of informing a route to be the fastest increases its probability to be chosen. In this chapter the increase was of around 10 percentage points with respect to the second fastest route. A MXL model was estimated considering the heterogeneous indifference bands that define the satisficing alternatives for each participant (*Model_1*). This model was compared to the estimates of the classical MXL model that takes into account all the alternatives (*Model_0*). The results show that, as expected, travel time information losses explanatory importance in the first model, while the rest of the variables maintain their same level. Thus, amongst the satisficing alternatives, participants put more stress on the physical route attributes rather than on travel time information for their route choices. These models were compared in terms of their predictive accuracy for out-of-sample observations, resulting in similar predictive accuracy. When measured in terms of erroneously assigned trips for a given OD pair, the errors are around 12.6%. However, *Model_1* is more in accordance than *Model_0* with the bounded rational behaviour observed in the descriptive analysis in this chapter. This result is promising, considering that in *Model_1* the definition of the exogenous indifference band is α_i^{max} is restrictive. A bounded rational model that considers more flexible definitions for the indifference band could improve the performance. This would require to investigate more complex models capable of inferring the indifference bands endogenously. Moreover, *Model_1* could be more advantageous in choice situations with many alternatives, in which some alternatives will not be satisficing for all individuals and thus they will have probability equal to zero of being chosen. These questions are left as the subject of future investigation.

The findings in this chapter may have practical implications that are left for future work. In traffic simulation, the estimates for the indifference band can be used to reduce the search space of choice set generation by discarding routes with travel time differences (with respect to the shortest time route) above α; or used as exogenous inputs in bounded rational models as the one proposed in Watling et al. (2018). In these cases, the impact of considering homogeneous versus heterogeneous indifference bands could be assessed to determine the trade-offs between the simplicity of the former and the realism of the later. Apart from these two practical applications, the estimates of the indifference band can shed some light on the boundaries in which the users' route choices could be influenced, with the objective of directing them towards the social optimum (van Essen et al. (2016) provides a complete review on this subject).

5.A Detailed results of the models estimates

Table 5.6: Results *Model_0*. Statistics on the posterior distribution for the mean and covariance matrix.

	Statistic*							
Parameter	Mean	s.d.	qt.2.5%	qt.25%	qt.50%	qt.75%	qt.97.5%	$\hat{R}^{**}$
$\hat{\beta}_{FRW}$	0.862	0.313	0.242	0.658	0.869	1.065	1.485	1.026
$\hat{\beta}_{DIR}$	1.377	0.640	0.018	0.969	1.414	1.838	2.502	1.042
$\hat{\beta}_{TRN}$	0.012	0.108	-0.204	-0.059	0.014	0.086	0.218	1.021
$\hat{\beta}_{INT}$	-0.044	0.029	-0.099	-0.063	-0.045	-0.024	0.011	1.007
$\hat{\beta}_{\%\Delta ITT}$	-3.285	0.356	-4.010	-3.512	-3.276	-3.045	-2.604	1.018
$\hat{\sigma}^2_{FRW}$	2.840	2.253	0.258	0.896	2.261	4.285	8.010	1.123
$\hat{\sigma}^2_{DIR}$	17.889	11.229	2.304	9.003	16.272	24.554	44.478	1.165
$\hat{\sigma}^2_{TRN}$	0.505	0.185	0.218	0.371	0.475	0.615	0.935	1.042
$\hat{\sigma}^2_{INT}$	0.052	0.011	0.034	0.044	0.051	0.059	0.077	1.008
$\hat{\sigma}^2_{\%\Delta ITT}$	17.121	3.230	11.608	14.842	16.833	19.134	24.192	1.001
$\hat{\sigma}_{FRW,DIR}$	4.660	4.628	-1.565	0.599	3.936	7.758	14.893	1.122
$\hat{\sigma}_{FRW,TRN}$	0.389	0.478	-0.379	0.039	0.298	0.714	1.460	1.091
$\hat{\sigma}_{FRW,INT}$	0.011	0.092	-0.173	-0.045	0.015	0.069	0.191	1.022
$\hat{\sigma}_{FRW,\%\Delta ITT}$	0.116	1.843	-3.147	-1.187	0.016	1.284	3.987	1.035
$\hat{\sigma}_{DIR,TRN}$	1.175	1.025	-0.309	0.363	1.011	1.840	3.468	1.081
$\hat{\sigma}_{DIR,INT}$	-0.340	0.227	-0.886	-0.465	-0.304	-0.178	0.014	1.004
$\hat{\sigma}_{DIR,\%\Delta ITT}$	9.290	3.853	2.868	6.492	8.991	11.662	17.941	1.008
$\hat{\sigma}_{TRN,INT}$	-0.029	0.032	-0.099	-0.048	-0.027	-0.007	0.028	1.007
$\hat{\sigma}_{TRN,\%\Delta ITT}$	-0.530	0.530	-1.582	-0.880	-0.535	-0.172	0.479	1.001
$\hat{\sigma}_{INT,\%\Delta ITT}$	-0.336	0.160	-0.675	-0.438	-0.328	-0.223	-0.042	1.005
lppd = -3307.05; WAIC*** =6,632.27 (p_{waic}=9.08)								

* Statistics based on 3,000 out of 80,000 samples (40,000 burn-in period) and saving 1/40 samples (thinning).
** Potential Scale Reduction. When the MCMC chains converge, it takes values close to 1.
*** WAIC is an estimate of expected predictive error (lower WAIC is better).

Table 5.7: Results *Model_1*. Statistics on the posterior distribution for the mean and covariance matrix.

	Statistic*							
Parameter	Mean	s.d.	qt.2.5%	qt.25%	qt.50%	qt.75%	qt.97.5%	$\hat{R}$
$\hat{\beta}_{FRW}$	0.622	0.340	-0.053	0.393	0.625	0.853	1.281	1.010
$\hat{\beta}_{DIR}$	1.863	0.753	0.387	1.331	1.868	2.395	3.311	1.001
$\hat{\beta}_{TRN}$	0.061	0.132	-0.197	-0.030	0.062	0.152	0.311	1.012
$\hat{\beta}_{INT}$	-0.076	0.032	-0.139	-0.098	-0.076	-0.053	-0.014	1.006
$\hat{\beta}_{\%\Delta ITT}$	0.366	0.373	-0.378	0.117	0.360	0.626	1.097	1.001
$\hat{\sigma}^2_{FRW}$	3.596	1.842	0.669	2.247	3.381	4.671	7.766	1.039
$\hat{\sigma}^2_{DIR}$	25.970	9.383	10.586	19.601	24.665	31.469	47.839	1.010
$\hat{\sigma}^2_{TRN}$	0.464	0.178	0.188	0.329	0.440	0.570	0.861	1.005
$\hat{\sigma}^2_{INT}$	0.058	0.014	0.035	0.048	0.056	0.066	0.089	1.001
$\hat{\sigma}^2_{\%\Delta ITT}$	10.867	2.575	6.357	9.056	10.690	12.499	16.354	1.004
$\hat{\sigma}_{FRW,DIR}$	7.470	3.564	1.634	4.831	7.176	9.757	15.380	1.039
$\hat{\sigma}_{FRW,TRN}$	0.543	0.451	-0.224	0.237	0.490	0.813	1.525	1.028
$\hat{\sigma}_{FRW,INT}$	-0.031	0.093	-0.219	-0.091	-0.033	0.031	0.154	1.019
$\hat{\sigma}_{FRW,\%\Delta ITT}$	0.882	1.596	-2.146	-0.141	0.874	1.831	4.347	1.011
$\hat{\sigma}_{DIR,TRN}$	1.622	1.057	-0.256	0.932	1.521	2.218	3.909	1.026
$\hat{\sigma}_{DIR,INT}$	-0.486	0.274	-1.075	-0.653	-0.465	-0.292	-0.017	1.002
$\hat{\sigma}_{DIR,\%\Delta ITT}$	8.525	3.696	1.888	5.925	8.295	10.930	15.974	1.003
$\hat{\sigma}_{TRN,INT}$	-0.025	0.035	-0.096	-0.047	-0.024	-0.002	0.043	1.009
$\hat{\sigma}_{TRN,\%\Delta ITT}$	0.021	0.543	-1.059	-0.333	0.034	0.393	1.069	1.024
$\hat{\sigma}_{INT,\%\Delta ITT}$	-0.382	0.150	-0.715	-0.477	-0.370	-0.276	-0.121	1.003
lppd = -3334.77; WAIC*** =6,688.20 (p_{waic}=9.33)								

* Statistics based on 3,000 out of 80,000 samples (40,000 burn-in period) and saving 1/40 samples (thinning).
** Potential Scale Reduction. When the MCMC chains converge, it takes values close to 1.
*** WAIC is an estimate of expected predictive error (lower WAIC is better).

6 Bounded rational choice set generation MXL model for route choice

In this chapter, a choice model that considers bounded rational behaviour in the individuals' choice set generation for route choice is developed (BRCS model). In the BRCS, the distribution of the indifference bands is inferred endogenously by jointly estimating the choice set generation and route choices. The model is proposed as an alternative to the MXL model in Chapter 5, where the individually estimated indifference bands are exogenously estimated and thus enter the model as independent variables. The BRCS model is compared, in terms of predictive accuracy, to the MXL model using synthetic and real data, obtained from the experiments with the MDG platform. The results show that the BRCS model is capable of inferring the distribution of the indifference bands for the synthetic generated data. Moreover, for this data, the BRCS shows higher predictive accuracy than the MXL model. In the case of the MDG data, the BRCS model exhibits higher predictive accuracy than both the MXL and the MXL with exogenously estimated indifference bands of Chapter 5.

6.1 Motivation

The model for bounded rationality adopted in this thesis assumes that decision-makers are indifferent to the travel times of the alternatives when they are below a threshold (indifference band), and that the thresholds are user-specific. The indifference band is defined to be relative to the travel time of the fastest route, which can be expressed as *travellers only consider routes with increments in travel time (with respect to the fastest route) below a threshold.* This user-specific threshold is given as a percentage increment in travel time: some travellers may not consider routes that are 10% slower than the fastest route, while for other travellers this threshold may be 30% or 0%. This latter value corresponds to the perfect rational individuals, i.e., individuals who chose always the fastest route. In Chapter 5, a MXL model conditional on the estimated individual indifference bands was introduced. In that model, the exogenously estimated indifference bands for each individual, α_i, entered the route choice model as an independent variable, allowing to study their impact on the choice probabilities. However, the estimation methods developed in Chapter 5 may suffer from some issues related to the latent nature of the indifference band. Furthermore, the estimators $\hat{\alpha}_i$ that can be used as exogenous variables in the conditional MXL model are limited. To avoid these issues, in this chapter a route choice model based on the MXL model that jointly considers the latent indifference bands is proposed.

The indifference band influences the choices of travellers by determining the choice set under consideration. Nonetheless, they are latent constructs that cannot be directly observed. What can be observed are the choices of individuals in different situations. Based on the repeated route choices of participants in the MDG platform, in Section 5.3.3 three estimators were proposed. The estimator $\hat{\alpha}_i^{max}$ is the maximum travel time difference of a chosen route amongst all the choice problems faced by individual i; $\hat{\alpha}_i^{95}$ the 95th percentile; and $\hat{\alpha}_i^{50}$ the median. However, the choice problems that participants face in the MDG cover only a limited number of situations (travel time differences) that may introduce biases to the estimates. For example, the estimated indifference band of a participant that faces three choice problems with all travel time differences below 5% will be at most 5%. A similar problem happens when all the travel time differences are large and, as a consequence, the participant chooses always the fastest route. In this case, the estimated indifference band would be 0%. In both examples, the indifference band is underestimated, as the estimators are inferred from partial observations. A third case is when an outlier is present in the data, that could be caused by a wrong manipulation of the participant in the MDG or because the participant was testing the tool. In this case, a chosen route with a large travel time difference may cause the estimator to overestimate the real value of the indifference band. These three cases are illustrated in Figure 6.1. In addition, not all of the proposed estimators can be used as inputs in a discrete choice model. The reason is that, except for $\hat{\alpha}_i^{max}$ the rest of the estimators violate the assumption in discrete choice models that the actual chosen alternative is part of the choice set. This is, for estimators $\hat{\alpha}_i < \hat{\alpha}_i^{max}$ the actual chosen alternative may be assigned a choice probability equal to zero, making it impossible to estimate the model. Therefore, the estimator of the indifference band used as input in the choice model must never discard the actual chosen alternatives from the choice sets; the only estimators that guarantee this are those with values greater than or equal to $\hat{\alpha}_i^{max}$. This issue is detailed in Section 5.3.4.

The bounded rational route choice model proposed in this chapter is a two-step process in which (1) decision-makers conform their own choice set by discarding not satisficing routes from the available alternatives, and then (2) they choose one of the satisficing alternatives. The first step pertains a bounded rational process, discarding the alternatives with travel times above a threshold (indifference band). The second step here is rational, since a MXL model

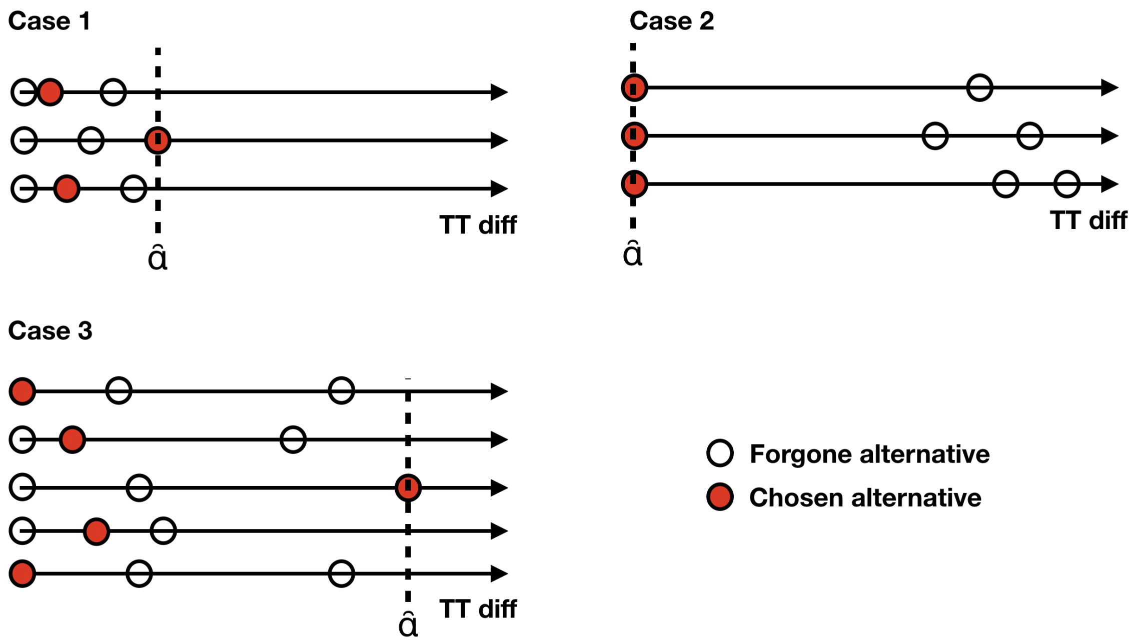

Figure 6.1: Issues in the individual estimation of the indifference bands α. The cases represent three choice problems, each with three alternative routes, faced by an individual.

is proposed to model the choice probability for the satisficing routes. The model is called bounded rational choice set generation mixed logit model (BRCS). This model is based on the ideas of probabilistic choice set model generation (Manski, 1977, Ben-akiva, 1987, Horowitz and Louviere, 1995, Haab and Hicks, 1997, Swait, 2001), in which the choice sets are endogenous to the choice model and they are decision-maker dependent. These models assume that the choice set that decision-makers consider are limited to a subset from all the available alternatives, but that the subset is not available to the researcher, i.e., it is a latent construct. Thus, a probability is assigned to every possible subset, and this probability is estimated jointly in the model. Also, since the travel time gives an ordering of the alternatives (from fastest to slower), the number of possible considered subsets according to the boundedly rational assumption is at most equal to the number of alternatives. By estimating the distribution of the indifference bands in the population, jointly with a MXL model for the satisficing routes, the BRCS model overcomes the issues in the individually estimated indifference bands. In the BRCS model the choices of all participants are considered together, thus the unobserved travel time differences for some individuals are inferred from the observed differences for other individuals. Moreover, in the BRCS model the probability of choosing an alternative depends on both the characteristics of the alternative and the probability that it belongs to the choice sets, which in this case are always greater than zero. Therefore, the BRCS model does not present the estimation issues of the conditional MXL model. In this chapter, the BRCS model's ability to infer the indifference band, as well as its prediction accuracy are tested using both synthetic and real data.

The rest of this chapter is organised as follows. The BRCS model is developed in Section 6.2. Then, the methodology to generate the synthetic data, and the methodology to assess the prediction accuracy of the model are presented in Section 6.3. The BRCS model's ability to infer the distribution of the indifference bands, as well as the prediction accuracy over the synthetic and real data are included in the results in Section 6.4. Finally, the last section of this chapter comprises the discussion and conclusions.

6.2 Bounded rational choice set generation model

According to the bounded rationality assumption, a traveller α_i considers only routes with travel time differences (with respect to the fastest route), that are below α_i. This problem is equivalent to the problem of heterogeneous choice sets in discrete choice models, in which individuals consider a subset of all the possible alternatives. Moreover, this subset is not observed by the bystander and, thus, it needs to be estimated endogenously. Let R be the set of all the possible alternatives, indexed by j, in a choice problem faced by a decision-maker i, and define $S_i \subseteq R$ as the choice set considered by i. S_i is the set of satisficing alternatives, i.e., $Pr(y_{ij} = 1 \,|\, S_i) = 0 \ \forall j \notin S_i$. Since the subset S_i is latent, it is considered a random variable and thus the probability that individual i chooses alternative j is given by

$$Pr(y_{ij} = 1) = \sum_{S \in \mathcal{P}(R)} Pr(y_{ij} = 1 \,|\, S) Pr(S)\,, \tag{6.1}$$

where $\mathcal{P}(R)$ is the power set of R, i.e., the set with all the possible subsets of R; $Pr(y_{ij} = 1 \,|\, S)$ is a discrete choice model; and $Pr(S)$ expresses the uncertainty on the actual choice set considered by the individual. The model in this section is formulated as a special case of the above model.

Consider the case in which an individual conform his/her choice set by including only the alternatives j in which the attribute z_j is under a threshold α_i, i.e., i considers j if and only if $z_j \leq \alpha_i$. Without loss of generality, it can be assumed that the alternatives are ordered with respect to z, i.e., $z_1 < z_2 < \cdots < z_J$, where $J = |R|$. Then, the only elements $S \in \mathcal{P}(R)$ that have non-zero probability of occurring are $\{1\}, \{1, 2\}, \dots, \{1, \dots, J\}$. Denote the set $S_k = \{1, 2, \dots, k\}$. Since α_i is unobserved by the bystander, then it can be considered as a continuous random variable, which implies that

$$\begin{aligned} Pr(S_k) &= Pr(\{1, 2, \dots, k\}) \\ &= Pr(z_{k'} \leq \alpha \ \forall k' \leq k \text{ and } z_l > \alpha \ \forall l > k) \\ &= Pr(z_k \leq \alpha < z_{k+1})\,. \end{aligned}$$

Noting that the probability of choosing a route not belonging to the choice set is zero, i.e., $Pr(y_{ij}=1\,|\,S_k)=0$ for all $k<j$, then expression (6.1) can be written as

$$\begin{aligned}Pr(y_{ij}=1\,|\,\boldsymbol{\theta}_\alpha) &= \sum_{k=j}^{J} Pr(y_{ij}=1\,|\,S_k)Pr(S_k)\\ &= \sum_{k=j}^{J-1}\{Pr(y_{ij}=1\,|\,S_k)Pr(z_k\le\alpha<z_{k+1})\} + Pr(y_{ij}=1\,|\,S_J)Pr(z_J\le\alpha)\\ &= \sum_{k=j}^{J-1}\left\{Pr(y_{ij}=1\,|\,S_k)\int_{z_k}^{z_{k+1}} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\right\} + Pr(y_{ij}=1\,|\,S_J)\int_{z_J}^{\infty} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\,,\end{aligned}$$

where $f_\alpha(\cdot\,;\boldsymbol{\theta}_\alpha)$ is the probability density function of α, and $\boldsymbol{\theta}_\alpha$ its set of parameters. If it is assumed that $Pr(y_{ij}=1\,|\,S_k)$ is given by a MXL model, then the probability that individual i chooses route j, conditioned on the parameters $\boldsymbol{\beta}_\mathbf{i}$ and the vector of attributes $\mathbf{x_{ij}}$ is given by

$$\begin{aligned}Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i,\boldsymbol{\theta}_\alpha;\mathbf{x_{ik}}\,\forall\,k) = \sum_{k=j}^{J-1}&\left\{\frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m=1}^{k} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}}\int_{z_k}^{z_{k+1}} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\right\}\\ &+\frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m=1}^{J} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}}\int_{z_J}^{\infty} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\,.\end{aligned}\tag{6.2}$$

Note that if the cumulative distribution function of α, $F_\alpha(z;\boldsymbol{\theta}_\alpha)=\int_{-\infty}^{z} f_\alpha(u;\boldsymbol{\theta}_\alpha)du$, has a closed form, then the probabilities $Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i,\boldsymbol{\theta}_\alpha;\mathbf{x_{ik}}\,\forall\,k)$ will also have a closed form.

The model in Eq. (6.1) is a two-step process: (1) the individual conforms the set S_i by discarding the non satisficing alternatives from a larger set and, then, (ii) he/she chooses an alternative from S_i. This process differs from other interpretations of bounded rationality, where individuals look for a satisficing alternative, and once found, they choose it. Note that in the interpretation given here, the first step pertains a boundedly rational process, and the second a rational process since $Pr(y_{ij}=1\,|\,S_i)$ is a RUM. The model in expression (6.2) can be then named bounded rational choice set mixed logit model, or BRCS model for short.

It can be shown that the BRCS model can be obtained by integrating out α from the conditional probability, $Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i;\alpha_i,\mathbf{x_{ik}}\,\forall k)$, defined in Section 5.3.4. To see this, the conditional probability is first rewritten as

$$\begin{aligned}Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i,\alpha_i;\mathbf{x_{ik}}\,\forall k) &= \begin{cases}\frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m|z_m\le\alpha_i} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}}, & z_j\le\alpha_i\\ 0, & \text{otherwise}\end{cases}\\ &= \frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}\mathbf{1}(z_j\le\alpha_i)}{\sum_{m=1}^{J} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}\mathbf{1}(z_m\le\alpha_i)}\,,\end{aligned}$$

where $\mathbf{1}(z_m\le\cdot)$ is the indicator function. Then, by the law of total probability and noting that the conditional probability is different to zero when $\alpha > z_j$,

$$\begin{aligned}
Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i,\boldsymbol{\theta}_\alpha;\mathbf{x_{ik}}\,\forall k) &= \int_{z_j}^{\infty} Pr(y_{ij}=1\,|\,\boldsymbol{\beta}_i,u;\mathbf{x_{ik}}\,\forall k) f_\alpha(u;\boldsymbol{\theta}_\alpha)du \\
&= \int_{z_j}^{\infty} \frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m=1}^{J} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}\mathbf{1}(z_m \le u)} f_\alpha(u;\boldsymbol{\theta}_\alpha)du \\
&= \sum_{k=j}^{J-1}\left\{\int_{z_k}^{z_{k+1}} \frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m=1}^{k} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\right\} \\
&\qquad + \int_{z_J}^{\infty} \frac{e^{V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})}}{\sum_{m=1}^{J} e^{V(\mathbf{x_{im}};\boldsymbol{\beta}_\mathbf{i})}} f_\alpha(u;\boldsymbol{\theta}_\alpha)du\,,
\end{aligned} \tag{6.3}$$

which is equal to expression (6.2). The last equality in the above expression is due to the additive property of integrals, and because when $z_k \le u \le z_{k+1}$, the maximum value that m can take is k, and thus $\mathbf{1}(z_m \le u) = 1$.

The model in expression (6.2) is now illustrated for the case of three alternatives, which is the case for the route choice experiments in this thesis. According to the bounded rational model assumed here, travellers do not consider in their choice set the routes with travel time difference (with respect to the shortest route) above α. Thus, $z_{ij} = (TT_{ij} - TT_{i(1)})/TT_{i(1)}$, where TT_{ij} is the travel time in route j when the problem is faced by individual i, and $TT_{i(1)}$ is the shortest time route. For ease of notation, define $V_{ij} = V(\mathbf{x_{ij}};\boldsymbol{\beta}_\mathbf{i})$. Then, the choice probabilities for each of the three routes is given by

$$\begin{aligned}
Pr(y_{i1}=1) &= Pr(0<\alpha<z_2) + Pr(z_2 \le \alpha < z_3)\tfrac{e^{V_{i1}}}{e^{V_{i1}}+e^{V_{i2}}} + Pr(z_3 \le \alpha)\tfrac{e^{V_{i1}}}{e^{V_{i1}}+e^{V_{i2}}+e^{V_{i3}}} \\
Pr(y_{i2}=1) &= Pr(z_2 \le \alpha < z_3)\tfrac{e^{V_{i2}}}{e^{V_{i1}}+e^{V_{i2}}} + Pr(z_3 \le \alpha)\tfrac{e^{V_{i2}}}{e^{V_{i1}}+e^{V_{i2}}+e^{V_{i3}}} \\
Pr(y_{i3}=1) &= Pr(z_3 \le \alpha)\tfrac{e^{V_{i3}}}{e^{V_{i1}}+e^{V_{i2}}+e^{V_{i3}}}\,.
\end{aligned} \tag{6.4}$$

In this case, the distribution of the values of α is interpreted as the distribution of indifference bands amongst the decision-makers.

6.2.1 Estimation

The estimation of the BRCS model requires to find plausible values for the parameters $\boldsymbol{\theta}_\beta$ and $\boldsymbol{\theta}_\alpha$ that determine the shapes of the distributions of the random coefficients $\boldsymbol{\beta}_\mathbf{i}$ in the MXL model, and the indifference bands α. Following the same notation as in Section 3.2.2, the BRCS can be written as the following hierarchical Bayesian model. For each individual i,

$$\begin{aligned}
\boldsymbol{y_{is}}|\boldsymbol{\beta}_{\mathbf{i}}, \boldsymbol{\theta}_\alpha \sim Cat(\boldsymbol{p_{is}}(\boldsymbol{\beta}_i, \boldsymbol{\theta}_\alpha)) \quad \forall s \\
p_{isj} = Pr(y_{ij} = 1 \,|\, \boldsymbol{\beta}_i, \boldsymbol{\theta}_\alpha; \mathbf{x_{ik}}\, \forall k) \;\forall\, j, s \\
\boldsymbol{\beta}_{\mathbf{i}} \sim f_\beta(\cdot\,;\, \boldsymbol{\theta}_\beta) \\
\boldsymbol{\theta}_\alpha \sim f_{\boldsymbol{\theta}_\alpha}(\cdot\,;\, \boldsymbol{\kappa}) \\
\boldsymbol{\theta}_\beta \sim f_{\boldsymbol{\theta}_\beta}(\cdot\,;\, \boldsymbol{\rho}) \,,
\end{aligned} \tag{6.5}$$

where $f_{\boldsymbol{\theta}_\alpha}$ and $f_{\boldsymbol{\theta}_\beta}$ are, respectively, the priors of the parameters' $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\beta$. The hierarchy is given by the coefficients $\boldsymbol{\beta}_{\mathbf{i}}$, which can be considered as individual-specific parameters. The parameters in $\boldsymbol{\theta}_\alpha$ are global, in the sense that it is shared by all individuals; they define the distribution of the indifference bands α in the population.

To estimate the MXL model in this study a Bayesian approach was adopted (see Train (2001)), relying on the Gibbs sampling method (Levin and Peres, 2017). The Gibbs sampler software JAGS (Plummer, 2003) and the R (R Core Team, 2018) package *rjags* were used to obtain samples from the posterior distribution of the parameters μ and Σ. The values of the hyperparameters $\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$, which define the priors, were chosen to be weakly-informative (very high variances). In other words, high uncertainty is assumed on the real values of the parameters that are being estimated.

6.3 Methodology

6.3.1 Synthetic data generation

The BRCS model is first tested for synthetic data, generated following the bounded rational choice set generation model. The test has two objectives: first, to see if the model is capable of finding the distribution of α and, second, to see the impact of different distributions of α in the models estimates and predictions. The results are compared with the results of a MXL model that considers no indifference band. The synthetic data comprises 100 fictional decision-makers, each facing 10 choices between three alternatives. The choices are based on two independent variables: z_{ij}, compared by the decision-maker against its own indifference band, α_i, to determine the choice sets in each problem; and x_{ij}, which enters the systematic part of the utility $V(x_{ij}; \beta_i)$. The values of the variables z and x are simulated from normal distributions with different means and variances. The process to generate the data is shown in Algorithm 1, which is based on the data generating process in expression (6.5).

Six data sets were generated considering values of α coming from six different parametrisations of the Weibull distribution. This distribution is the power transformation of the expo-

Algorithm 1: Pseudo-code for the generation of the synthetic data set with boundedly rational choices.

Input: Parameters of the distributions of the indifference bands α, k and λ.
Output: Realisations, $y_{i,c}$, of the bounded rational choice set MXL model.
for *individual* i **do**
  $\beta_{x,i} \sim \mathcal{N}(3,1)$
  $\alpha_i \sim \text{Weibull}(k,\lambda)$;
  for *choice* c **do**
    $x_{i,c,1} \sim \mathcal{N}(3,4)$; $x_{i,c,2} \sim \mathcal{N}(4,4)$; $x_{i,c,3} \sim \mathcal{N}(4,1)$
    $z_{i,c,1} \sim \mathcal{N}(20,16)$; $z_{i,c,2} \sim \mathcal{N}(25,16)$; $z_{i,c,3} \sim \mathcal{N}(35,64)$
    $z_{i,c,j} \leftarrow z_{i,c,j} / \min(z_{i,c,1}, z_{i,c,2}, z_{i,c,3})$, $j = 1,2,3$
    $S \leftarrow \{j \,|\, z_{i,c,j} \leq \alpha_i \,\forall\, j\}$
    $P_j \leftarrow 0$
    for $j \in S$ **do**
      $P_j \leftarrow \exp(\beta_{x,i} x_{i,c,j})$
    end
    $P_j \leftarrow P_j / \sum_{k=1}^{3} P_k$, $\forall j$
    $y_{i,c} \leftarrow Cat(P)$
  end
end

* The second parameter in the normal distribution is the variance.

nential distribution, it takes non-negative values and has two parameters: the shape parameter $k > 0$ and the rate parameter $\lambda > 0$; the exponential distribution is the special case when $k = 1$. The Weibull distribution was chosen for its flexibility (it can take several shapes), because it has a closed form, and because for certain parameters it accumulates values close to zero rapidly. The six probability density functions used to generate the data are shown in Figure 6.2. Three of the six distributions correspond to the exponential distribution ($k = 1$). The density, f, and cumulative, F, functions of the Weibull distribution are

$$\begin{aligned} w(t;k,\lambda) &= k\lambda t^{k-1} e^{-\lambda t^k} \\ W(t;k,\lambda) &= 1 - e^{-\lambda t^k} . \end{aligned} \tag{6.6}$$

with $t \geq 0$.

The distributions of the generated choices over the three alternatives are shown in Figure 6.3. Observe that the number of choices for the alternative $R1$ generally increases when the values of α are distributed closer to zero, i.e., when the parameter λ increases for the exponentially distributed α ($k = 1$), or when k decreases. For small values for α_i decision-makers discard more alternatives from the choice set, thus giving more probability to $R1$ to be chosen.

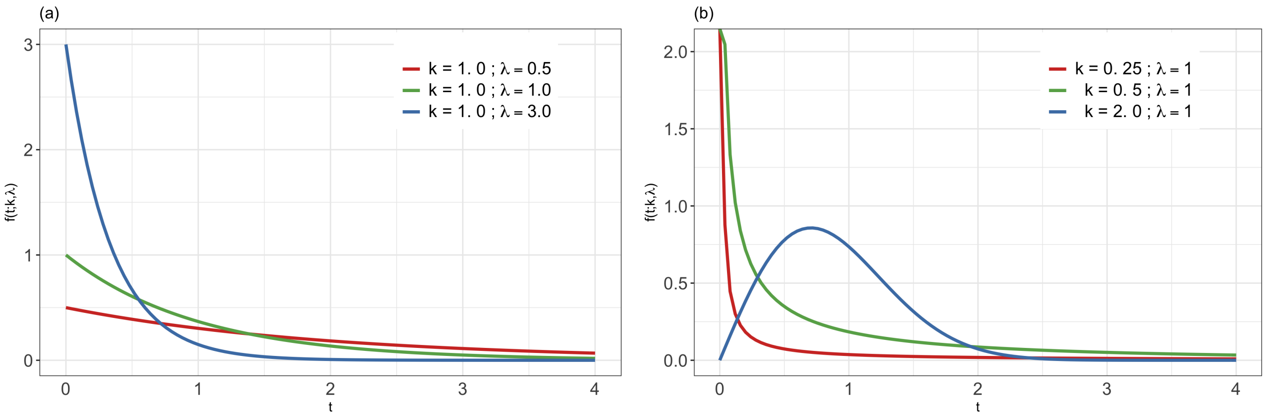

Figure 6.2: Weibull probability density functions used to generate the data with (a) three different rate parameters and $k = 1$ (exponential distribution), and (b) three different shape parameters and same rate $\lambda = 1$. Note in this last plot how when $k < 1$ the values are accumulated closer to zero.

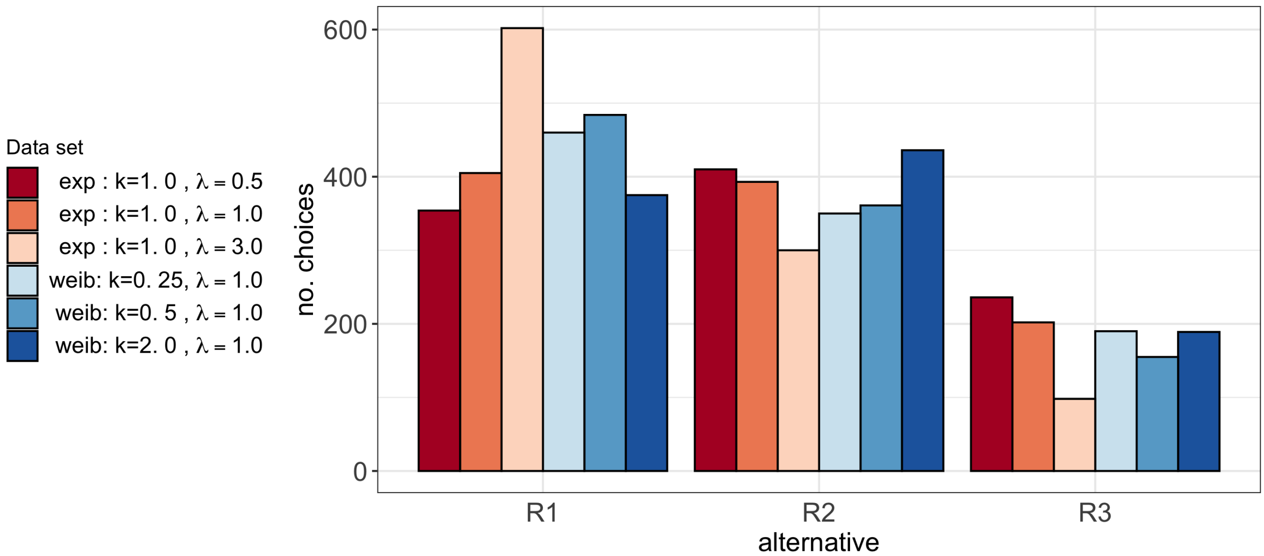

Figure 6.3: Choice distributions for the synthetic data.

6.3.2 Prediction accuracy

Cross-validation is used to compare the MXL and BRCS models in terms of prediction accuracy. In this study, cross-validation is performed by randomly removing one third of the observations from the data set at each iteration. The removed observations conform the test set and the remaining observations the training set. The models are estimated for the training set and they are used to predict the responses in the test set. This process is repeated ten times. The prediction error is defined as the absolute mismatch between the observed and the predicted choice probabilities, given by

$$err(P, \hat{P}) = \sum_{j=1}^{3} \max(0, P_j - \hat{P}_j)\,, \tag{6.7}$$

where P_j and $\hat{P}_j$ are the observed and predicted probabilities of choosing alternative j. This error function is a metric.

Seeing that the MXL and BRCS models have a hierarchical structure in which each decision-maker makes several choices, two cross-validation schemes are proposed:

- CV1, at each iteration, randomly remove 1/3 of the choices from the data set;
- CV2, at each iteration, randomly remove 1/3 of the decision-makers from the data set.

The schema CV1 is performed on both the synthetic and the MDG experiments' data sets; schema CV2 is performed only on the synthetic data sets. The reason is because in the synthetic data sets there is a sufficient number of choices per individual (10 choices), so it is likely that in each iteration the same individuals will remain in both the training and test sets under CV1, in which 30% of the choices are removed. Since the value of α is per individual and not per choice, then the latter schema entails greater uncertainty.

6.4 Results

6.4.1 Synthetic data

The BRCS model is fitted to the six simulated data sets. For the exponentially distributed α, the shape parameter of the distribution $w(t; k, \lambda)$ is fixed to $k = 1$; and for the Weibull distributed α, the rate parameter is fixed to $\lambda = 1$. This is, only one parameter is estimated. The specification of the MXL part of the model is

$$V_{ij} = \beta_{x,i} x_{ij}\,, \tag{6.8}$$

with $\beta_{x,i} \sim \mathcal{N}(\beta_x, \sigma_x^2)$. This is, the variable z_{ij} does not enter the specification of the MXL part of the model. First, the fit of the BRCS model is analysed for the six generated data sets, with particular attention to the estimated distribution for the values of α. Then, the BRCS model is compared, in terms of predictive accuracy, to the MXL model given by

$$V_{ij} = \beta_{x,i} x_{ij} + \beta_{y,i} z_{ij}\,, \tag{6.9}$$

with $\beta_{x,i} \sim \mathcal{N}(\beta_x, \sigma_x^2)$ and $\beta_{y,i} \sim \mathcal{N}(\beta_y, \sigma_y^2)$.

The estimation results of both models for the six data sets are presented in Table 6.1. The results show that the BRCS estimates for the parameters k and λ are close to their real values, suggesting that the BRCS can be used to make inference about the indifference bands of decision-makers. This can be best seen in Figure 6.4, where the simulated and theoretical values of α are compared against the estimated distribution. The estimates for the β_z coefficient in the MXL models show a clear pattern: the higher the value of λ in the exponential case or the smaller the value of k in the Weibull case, the smaller the coefficients β_z are and the larger their variance $\hat{\sigma_z}$. This is explained because higher values of λ or smaller values of k entail values of α closer to zero, thus leaving out more alternatives from the choice set and making this variable more important in the choice probabilities. The higher variances are explained because the MXL model tries to compensate for higher values of α.

Table 6.1: Estimation results for synthetic data. The real values of the coefficients β is shown in parenthesis.

	Exponentially distributed α ($k = 1.0$)						Weibull distributed α ($\lambda = 1.0$)					
Params.	$\lambda = 0.5$		$\lambda = 1.0$		$\lambda = 3.0$		$k = 0.25$		$k = 0.5$		$k = 2.0$	
Coeff.	BRCS	MXL	BRCS	MXL	BRCS	MXL	BRCS	MXL	BRCS	MXL	BRCS	MXL
$\hat{\beta_x}$ (=3)	3.022	2.206	3.549	1.681	3.228	1.060	3.181	1.932	3.665	1.510	2.518	1.704
$\hat{\beta_z}$	-	-3.576	-	-5.641	-	-12.990	-	-9.450	-	-11.287	-	-4.681
$\hat{\sigma_x}$ (=1)	1.603	1.305	1.970	1.074	1.007	0.678	2.957	1.780	3.134	1.303	1.254	0.754
$\hat{\sigma_z}$	-	3.595	-	5.234	-	8.039	-	11.233	-	11.168	-	2.866
$\hat{\lambda}$	0.539	-	0.972	-	2.866	-	-	-	-	-	-	-
$\hat{k}$	-	-	-	-	-	-	0.571	-	0.592	-	2.332	-

In the results, it is clear that the estimates of the coefficients $\hat{\beta_x}$ differ more between the BRCS and MXL models, but less for the BRCS model fitted to the different data sets. Furthermore, the estimates $\hat{\beta_x}$ of the BRCS model are closer to their real values, compared to the estimates of the MXL model. As a consequence, the MXL model may lead to erroneous conclusions in the interpretation of the coefficients. In the example here, the $\hat{\beta_x}$ in the MXL model underestimates the effect that x has on the choices. However, it can be seen that the estimates for the standard deviations are large compared to the real values in the case of the BRCS model. This may be due to some problem of identification between the variances of β_x and the indifference band. Further research need to be done in order to identify the cause of larger variances.

The prediction error for each iteration in the two cross-validation schemes is presented in Figure 6.5, and the aggregated results in Table 6.2. The mean prediction errors for the BRCS model are 2.6% and 4.1% for CV1 and CV2, respectively; for the MXL model these errors are of 5.6% and 8.2%, in the same order. The results show that for both schemes, the BRCS model outperforms the MXL model: the BRCS model performed better than the MXL model in 85% of the test cases for the schema CV1, and in 82% in CV2. The results are confirmed

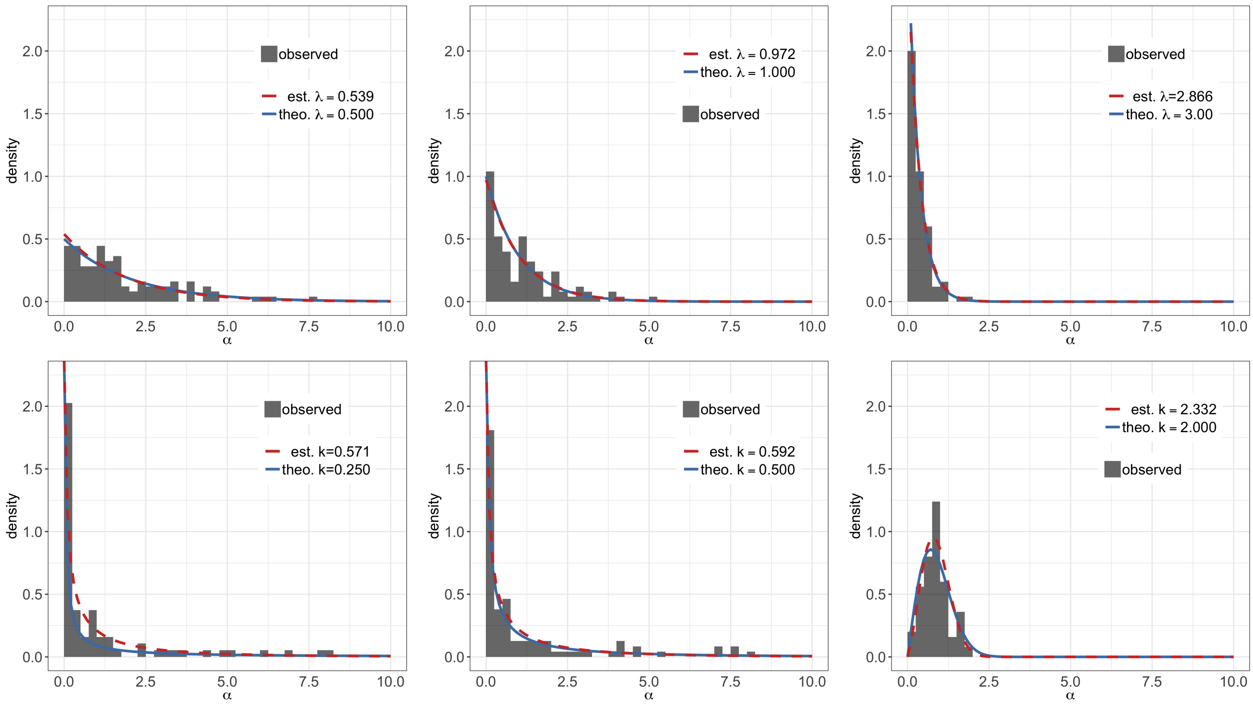

Figure 6.4: Observed vs estimated distributions for the values of α. The histogram depicts the distribution of the observed values, the blue line the theoretical distribution used to obtain the simulated values, and the red line is the estimated distribution.

by the paired Wilcoxon signed-rank test (Wilcoxon, 1945), which rejects the null hypothesis *H0: the prediction error for the MXL model is smaller than for the BRCS model* for both the CV1 and CV2 schemes with p-values $< 9.662e-10$ and $1.267e-09$, respectively. These results are expected since the data was generated following a boundedly rational model. However, there are some patterns that are worth describing as they illustrate the effect that the values of α have in the models performance. The first is that when the distribution of α accumulates values closer to zero (small values for k), the MXL model predictions get worse. This is more evident in the CV2 case, in which in some test cases the predictions of the MXL model are considerably higher (more than 15%). On the contrary, when the values of α tend to be larger, the accuracy of the two models becomes similar. This is explained because larger values of α imply that fewer alternatives are removed from the choice sets and thus the effect of z on the choice probabilities diminishes. A second pattern is that the prediction errors of both models are higher in CV2 than in CV1, an expected outcome, since less information about the distribution of α is contained in the training sets in CV2. This result suggests that observing a few choices of many individuals has a larger positive effect on the prediction accuracy of the models, compared to observing many choices from a small number of individuals. As a final thought in this section, it can be said that when a bounded rational process is suspected, then the BRCS model will both reduce the prediction error and entail to the correct interpretation of the coefficients.

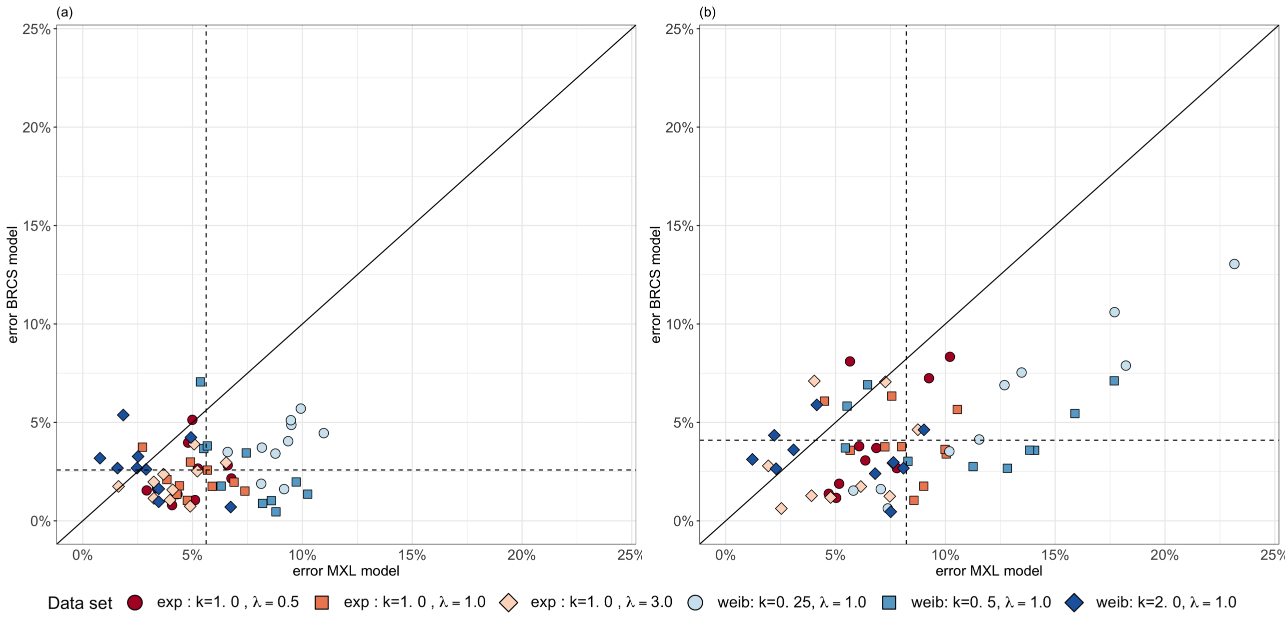

Figure 6.5: Comparison of the prediction error of the MXL model against the BRCS model for the (a) CV1 and (b) CV2 schemes. Each point represents the error of each of the 60 test sets (10 for each of the 6 generated data sets). The dotted lines are the average for each model.

Table 6.2: Prediction error for each of the six synthetic data sets. The Wilcoxon test is the one-tailed test with *H0: the prediction error for the MXL model is smaller than for the BRCS model.*

	CV1				CV2			
	BRCS		MXL		BRCS		MXL	
Data set	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
$k = 1.0,\ \lambda = 0.5$	0.0231	0.0137	0.0482	0.0124	0.0413	0.0275	0.0671	0.0185
$k = 1.0,\ \lambda = 1.0$	0.0208	0.0082	0.0507	0.0141	0.0390	0.0173	0.0811	0.0195
$k = 1.0,\ \lambda = 3.0$	0.0200	0.0097	0.0415	0.0135	0.0306	0.0242	0.0544	0.0234
$k = 0.25,\ \lambda = 1.0$	0.0383	0.0132	0.0900	0.0119	0.0574	0.0414	0.127	0.0559
$k = 0.5,\ \lambda = 1.0$	0.0254	0.0200	0.0758	0.0180	0.0446	0.0170	0.111	0.0445
$k = 2.0,\ \lambda = 1.0$	0.0273	0.0142	0.0307	0.0173	0.0327	0.0147	0.0520	0.0290
Total	0.0258	0.0145	0.0562	0.0249	0.0409	0.0258	0.0822	0.0437
Wilcoxon test	V = 99, p-value = 9.662e-10				V = 105, p-value = 1.267e-09			

6.4.2 Route choice experiments

The BRCS model is fitted to the data obtained in the route choice experiments carried out with the MDG. The data is the same as the one used in Chapter 5, considering exclusively participants who received travel time information. This is 3,664 choices from 353 participants. The results of the fitted model are compared to the results of the conditional, $Pr(y_{ij} = 1 \mid \boldsymbol{\beta}_i; \alpha_i, \mathbf{x_{ik}} \,\forall k)$ and the MXL models in Section 5.4.5. The variables contemplated in the systematic utilities of the three models are

- FRW_j: the % of freeway that composes the routes;
- DIR_j: the directness of the trip, defined as the Euclidean distance divided by the length of the route;
- TRN_j: the number of turns per kilometre;
- INT_j: the number of intersections per kilometre;
- $ITT_{ij}/ITT_{i(1)}$: relative travel time difference between route j and the fastest route.

Analogous to the analysis of the synthetic data, the model specification in the BRCS does not include the variable that determines the satisficing sets, given, in this case, by the relative travel time information $z_{ij} = (ITT_{ij} - ITT_{i(1)})/ITT_{i(1)}$. The BRCS model is estimated for Weibull distributed indifference bands α. The reason is that this distribution allows for values close to zero, which may better capture the indifference band of the perfect rational participants, i.e., the participants who always chose the fastest route. The model's estimates are presented in Table 6.3. For the sake of completeness, the estimates for the MXL and the conditional model, $Pr(y_{ij} = 1 \mid \boldsymbol{\beta}_i; \alpha_i, \mathbf{x_{ik}} \,\forall k)$ of Section 5.4.5 are also included in the table.

The estimation results show that, apart from $\hat{\beta}_{\%\Delta ITT}$, the estimated values for the mean of the coefficients are similar in the three models. The travel time information does not enter the BRCS model and it is small (in absolute value) in the conditional MXL, which means that the role played by the other four route attributes in the participants' choices is greater in these two models than in the MXL model. This is the expected behaviour in a bounded rational model, where within the indifference band the travel time information does not influence the choices of travellers and thus the rest of the attributes have a higher relevance. Suggesting, therefore, that the BRCS model is capable of inferring the distribution of the indifference bands. In this case, the indifference bands follow a Weibull distribution with shape parameter $\hat{k} = 0.736$ and scale parameter $\hat{\lambda} = 1.032$, which has a mean value of 1.159 and accumulates 10% and 50% of its values at 0.045 and 0.583, respectively. Hence, according to the estimated BRCS model, 10% of the decision-makers do not consider routes with travel times 4.5% higher than in the shortest-time route, and that half of decision-makers do not consider travel times 58% higher. The mean decision-maker will still consider routes twice longer than the shortest-time route, nonetheless, this interpretation may be misleading since the Weibull distribution with $0 < k < 1$ is heavy-tailed, moving its mean value to the right of the distribution (in this

Table 6.3: Estimation results of the BRCS, the MXL and the conditional model. The standard deviations and the correlation coefficients are shown in parenthesis.

	BRCS		MXL		*Cond. MXL*	
Coefficient	mean	s.error	mean	s.error	mean	s.error
$\hat{\beta}_{FRW}$	0.824	0.337	0.862	0.313	0.622	0.340
$\hat{\beta}_{DIR}$	1.571	0.681	1.377	0.640	1.863	0.753
$\hat{\beta}_{TRN}$	0.032	0.122	0.012	0.108	0.061	0.132
$\hat{\beta}_{INT}$	-0.054	0.035	-0.044	0.029	-0.076	0.032
$\hat{\beta}_{\%\Delta ITT}$	-	-	-3.285	0.356	0.366	0.373
$\hat{\sigma}^2_{FRW}$	4.258 (2.064)	2.255	2.840 (1.685)	2.253	3.596 (1.896)	1.842
$\hat{\sigma}^2_{DIR}$	26.505 (5.148)	10.585	17.889 (4.230)	11.229	25.970 (5.096)	9.383
$\hat{\sigma}^2_{TRN}$	0.551 (0.742)	0.264	0.505 (0.711)	0.185	0.464 (0.681)	0.178
$\hat{\sigma}^2_{INT}$	0.085 (0.291)	0.021	0.052 (0.228)	0.011	0.058 (0.240)	0.014
$\hat{\sigma}^2_{\%\Delta ITT}$	-	-	17.121 (4.138)	3.230	10.867 (3.297)	2.575
$\hat{\sigma}_{FRW,DIR}$	9.518 (0.896)	4.261	4.660 (0.654)	4.628	7.470 (0.773)	3.564
$\hat{\sigma}_{FRW,TRN}$	0.994 (0.649)	0.602	0.389 (0.324)	0.478	0.543 (0.420)	0.451
$\hat{\sigma}_{FRW,INT}$	-0.259 (-0.432)	0.131	0.011 (0.028)	0.092	-0.031 (-0.069)	0.093
$\hat{\sigma}_{FRW,\%\Delta ITT}$	- -	-	0.116 (0.017)	1.843	0.882 (0.141)	1.596
$\hat{\sigma}_{DIR,TRN}$	2.376 (0.622)	1.323	1.175 (0.391)	1.025	1.622 (0.467)	1.057
$\hat{\sigma}_{DIR,INT}$	-0.773 (-0.517)	0.377	-0.340 (-0.352)	0.227	-0.486 (-0.397)	0.274
$\hat{\sigma}_{DIR,\%\Delta ITT}$	- -	-	9.290 (0.531)	3.853	8.525 (0.507)	3.696
$\hat{\sigma}_{TRN,INT}$	-0.107 (-0.496)	0.058	-0.029 (-0.180)	0.032	-0.025 (-0.151)	0.035
$\hat{\sigma}_{TRN,\%\Delta ITT}$	-	-	-0.530 (-0.180)	0.530	0.021 (0.009)	0.543
$\hat{\sigma}_{INT,\%\Delta ITT}$	-	-	-0.336 (-0.356)	0.160	-0.382 (-0.482)	0.150
k	0.736	0.075	-	-	-	-
λ	1.032	0.110	-	-	-	-
Deviance	6,165	37.280	5,679	56.951	5,066	41.598
DIC	6,848.8		7,205.0		5,919.8	

case 32% of the values of the distribution lie to the right side of the mean). The density and cumulative probability functions are shown in Figure 6.6.

The distributions of the indifference bands obtained here are now compared to those in Section 5.4.4. In that section, various definitions were used to compute the individual indifference bands: $\hat{\alpha}_i^{max}$, $\hat{\alpha}_i^{95}$ and $\hat{\alpha}_i^{50}$ (see Section 5.3.3 for the methodology). The distributions of these estimators are presented in Figure 6.7 along with the distribution found with the BRCS model. For ease of exposition, in the following analysis the indifference bands for the BRCS model (Weibull distributed) are denoted as α^{BRCS}. The first difference to be noticed is that the distributions of the individually estimated indifference bands accumulate values more rapidly than α^{BRCS}. This is explained, on the one hand, because $\hat{\alpha}_i = 0$ in approximately 10% of the individuals for $\hat{\alpha}_i^{max}$ and $\hat{\alpha}_i^{95}$; a behaviour that cannot be observed in the Weibull case, as it is a continuous distribution. On the other hand, the slower accumulation rate of α^{BRCS} is explained because of its heavy-tail. As a result, α^{BRCS} appears to be underestimating the

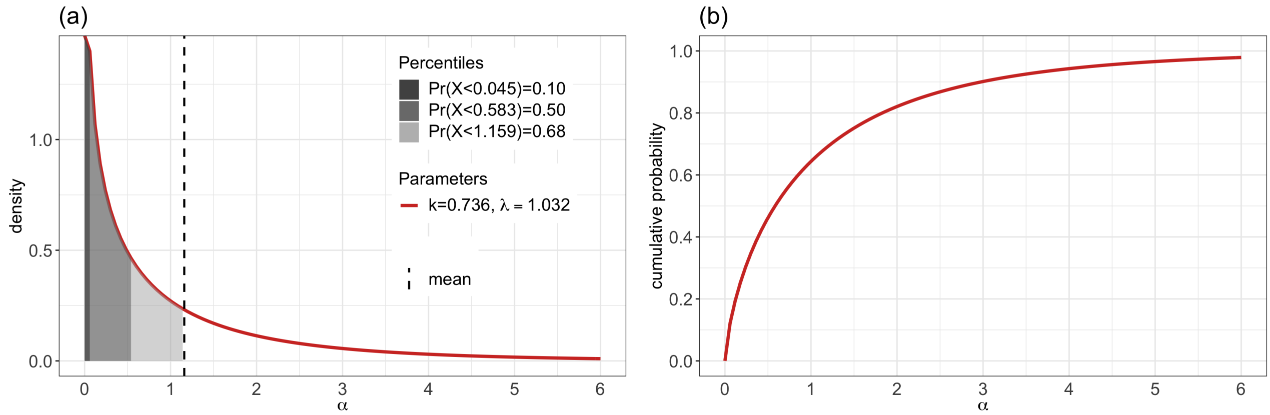

Figure 6.6: Estimated (a) density and (b) cumulative probability distributions for the indifference bands in the BRCS model.

number of perfect rational decision-makers, while overestimating the number of large indifference bands. The indifference band is a latent value that cannot be directly observed in the data, there is no ground truth with which to compare these estimates in order to determine which better approximates the real indifference bands. However, since the prediction errors of the BRCS model are systematically smaller than the Cond. MXL and MXL models, it can be said that the former explains better the observed choices. As exposed in the motivation to this study, the advantage of estimating the indifference bands individually is that they are directly inferred from each individual's choices and no parametric shape for its distribution is assumed, the downside of this approach is that the information is only partially observed, which may derive in biased estimates. Furthermore, not all the estimators of the individual indifference bands can be used as exogenous inputs to a discrete choice model. The reason is that in discrete choice models it is necessary for the actual choice to be part of the choice set, and the estimators $\hat{\alpha}_i$ such that $\hat{\alpha}_i \geq \hat{\alpha}_i^{max}$ are the only ones that guarantee the actual observed choices to be part of the choice sets in all decision problems (see Section 5.3.4). None of these problems are present in the BRCS approach, in which alternatives are never assigned a zero probability, and information from all decision-makers is taken into account to estimate the distribution of α. This last point is important when data is partially observed, since it takes into account information from the choices of all individuals to make inference for the individuals that are observed only in situations with small (or high) travel time differences. The disadvantage of the BRCS approach is that it is parametric, in the sense that a distribution family is assumed for the values of α.

The prediction accuracy of the three models is now analysed. As mentioned before, only the CV1 scheme is used to determine the training and test sets at each iteration. Since the number of choices per decision-maker is small (see Figure 5.2), CV1 guarantees the removal of decision-makers from the training sets: on average, 25 decision-makers, representing 7% of

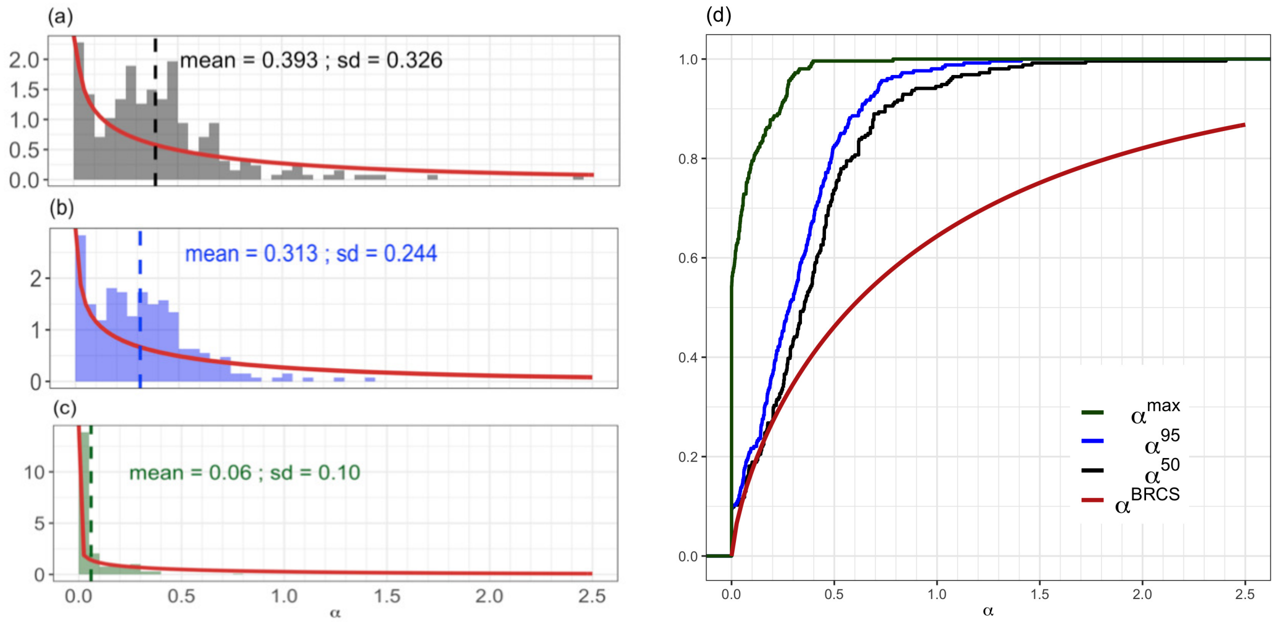

Figure 6.7: Distributions of the individually estimated indifference bands (a) $\hat{\alpha}_i^{max}$, (b) $\hat{\alpha}_i^{95}$ and (c) $\hat{\alpha}_i^{50}$; and the Weibull distribution inferred by the BRCS model (red line). (d) the cumulative probability distributions.

the individuals are removed from the training set in each iteration. Contrary to the synthetic data case, were the generated data corresponded to a unique OD pair, the data here comprises the choices over 41 OD pairs. Therefore, the prediction errors can be aggregated both at CV iteration level and at OD pair level. In the former, the errors in the 41 OD pairs are averaged for each CV iteration, while in the later the errors in each OD pair are averaged across iteration. The results are presented in Figure 6.8. It can be seen that the BRCS model's errors per iteration are smaller in the great majority of the cases, but the same is not true when the aggregation is at OD level. The average errors are 0.146, 0.150 and 0.155 respectively for the BRCS, the MXL and the Cond. MXL models. The paired Wilcoxon signed-rank test (one tail) confirms these results. At CV iteration level, the test rejects the null hypothesis *H0: the prediction error for the MXL (or Cond. MXL) model is smaller than for the BRCS model* with p-value 0.006836 (0.01367). However, when aggregating the errors by OD pair, the test does not reject the null hypotheses in either case, with p-value of 0.2243 for the test between the MXL and the BRCS models, and p-value of 0.1674 for the test between the Cond. MXL and the BRCS models. The higher prediction accuracy of the BRCS model when aggregating by iteration, but not by OD pair, can only be explained by a compensatory behaviour: large prediction errors in some OD pairs are compensated with low prediction errors in other OD pairs. Hence, it can be concluded that at network level, i.e., averaging the error of the 41 OD pairs, the BRCS model performs better than the other two options in predicting the choices of travellers, but at OD level this model represents no gain.

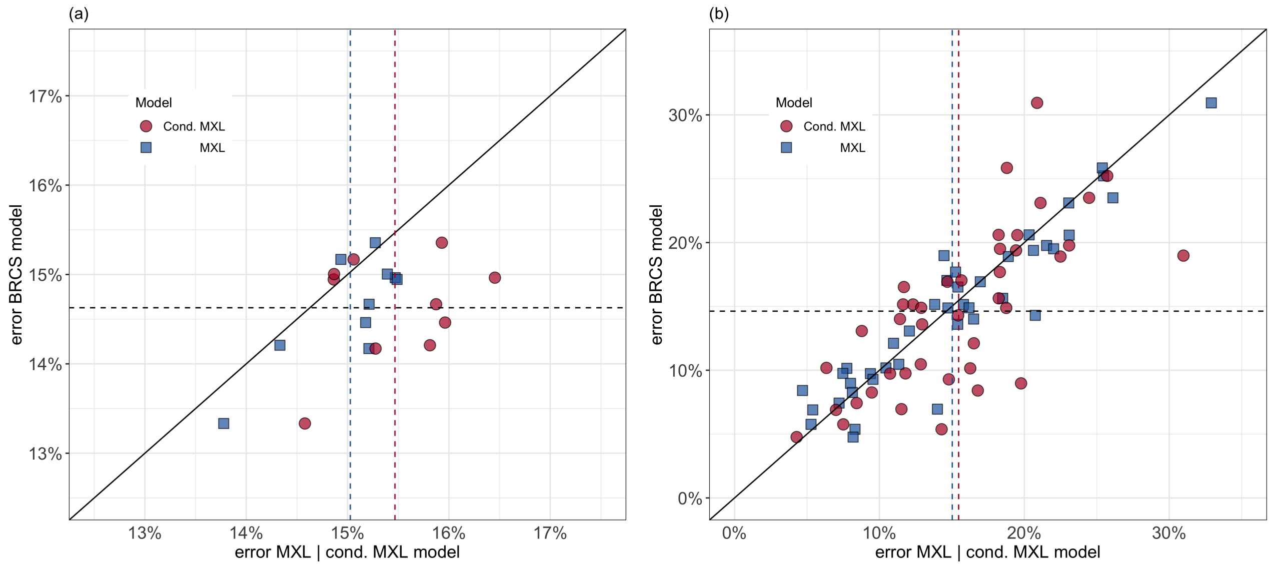

Figure 6.8: Comparison of the prediction error of the BRCS model against both the prediction errors of the MXL and cond. MXL models. (a) aggregated by CV iteration and (b) aggregated by OD pair.

6.5 Conclusions and discussion

In this chapter, a model for boundedly rational route choice model was proposed: the bounded rational choice set generation mixed logit model (BRCS). The model considers (i) a bounded rational choice set generation process in which the alternative routes with travel time differences above a certain threshold (indifference band) are removed from the available alternatives, and (ii) a rational choice process for the alternatives in the generated choice set. Both the choice set and the route choice are jointly estimated, allowing for the BRCS to implicitly infer the latent population's indifference band distribution. The choice set generation process assumes a parametric distribution for the indifference bands. This distribution can take any form, but in this work, Weibull distributed indifference bands were assumed. The ability of the model to infer the distribution of the indifference bands was tested using synthetic data. The prediction accuracy of the BRCS model was tested against the predictions of the MXL model and the MXL model conditional on exogenously estimated indifference bands using real data, coming from a series of computer route choice experiments carried out with the MDG.

The results of the estimation for the synthetic data show that the BRCS model is capable of inferring the mean parameters of the underlying distribution of the indifference bands. This result is specially evident when the indifference bands are highly discriminative, i.e., when they are small (close to zero) and thus they discard more alternatives from the choice sets. In such cases, the BRCS model was capable of inferring the real parameters of the choice model, contrary to the MXL model (mispecified model). This suggests that when a bounded rational process generates the data, the interpretation of the coefficients of the MXL model may be misleading.

For the real data, the BRCS model was compared to both the MXL and the conditional MXL model of Chapter 5. The estimated coefficients for the route attributes (other than the travel time) are similar for the three models. However, these attributes play a larger role in the BRCS and the conditional MXL model than in the MXL model. The estimated (Weibull) distribution for the indifference bands has a 10th percentile value of 4.5% and a median value of 58%. This means that 10% of the participants did not consider routes with travel times 4.5% higher than the shortest-time route, and half of them did not consider routes with travel times 58% slower. These estimates are higher than those obtained by individually estimating the indifference bands of the participants. The prediction accuracy of the BRCS model showed slightly better results than both the MXL and conditional models, suggesting, on the one hand, the existence of a boundedly rational process in the collected route choice data and, on the other hand, that the estimates of the indifference band distribution of the BRCS adjust better to the *real* distribution than the individually estimated $\hat{\alpha}_i^{max}$. Even though the prediction accuracy gains of the BRCS model are moderate compared to the MXL model, they are systematic. Moreover, the method is able to capture the boundedly rational behaviour of travellers, and thus deserves further investigation. For example, by testing different family of distributions for the indifference bands, including mixed distributions that are capable of capturing indifference bands equal to zero (perfect rational decision-makers). The choice of the underlying distribution of the indifference band may improve significantly the prediction accuracy of the model.

Conclusions of part I: route choice model selection

Part I of this thesis is composed of three research chapters concerned with the study of travellers' route choice behaviour through computer experiments. In Chapter 4, a methodology to perform route choice experiments was proposed. The methodology is about determining a suitable subset of OD pairs and routes to use in the route choice experiments, such that the responses of participants with respect to predefined routes and attributes in this subset can be generalised to all situations in the network. For the city of Lyon in France, 9 OD pairs and their three connecting routes were used as representative of 624,490 OD configurations. These nine representative OD configurations cover around 83% of the values of the attributes of the OD-routes in the network. In Chapter 5, the influence that travel time information has on the route choices of travellers was investigated, from the perfect rational and bounded rational perspectives. It was found that travellers minimise travel time with respect to a reference point, given in this case by the travel time in the fastest route, and that the reference point is context-dependent, since it is evaluated in each route choice problem. On that chapter, evidence was found that points to a bounded rational behaviour in route choice: only a small percentage of participants (10%) chose always the fastest route, and the average participant did not consider routes with travel time differences 1.3 times slower than the fastest alternative. Heterogeneity in the indifference bands was observed. In Chapter 6, these findings are translated into a discrete choice model (BRCS) that considers (i) a bounded rational choice set generation process in which the alternative routes with travel time differences above a certain threshold (indifference band) are removed from the available alternatives, and (ii) a rational choice process for the alternatives in the generated choice set. The BRCS model resulted superior in predicting the choices of travellers in simulated and real data. An important limitation in the study of route choice in this thesis is that, in its current state, the MDG does not allow to study the learning process of travellers. This was discussed in Section 5.2. Therefore, the study of route choice in this thesis relies heavily on the travel time information given to the participants (close to the perfect information scenario). Travel time information is used as a proxy to study how travellers react to the travel time in the different routes. However, in real-life settings travellers base their choices on the perceptions they construct from both experience and information. In this sense, travel time uncertainty and variability, as well as the accuracy of the information have an important role on their choices. This is an important limitation, since the interpretation needs to be done considering the travel time information and not the perceived information.

As a closing section of part I of this thesis, model selection is carried out by putting into practice the lessons learned from the previous chapters. The objective is to find the route choice model that better predicts the choices of the participants, while maintaining its representativeness to the whole network. In order to guarantee this, the models are estimated with a balanced version of the data, obtained by considering (i) the importance of each cluster to the network, and (ii) the importance of each OD pair in a given cluster. The original data set comprises of 5,535 choices from 496 participants in 41 OD pairs; the balanced data set comprises of 8,384 choices by the same number of participants in the same 41 OD pairs. The details on how the balanced data set was obtained are presented at the end of this section. All models were estimated considering the participants who received and did not receive information on the travel time. As in the previous chapters, the explanatory variables that enter the models are:

- DIR_j, the directness of the route j, defined as the length of j divided by the euclidean distance between origin and destination;
- TNR_j, the number of turns per kilometre in the route j;
- FRW_j, the percentage of freeway that composes the route j;
- INF_i, binary variable indicating if participant i received information;
- ITT_{js}, the informed travel time in the route j in OD pair and moment s, the variable is normalised by dividing the informed travel time between the minimum informed travel time in the choice situation s;
- INF_i, binary variable that indicates if participant i received the travel time information;
- $FAST_{js}$, binary variable indicating if route j is the fastest in situation s.

For ease of exposition, denote the alternative specific variables (those indexed only by j) by the vector $\mathbf{x_j}$. From the six models that are tested, four are MXL models with different specifications of the systematic part of the utility and correlation structures. The fifth and the sixth models are the MXL-Cond. and BRCS models introduced in Chapter 6. The former considers an exogeneously estimated indifference band in order to constitute the choice set, the latter estimates the distribution of the indifference bands endogenously. The models' specifications are given by:

$$
\begin{aligned}
\text{MXL-1: } & V_{ijs} = \mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i} + \beta_{i,ITT} \cdot ITT_{js} \cdot INF_i & \beta_{i,p} \sim \mathcal{N}(\beta_p, \sigma_p^2) \\
\text{MXL-2: } & V_{ijs} = \mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i} + \beta_{i,ITT} \cdot ITT_{js} \cdot INF_i & \boldsymbol{\beta}_\mathbf{i} \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta) \\
\text{MXL-3: } & V_{ijs} = \mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i} + (\beta_{i,ITT} \cdot ITT_{js} + \beta_{i,FAST} \cdot FAST_{js}) \cdot INF_i & \boldsymbol{\beta}_\mathbf{i} \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta) \\
\text{MXL-4: } & V_{ijs} = \mathbf{x_j}^T (\boldsymbol{\beta}_\mathbf{i} + \boldsymbol{\gamma} \cdot INF_i) + \beta_{i,ITT} \cdot ITT_{js} \cdot INF_i & \boldsymbol{\beta}_\mathbf{i} \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta) \\
\text{MXL-Cond.: } & V_{ijs} = \mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i} + \beta_{i,ITT} \cdot ITT_{js} \cdot INF_i & \boldsymbol{\beta}_\mathbf{i} \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta) \\
\text{BRCS: } & V_{ijs} = \mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i} & \boldsymbol{\beta}_\mathbf{i} \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_\beta)\,.
\end{aligned}
$$

Note that all models, except for the MXL-1 model, consider the correlations between the coefficients $\boldsymbol{\beta}_\mathbf{i}$. Model MXL-3 introduces a binary variable indicating the fastest route in each choice problem. Model, MXL-4 introduces the coefficients $\boldsymbol{\gamma}$, which are estimated for all the participants (they are not indexed by i). The coefficients $\boldsymbol{\gamma}$ allow the mean preferences for the attributes to be different between the uninformed and the informed participants: the preferences for the not informed participants are given by $\boldsymbol{\beta}_\mathbf{i}$, whereas for the informed participants are given by $\boldsymbol{\beta}_\mathbf{i} + \boldsymbol{\gamma}$. The estimated parameters for these models are included in Table 6.4. As it was discussed in Section 6.4.1, the deviance of the models (a measure of goodness-of-fit) and the deviance information criterion (DIC) are not useful for model selection for these models. Therefore, model selection needs to be based on the predictive accuracy for out-of-sample observations. The prediction accuracy is assessed with 10 iterations of bootstrapping without replacement. At each iteration, two-thirds of the observations are sampled and used as training set; the predictions of the models are tested on the remaining one-third of the observations. The mean prediction error of the 10 iterations is computed for each model; the results are shown in Figure 6.9.

The first pattern to be noticed is that all models show a higher accuracy in predicting the choices of the participants who received travel time information, with a mean prediction error difference between the groups of around 5%. This result is explained because travel time information is a more important variable in the choices of travellers and, thus, it has a higher explanatory power. Even though the models' specification for the travellers that received no information is the same in all the models (i.e., $\mathbf{x_j}^T \boldsymbol{\beta}_\mathbf{i}$), the predictive error for the travellers without information vary across them. This is explained because the both the choices of the informed and the not informed travellers contribute to the estimates of the parameters of $\boldsymbol{\beta}_\mathbf{i}$. For the not informed participants, the models that show the highest accuracy are the MXL-4 and MXL-Cond. models. In the case of the MXL-4 model, the improvement is because the coefficients of the route specific variables (% of freeway, directness, no. of turns and intersections per kilometre) are treated differently for the informed and not informed participants.

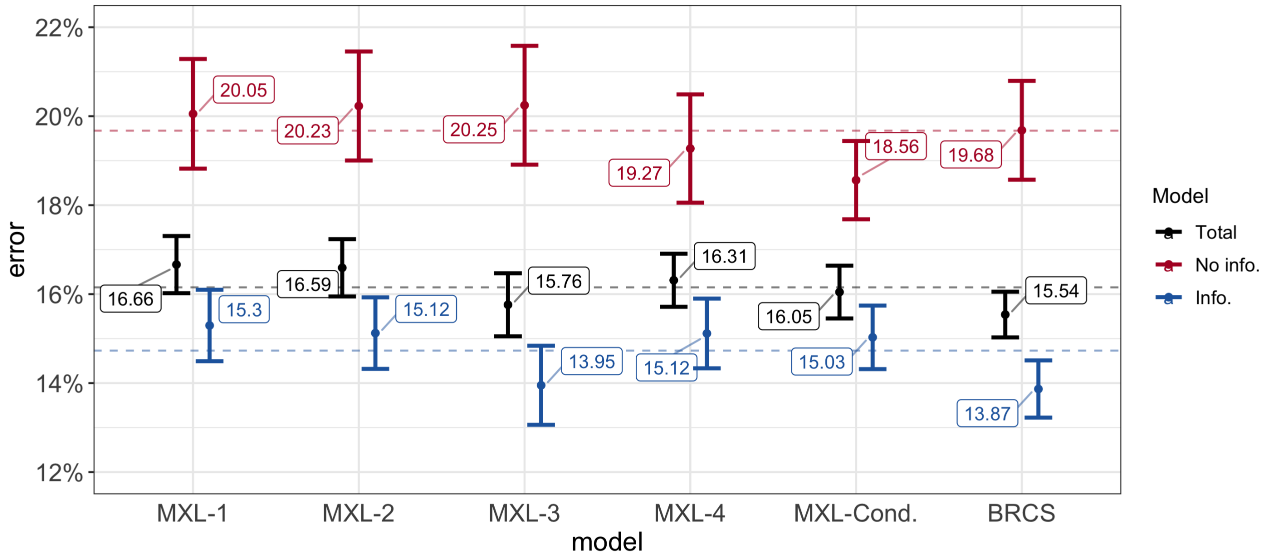

Figure 6.9: Mean and standard deviation of the prediction error for the six route choice models.

In the case of the MXL-Cond. model, the higher accuracy for the participants who did not receive information could be explained by the smaller role of travel time (coefficient smaller in magnitude) within the alternatives that are considered in the choice set. The BRCS model, which considers a bounded rational behaviour, is the one with the highest prediction accuracy, which is mainly explained by the predictions for the participants that received information. For the participants who did not receive information, this model has the average accuracy. The MXL-3 model has the second highest overall accuracy, which as in the BRCS case, it is due to a smaller error for the informed participants. Nonetheless, the MXL-3 model also shows the larger error for the not informed participants, evidencing a compromise in accuracy between the two groups. The specification of the MXL-3 model includes a variable that indicates when a route was the fastest. Therefore, supporting the finding discussed in the conclusion of Chapter 5: the fact of informing a route to be the fastest increases its probability to be chosen. The mean accuracy of the best model obtained here, the BRCS model, is 15.54%. Since this error measure has a direct interpretation, it means that the BRCS model will assign correctly around 85% of the trips in the network. The above observations can be summarised as:

- there is a bounded rational behaviour in the choice set generation process of travellers;
- no matter the travel time difference between the routes, there is an effect of informing that a route is the fastest;
- the preferences of travellers between the route attributes and the travel time are correlated;

- receiving travel time information changes the preferences of travellers for the physical attributes of the routes;
- the best route choice model in this thesis, the BRCS model, will assign correctly around 85% of the trips in the network.

Methodology for balancing the data

For the model selection, the data collected in all the MDG experiments is combined. However, the number of observations in each OD pair varies significantly. This implies that if the observations are all taken equally into account, then the estimated choice models will be biased towards the OD pairs with more elements. To avoid this issue, the observations need to be weighted. To this purpose, there are two aspects that need to be taken into account when weighting the observations:

1. **The importance of each cluster to the network**. The clusters obtained in Section 4 differ in the number of elements that they contain. In order to estimate a model that approximates well the route choices in a randomly selected OD pair in the whole network, then each cluster is weighted according to its number of elements. The weight of the cluster k in the data set is
$$w'_k = \frac{N_k}{\sum_{l=1}^{9} N_l}\,,$$
where N_a is the number of elements in cluster a.

2. **The importance of each OD pair in a cluster**. The most important OD pair in a cluster, in terms of its representativeness, is its centroid. The importance of the rest of the OD pairs in the data set can then be considered as a function of their distance to the centroid. Define the importance of an OD pair od as its similarity with the centroid of the cluster that contains it. That is,
$$s_{k,od} = \begin{cases} 1 - \frac{\mathrm{dist}(od,c^*)}{\max_{a\in OD_k}\{\mathrm{dist}(a,c^*)\}}, & od, c^* \in OD_k \\ 0, & \text{otherwise}\,, \end{cases}$$
where c^* is the centroid of cluster k and OD_k the set of OD pairs in cluster k. The weight of the importance of an OD pair is obtained by normalising for the different number of OD pairs that are found in the data set, i.e.,
$$w''_{k,od} = \frac{s_{k,od}}{\sum_{a\in OD_k} s_{k,a}}\,.$$

Considering the above points, the total weight that an OD pair in the data set must have is therefore given by

$$w_{k,od} = w'_k \times w''_{k,od} \,. \tag{6.10}$$

Note that $\sum_k \sum_{od} w_{k,od} = 1$, since $\sum_{od \in OD_k} w''_{k,od} = 1$ and clearly $\sum_k w'_k = 1$.

In this thesis, the inference method to estimate the choice models is bayesian. To be specific, a Gibbs sampler is used to sample the posterior distribution of the model's parameters. Contrary to maximum likelihood estimation, where the weights enter the model by multiplying the contribution of each observation to the likelihood, in bayesian methods there is no easy way to weight the observations. The solution in this case is to resample the data so that the number of choices in each OD attain the necessary number of observations, given by the weights. The actual number of observations in the different OD pairs and the number needed to respect the weights $w_{k,od}$ are presented in Fig. 6.10 (the weights are multiplied by N, the number of choices in the data set). It can be seen that some OD pairs, in particular those of the Lyon-36V network, are over represented in the data set, while other OD pairs are sub represented. This implies that there are cases in which it is necessary to resample the observations, while in other cases it is necessary to remove them. However, with this strategy many observations would be lost. On the other hand, it is possible to find a number of observations for the resampled data set N^*, such that no original observations are lost. The problem with this strategy is that the number of resampled observations needs to be high for certain OD pairs. Moreover, since the observations are made by participants, this later strategy would lead to some participants with many repetitions. From this analysis, it can be concluded that a *good* data set size N^* would be such that reduces the amount of lost information, but without creating many new observations. The size of the data set used to resample and remove observations is $N^* = 1.5N$. The effect of choosing this number is also shown in Fig. 6.10. It is worth mentioning that the observations that were removed are from the repetitions of participants within the same OD pair, and that no participant were completely removed from any of the OD pairs.

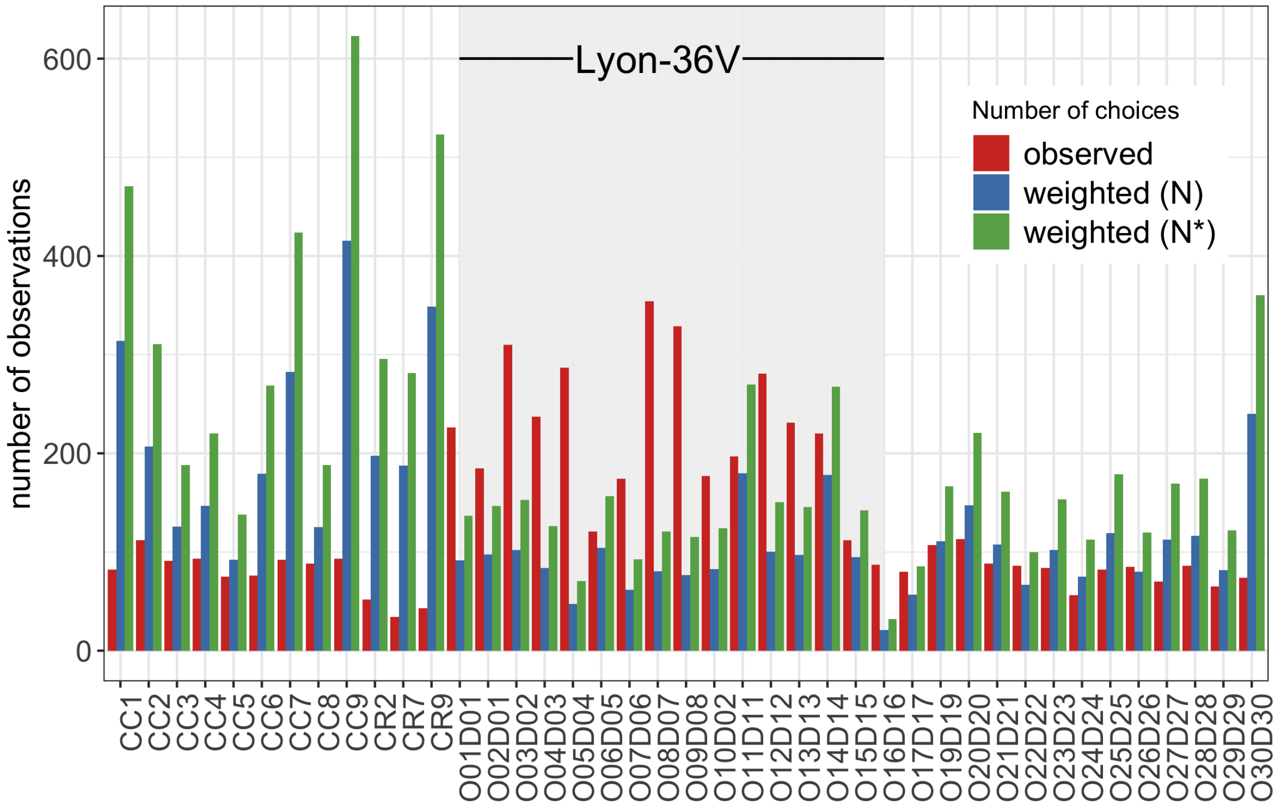

Figure 6.10: Number of observations in the data set. The actual number of observations (N in total) in each OD pair is compared against the weighted number of observations considering N and $1.5N$ observations in the resampled data set.

Table 6.4: Estimates for the mean and standard error of the parameters of the six models. The standar deviation and the correlations of the coefficients $\boldsymbol{\beta}_\mathbf{i}$ are shown in parenthesis.

	MXL-1		MXL-2		MXL-3		MXL-4		MXL-Cond.		BRCS	
Parameter	mean	s.error	mean	s.error	mean	s.error	mean	s.error	mean	s.error	mean	s.error
$\hat{\beta}_{FRW}$	0.57	0.30	0.55	0.34	0.37	0.34	1.02	0.53	0.72	0.38	0.63	0.43
$\hat{\beta}_{DIR}$	4.01	0.56	3.94	0.64	3.88	0.68	6.25	0.98	4.71	0.68	5.46	0.77
$\hat{\beta}_{TRN}$	0.04	0.12	0.10	0.13	0.09	0.14	-0.02	0.21	0.15	0.14	0.18	0.17
$\hat{\beta}_{INT}$	-0.15	0.03	-0.16	0.03	-0.16	0.04	-0.15	0.06	-0.15	0.04	-0.23	0.04
$\hat{\beta}_{ITT}$	-4.31	0.51	-4.11	0.50	-2.94	0.48	-4.41	0.50	2.68	0.54	-	-
$\hat{\beta}_{FAST}$	-	-	-	-	0.18	0.15	-	-	-	-	-	-
$\hat{\gamma}_{FRW}$	-	-	-	-	-	-	-0.82	0.66	-	-	-	-
$\hat{\gamma}_{DIR}$	-	-	-	-	-	-	-3.81	1.26	-	-	-	-
$\hat{\gamma}_{TRN}$	-	-	-	-	-	-	0.19	0.25	-	-	-	-
$\hat{\gamma}_{INT}$	-	-	-	-	-	-	-0.01	0.07	-	-	-	-
$\hat{\sigma}^2_{FRW}$	13.44 (3.67)	2.20	21.77 (4.67)	3.73	20.20 (4.49)	3.80	21.41 (4.63)	3.56	25.49 (5.05)	4.56	40.61 (6.37)	6.46
$\hat{\sigma}^2_{DIR}$	56.49 (7.52)	8.87	81.98 (9.05)	13.33	90.03 (9.49)	14.80	75.65 (8.70)	12.75	77.27 (8.79)	13.26	142.14 (11.92)	21.94
$\hat{\sigma}^2_{TRN}$	2.51 (1.58)	0.44	3.07 (1.75)	0.60	3.60 (1.90)	0.75	3.13 (1.77)	0.61	3.05 (1.75)	0.60	5.98 (2.44)	1.16
$\hat{\sigma}^2_{INT}$	0.13 (0.36)	0.02	0.26 (0.51)	0.04	0.29 (0.53)	0.05	0.26 (0.51)	0.04	0.28 (0.53)	0.05	0.39 (0.62)	0.06
$\hat{\sigma}^2_{ITT}$	53.25 (7.30)	8.19	48.26 (6.95)	7.61	31.95 (5.65)	6.94	47.58 (6.90)	7.17	36.26 (6.02)	6.44	-	-
$\hat{\sigma}^2_{FAST}$	-	-	-	-	3.55 (1.88)	0.74	-	-	-	-	-	-
$\hat{\sigma}_{FRW,DIR}$	-	-	13.71 (0.32)	4.91	13.93 (0.33)	5.44	12.21 (0.30)	4.67	8.42 (0.19)	5.34	33.85 (0.45)	8.55
$\hat{\sigma}_{FRW,TRN}$	-	-	1.59 (0.19)	0.97	1.02 (0.12)	1.06	1.49 (0.18)	0.96	1.88 (0.21)	1.09	5.63 (0.36)	1.78
$\hat{\sigma}_{FRW,INT}$	-	-	1.05 (0.44)	0.29	0.98 (0.41)	0.28	1.07 (0.45)	0.29	1.31 (0.49)	0.34	1.10 (0.28)	0.42
$\hat{\sigma}_{FRW,ITT}$	-	-	-1.47 (-0.05)	3.83	2.16 (0.08)	3.70	-1.90 (-0.06)	3.56	1.35 (0.04)	3.80	-	-
$\hat{\sigma}_{FRW,FAST}$	-	-	-	-	-0.85 (-0.10)	1.19	-	-	-	-	-	-
$\hat{\sigma}_{DIR,TRN}$	-	-	3.16 (0.20)	1.98	3.36 (0.19)	2.26	2.98 (0.19)	1.91	2.21 (0.14)	1.92	11.24 (0.39)	3.74
$\hat{\sigma}_{DIR,INT}$	-	-	-2.10 (-0.45)	0.60	-2.62 (-0.52)	0.68	-2.06 (-0.46)	0.58	-2.40 (-0.52)	0.61	-3.74 (-0.50)	0.94
$\hat{\sigma}_{DIR,ITT}$	-	-	12.27 (0.20)	7.60	10.43 (0.19)	7.80	11.82 (0.20)	6.98	15.91 (0.30)	6.73	-	-
$\hat{\sigma}_{DIR,FAST}$	-	-	-	-	-3.79 (-0.21)	2.55	-	-	-	-	-	-
$\hat{\sigma}_{TRN,INT}$	-	-	-0.23 (-0.25)	0.12	-0.29 (-0.28)	0.14	-0.24 (-0.26)	0.12	-0.14 (-0.15)	0.12	-0.55 (-0.36)	0.20
$\hat{\sigma}_{TRN,ITT}$	-	-	-1.38 (-0.11)	1.56	-1.01 (-0.09)	1.67	-1.46 (-0.12)	1.56	0.07 (0.01)	1.49	-	-
$\hat{\sigma}_{TRN,FAST}$	-	-	-	-	0.20 (0.06)	0.61	-	-	-	-	-	-
$\hat{\sigma}_{INT,ITT}$	-	-	-0.37 (-0.10)	0.39	-0.20 (-0.07)	0.39	-0.39 (-0.11)	0.38	-0.58 (-0.18)	0.38	-	-
$\hat{\sigma}_{INT,FAST}$	-	-	-	-	0.19 (0.19)	0.14	-	-	-	-	-	-
$\hat{\sigma}_{ITT,FAST}$	-	-	-	-	2.51 (0.24)	1.57	-	-	-	-	-	-
$\hat{k}$	-	-	-	-	-	-	-	-	-	-	0.43	0.07
$\hat{\lambda}$	-	-	-	-	-	-	-	-	-	-	0.46	0.07

Part II

Route and departure time choice behaviour

7 Route and departure time choice behaviour

Considering an uni-modal network of car trips, route and departure time are two of the main decisions that travellers make to complete a trip. At an aggregated level, this implies that the traffic patterns in the uni-modal network are mainly explained by the sum of these two individual choices. In this chapter, the route and departure time choice behaviour of travellers is investigated. To this purpose, an experiment that considers both decisions simultaneously was carried out using the MDG platform. The objective is to understand which variables influence the joint route and departure time choices of travellers, and to test the appropriateness of the MXL model to explain and predict these choices. The joint model proposed here introduces time-dependent correlations in the specification of a MXL model. It is important to mention that this is an investigation in progress, and that the results are not definitive nor complete.

7.1 Motivation

There are two difficulties that arise when considering the joint route and departure time choices. The first is related to the dependency that the route and departure time choices have on the travel time, which does not allow to decompose the joint route and departure time choice probability in the product of two simpler probabilities. This can be formally argued treating the dependencies of these variables as a probabilistic graphical model (Barber, 2011). Let DT, TT and R be random variables denoting, respectively, the departure time choice, the travel time, and the route choice. The dependency structure between these variables is given by $DT \rightarrow TT \leftarrow R$ or, in other words, the travel time depends on both the route and departure time choices. The above structure is known as the *v-shape* in probabilistic graphical models and implies that if the route and departure time choices are known, then the travel time is (theoretically) fully determined. This assumption makes sense since travel time is the result of both departure time and route choices. Modelling the choices of travellers (from a RUM perspective), is to find an expression for the choice probability $Pr(R, DT \mid TT)$, where TT is the known (or independent) variable. However, the *v-shape* dependency structure between the variables imply that if TT is given, then R and DT are dependent. As a consequence, the joint probability $Pr(R, DT \mid TT)$ cannot be decomposed as the product $Pr(R \mid TT)Pr(DT \mid TT)$. Therefore, the joint route and departure time choices must be necessarily considered simultaneously. The second difficulty comes from the continuous nature of time which, by definition, cannot be the outcome of discrete choice models. This point is discussed below.

Two major approaches are found in literature to model departure time choices, depending on whether departure time is treated as a continuous or discrete variable. To the first category belong, for example, the works of Bhat and Steed (2002), Nurul Habib et al. (2009), Habib (2013), who propose to model the continuous departure time using hazard models. Hazard models belong to a branch of statistics, *survival analysis*, in which the expected time duration until an event occurs (in this case the trip departure) is analysed (Hougaard, 1999). Hazard models are appealing for their simplicity and because they can be stated as a regression problem, allowing for the inclusion of independent variables to explain the time of departure. Moreover, a large literature on these models exists, coming mainly from actuarial sciences, and hence the methods to estimate hazard models and their properties are well-known. Nonetheless, hazard models lack of a behavioural basis, thus preventing their interpretation from a behavioural angle. This shortcoming is not present when travel time is discretised and modelled as the outcome of a discrete choice model. However, this latter approach has also some drawbacks, consequence of the discretisation of the continuous departure time. There is a natural correlation between the alternatives that are close in time, thus, it makes sense to consider the choice probability of an alternative route j in time interval t to be correlated with the choice probability of j in time $t-1$. These correlations can be modelled in discrete choice models. Small (1987) developed the Ordered Generalised Extreme Value (OGEV) model that considers correlations between ordered alternatives (such as time intervals), in which alternatives that are close in time have a strong correlation and the correlation diminishes with more distant alternatives. Bhat (1998) used an OGEV model to analyse the travel time and mode choices in urban shopping trips. Complex correlation structures can be also specified in the more general mixed logit model. Bajwa et al. (2009) and de Jong et al. (2003) used a mixed logit model to study the mode and departure time choices. Notwithstanding the capability of the MXL model to specify complex correlation structures, there are still some issues related to the discretisation of the departure time, as pointed out by Bhat (1998). Notably, the selection of the time interval size and the correlations of departure times close to the boundaries of the intervals.

As exposed above, there is a trade-off between the behavioural justification when considering a continuous departure time model and the biases introduced when discretising this variable. The approach followed in this chapter is to model the joint route and departure time as a discrete choice, using a MXL model, thus, maintaining its behavioural interpretability. For this purpose, the departure time is discretised in intervals of five minutes. Different correlation structures are tested in order to see the effect that they have in the predicted choices. The data used in this chapter comes from computer experiments using the MDG platform, designed to observe both the route and departure time choices of participants. 1,145 choices of 177 participants were recorded over five OD pairs connected by three alternative routes. In the experiments, all participants received travel time information on the three alternatives routes, but contrary to the route choice experiments, the information changes depending on the departure time of the trip.

7.2 Methodology

7.2.1 Route and departure time experiment

The experiment consists on decision problems in which the participants made decisions on both the route and the departure time to complete a home-work trip. The experiments are placed on a simulated environment of the Lyon-full network from 6:45 to 9:00 hours, recording the choices of participants on 5 OD pairs connected by three routes: CC1, CC2, CC5, CC7, and CC8 (see appendix 2.A and appendix 2.B). At the beginning of each choice problem, the objective arrival time is informed to the participants. The experiments are similar to those of route choice, with the difference that the travel time information of each route changes with the selected departure time, i.e., the travel time information depends on both the departure time period and the route. Thus, a choice problem in the experiment requires participants to test different departure times to see the travel time information in the different routes. Once a desired departure time and route are found, the participant makes the choice. As in the route choice experiments, the choices are mainly based on the travel time information. In this experiment, all participants received travel time information for the three alternative routes. After a trip is finished, participants received a score which depends on the arrival time. The score follows a trapezoid shape (see Figure 2.6) that penalises early and late arrivals. The experiment was configured such that if no change on the departure time is made, then the trip will be late for more than 10 minutes, forcing participants to choose a departure time if they want to achieve the objective. This was done in order to discriminate not engaged participants: participants who did not changed the departure time from its initial value are not taken into account. For the experiments, 177 students of the University of Lyon were recruited. In total, 1,145 choices were collected, making it an an average of 6.5 choices per participant.

7.2.2 Definition of the variables related to the time for travel

Intervals of size of 5 minutes are considered in the discretisation of the departure time. As a result, each alternative in a choice problem corresponds to a departure time interval $[t_1, t_2]$ and a route j, i.e., it corresponds to the pair $s = (j, [t_1, t_2])$. The departure time choice enables travellers to choose how early they desire to arrive to their destination. Therefore, the expected earliness (lateness) of the trips can be considered as an explanatory variable of their choices. Define DT_s, AT^*_s and ETT_s as the departure time, the objective arrival time and the expected travel time of the trip s. The variable ETT here is equivalent to the travel time information (ITT) in the previous chapters. The expected time at arrival is given by $ETA_s = DT_s + ETT_s$. This latter value represents the time at which participants intend to arrive to the destination. The expected lateness, $ELATE_s$, can be then obtained as the difference between the objective and the expected arrival time, i.e., $ELATE_s = ETA_s - AT^*_s$, thus $ELATE_s > 0$ are the expected late arrivals and $ELATE_s < 0$ the expected early arrivals. Since late arrivals have worse consequences than early arrivals (see Figure 7.3), it makes sense to consider them separately, i.e.,

$$
\begin{aligned}
ELATE^-_s &= \max(AT^*_s - ETA_s, 0) \\
ELATE^+_s &= \max(ETA_s - AT^*_s, 0)\,.
\end{aligned}
$$

A last variable, the allocated time for travel, ALT_s, is defined as the time that a traveller spends in travel plus the time that he/she has to wait before the objective arrival time in the case of an early arrival, i.e.,

$$ALT_s = ETT_s + ELATE^-_s\,.$$

Note that when the trip is expected to arrive late, $ETA_s > AT^*_s$ then $ALT_s = ETT_s$. The variables defined above are represented in Figure 7.1 for a better understanding.

In the experiments, the route and departure time choices were observed in 5 OD pairs that differ in the magnitude of the ETT. This implies that the variables defined above need to be normalised before estimating a discrete choice model, otherwise the estimates would be biased towards the OD pairs with higher magnitude of ETT. In the previous chapters, the normalisation was done by dividing the informed travel time by the informed travel time in the fastest route. In the same manner, the normalisation for the ETT here is done considering the minimum ETT for all the alternatives in a choice problem, i.e,

$$NETT_s = \frac{ETT_s}{\min_s(ETT_s)}\,.$$

The $ELATE^-$ and $ELATE^+$ are normalised so that they represent the percentage of the trip

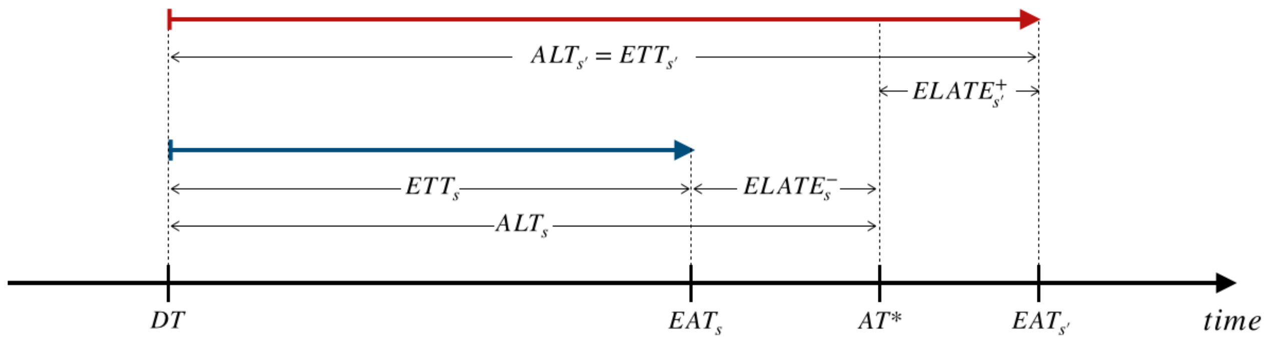

Figure 7.1: Representation of the variables ETT, $ELATE^-$ and $ELATE^+$ for two trips, s and s', with expected arrival time before and after the objective arrival time, respectively.

that is expected to arrive early or late, i.e,

$$NELATE^-_s = \frac{ELATE^-_s}{ALT_s}$$
$$NELATE^+_s = \frac{ELATE^+_s}{ALT_s} .$$

Note that the normalisation of ETT_s is done relative to the choice problem, which allows it to be comparable between choice problems and between OD pairs. The normalisation of $ELATE^-$ and $ELATE^+$ is done relative to the alternative s.

7.2.3 Joint route and departure time model

The joint route and departure choice model considered here is a MXL model with lagged errors. To include the lagged errors, it is convenient to represent the MXL model as an error components (EC) model. The alternatives are given by the pair (j,t), where j represents the route, and t the departure time interval. That is, (j,t) represents an alternative in the discrete choice model. The utility that individual i obtains from an alternative can be written as

$$U_{i,(j,t)} = \mathbf{x_{i,(j,t)}}^T \boldsymbol{\beta}_\mathbf{i} + \nu_{i,(j,t)} + \sum_{l=1}^{h} \theta_l \nu_{i,(j,t-l)} + \varepsilon_{i,(j,t)} \tag{7.1}$$

where $\mathbf{x_{i,(j,t)}}$ is the vector of explanatory variables, $\boldsymbol{\beta}_\mathbf{i}$ are the random coefficients indexed by individual i and $\varepsilon_{i,(j,t)}$ is the unobserved part of the utility. $\nu_{i,(j,t)}$ are the lagged errors and they are independent and identically distributed $\mathcal{N}(0,1)$; these errors are also independent of $\varepsilon_{i,(j,t)}$. The parameters θ_l account for the amount of correlation between neighbouring time intervals and they need to be estimated. To derive the covariance of the model in expression (7.1), let

$\eta_{i,(j,t)} = \nu_{i,(j,t)} + \sum_{l=1}^{h} \theta_l \nu_{i,(j,t-l)} + \varepsilon_{i,(j,t)}$. Then,

$$Cov(U_{i,(j,t)}, U_{i,(k,t')}) = \mathbf{x_{i,(j,t)}}^T \Sigma_\beta \mathbf{x_{i,(k,t')}} + Cov(\eta_{i,(j,t)}, \eta_{i,(k,t')}) \,, \tag{7.2}$$

where Σ_β is the covariance matrix of the random coefficients $\boldsymbol{\beta}_\mathbf{i}$. The second term in the right hand side of equation (7.2) is given by

$$Cov(\eta_{i,(j,t)}, \eta_{i,(k,t')}) = \begin{cases} 1 + \sum_{l=1}^{h} \theta_l^2 + \sigma_\varepsilon^2 & \text{if } k = j \,,\ t' = t \\ \theta_1 + \sum_{l=2}^{h} \theta_l \theta_{l-1} & \text{if } k = j \,,\ t' = t-1 \text{ or } t' = t+1 \\ \theta_2 + \sum_{l=3}^{h} \theta_l \theta_{l-2} & \text{if } k = j \,,\ t' = t-2 \text{ or } t' = t+2 \\ \vdots & \\ \theta_{h-1} + \theta_h \theta_1 & \text{if } k = j \,,\ t' = t-(h+1) \text{ or } t' = t+(h+1) \\ \vdots & \\ \theta_h & \text{if } k = j \,,\ t' = t-h \text{ or } t' = t+h \\ 0 & \text{otherwise.} \end{cases}$$

Note that the alternatives in the model in expression (7.1) have two sources of correlation: the first given by the covariance of the random coefficients $\boldsymbol{\beta}_\mathbf{i}$, i.e., Σ_β and the second given by the lagged errors ν. Thus, the correlation explained by the term $\mathbf{x_{i,(j,t)}}^T \Sigma_\beta \mathbf{x_{i,(k,t')}}$ in equation (7.2) could be confounded with the correlation induced by $Cov(\eta_{i,(j,t)}, \eta_{i,(k,t')})$. Therefore, to observe the effect of the lagged errors, in this chapter the coefficients $\boldsymbol{\beta}_\mathbf{i}$ are considered to be equal across all participants, i.e., $\boldsymbol{\beta}_\mathbf{i} = \boldsymbol{\beta}$. With this assumption,

$$Cov(U_{i,(j,t)}, U_{i,(k,t')}) = Cov(\eta_{i,(j,t)}, \eta_{i,(k,t')}) \,. \tag{7.3}$$

7.3 Results

7.3.1 Exploratory analysis

The route choice distribution is shown in Fig. 7.2, where it can be seen a strong relationship between the preferred route and the proportion of times that the routes were informed to be the fastest. Considering the 5 OD pairs, the percentage of choices that were for the fastest route is 57%, for the second fastest 24% and 19% for slow route. These percentages are similar to those found in Chapter 5 (60%, 24% and 16%). This means that participants are, in general, minimising their travel time when selecting a route, but that there is still a large number of choices that are *suboptimal* in the sense that they are not minimising travel time. Suboptimal behaviour is specially noted in OD pairs CC5 and CC8, where the percentage of times the fastest route was chosen is small. As argued in Chapter 5, this behaviour is likely due to boundedly rational behaviour, and it depends on both the relative differences in travel time.

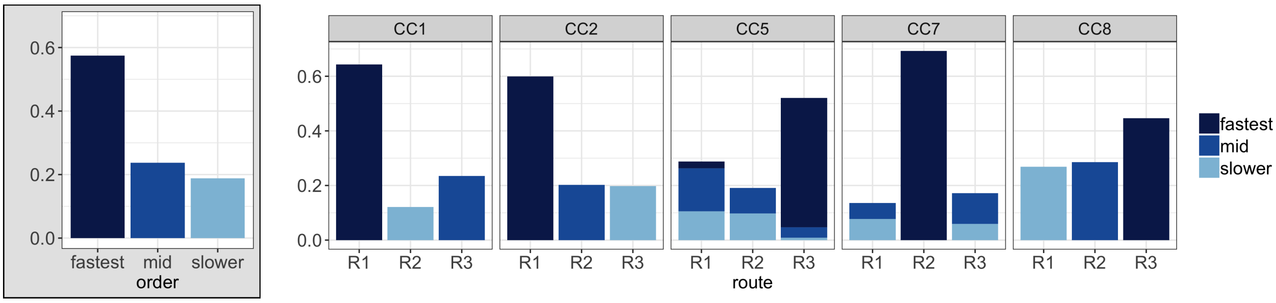

Figure 7.2: Route choice distribution. The percentage of choices for each of the three routes in the five OD pairs. The left plot represents the total percentage of choices for the fastest, second fastest and slowest routes. The colour indicates the percentage of the times that the route was the fastest informed route.

The choices of the participants are based mainly on the travel time estimates given for the different routes and departure times. A natural question that arises here is if participants minimise the difference between the estimated arrival time and the objective arrival time. To see this, the distribution of *ELATE* is shown in Fig. 7.3, where it can be seen that the mode of the distribution is placed to the left side of the origin, around -2.5 minutes. This means that the most frequent departure time choices are such that participants intend to arrive to their destination just before the objective arrival time. Moreover, for almost half of the choices *ETA* is in the interval $[-5, 5]$. As expected, the $ELATE$ distribution is asymmetrical, accumulating more observations before the objective arrival time: 80% of the trips were planned to arrive before the objective arrival time. This is explained because participants regard late arrivals to have more negative consequences than early arrivals. These results reveal that, globally, participants are optimising the ETA, in the sense that they minimise the expected travel time, ETT, at the same time that they try to minimise the risk of late arrivals $ELATE^{+}$. These patterns are observed on the five OD pairs.

The variability in the *ELATE* distribution may be explained by the heterogeneity of the risk profiles of participants. To obtain the risk profile of the participants, the distribution of *ELATE* is obtained at participant level. In order to approximate the distributions only participants with more than 4 recorded choices are taken into account in this analysis. These distributions are shown in Fig. 7.4, where it can be seen that the participants are heterogeneous in their *ETA* choice. Participants are categorised according to their risk profile depending on the 75th percentile of their *ELATE* distribution. Risk-averse participants are those whose 75th percentile of *ELATE* is less than 5 minutes, that is, participants who planned at least 75% of their trips to arrive more than 5 minutes before the objective arrival time. The risk-neutral are

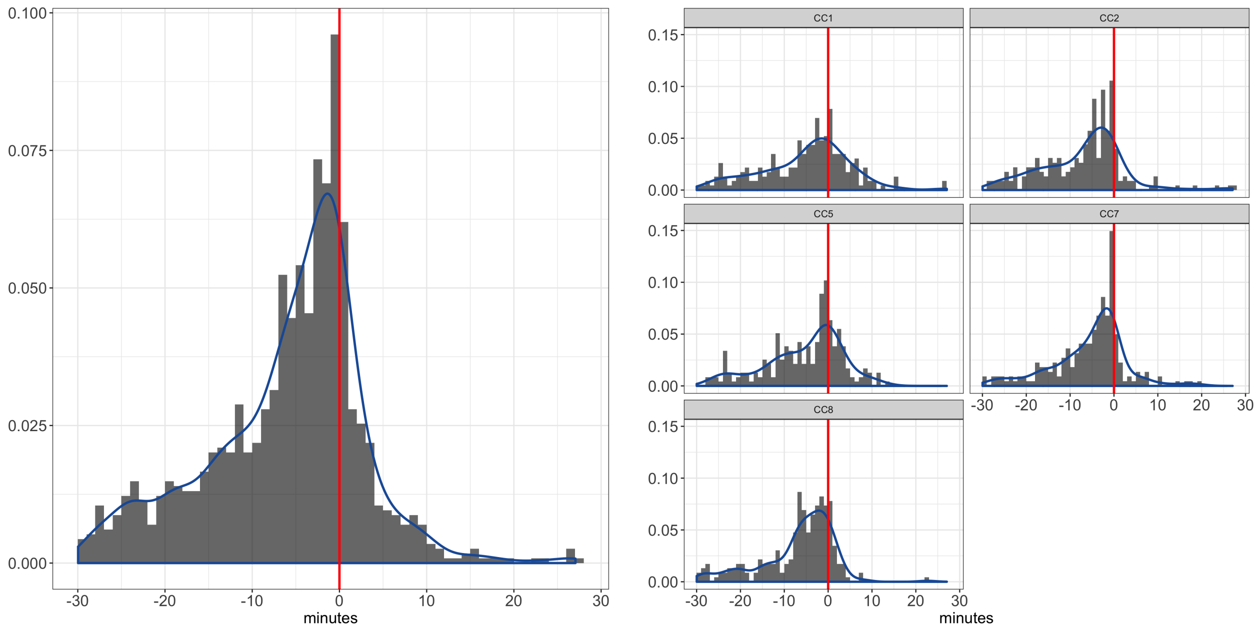

Figure 7.3: Expected lateness distribution. The mode of the distribution is placed just before $ELATE = 0$ and has a long negative tail. This means, on the one hand, that participants consider that late arrivals have worst consequences than early arrivals. On the other hand, they try to minimise the early arrivals.

defined as those participants with 75% of their trips arriving between 0 and 5 minutes before the objective arrival time, and the risk-prone are those who arrive late in at least 25% of the trips. Note that this last group has mean expected arrival times close to the objective arrival time. With this definition, the percentage of risk-averse, risk-neutral and risk-prone participants is 24%, 42% and 34%, respectively.

7.3.2 Route and departure time model estimates

Two variants of the model in expression (7.1) are estimated. In both models, the systematic part of the utility is given by

$$V_{i,(j,t)} = \beta_1 \cdot NETT_{i,(j,t)} + \beta_2 \cdot NELATE^{-}_{i,(j,t)} + \beta_3 \cdot NELATE^{+}_{i,(j,t)} \,,$$

and the correlation structures by

$$\eta^1_{i,(j,t)} = \theta_1 \nu_{i,(j,t-1)} + \nu_{i,(j,t)} + \varepsilon_{i,(j,t)} \tag{RDT1}$$

$$\eta^2_{i,(j,t)} = \theta_2 \nu_{i,(j,t-2)} + \theta_1 \nu_{i,(j,t-1)} + \nu_{i,(j,t)} + \varepsilon_{i,(j,t)} \,. \tag{RDT2}$$

Model RDT1 considers a lag of $h = 1$, whereas model RDT2 a lag $h = 2$. The estimated models are compared to that of the MNL model in order to determine if including lagged errors improves

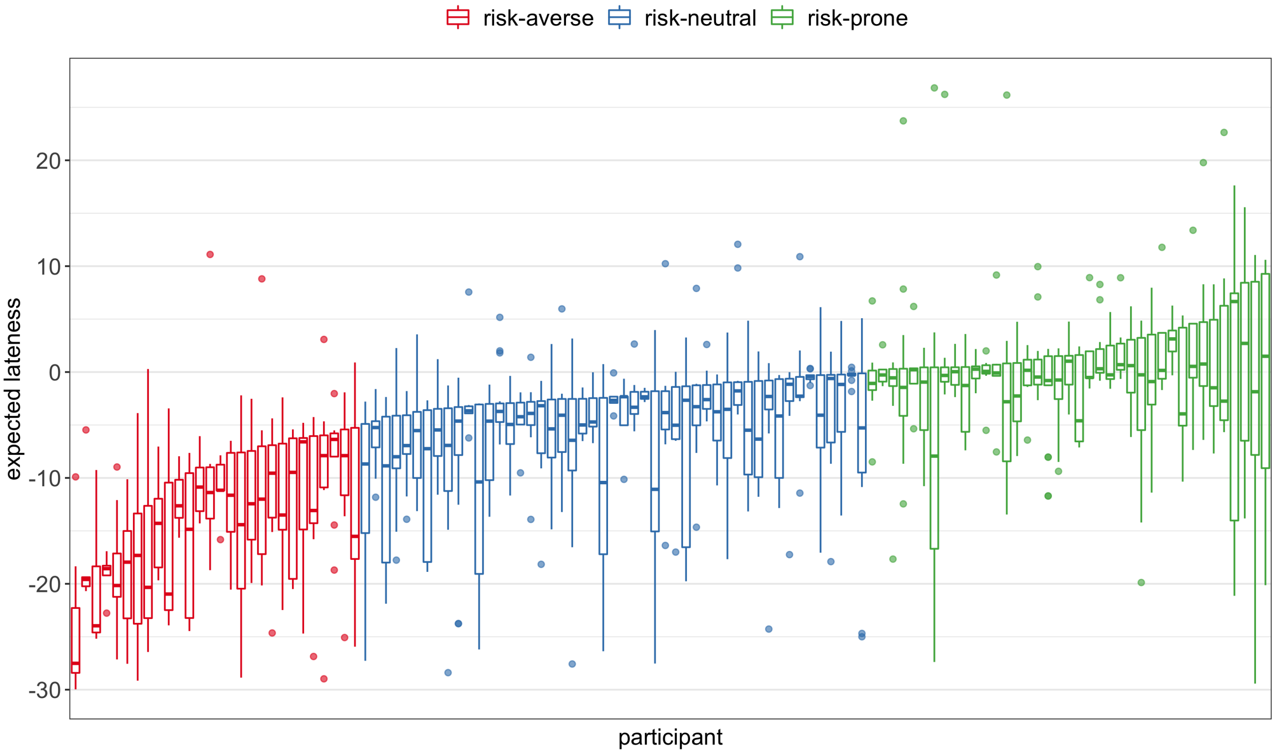

Figure 7.4: Distribution of *ELATE* per participant. The distributions are ordered by the 75th percentile of the distribution.

the quality of the joint route and departure time models. The estimates of the three models are included in Table 7.1, where it can be seen that the estimated parameter $\hat{\beta}_1 < 0$, meaning that participants prefer shorter routes in time. The variables $NELATE^{-}$ and $NELATE^{+}$ are the percentage of the allocated time for travel that early or late arrivals represent. Since these variables are expressed in the same units, then the magnitude of their estimated coefficients can be compared directly. Both parameters have negative sign, meaning that there is a disutility associated with early and late arrivals. However, $\hat{\beta}_3/\hat{\beta}_2 > 2$, which implies that for the average participants late arrivals are twice less desirable than early arrivals. In the case of the estimated travel time. These results considered altogether suggest that participants make choices in order to minimise their travel time, while avoiding late arrivals. The same conclusion was found in the interpretation of the results in the exploratory analysis.

The goodness of fit of the three models is assessed with the training error. This is, the prediction error for the same data that was used to fit the models. The training error is compared for the marginal probabilities for the route and departure time choices, i.e, $Pr(R)$ and $Pr(DT)$. To obtain the training error for the route choices, the methodology of the previous chapters, where the observed and predicted route choice distributions were compared, is replicated (see Section 6.3.2 for more details). Note that this methodology requires to compute the observed

Table 7.1: Estimates of the mean parameters of the three joint route and departure time models.

	MNL		RDT1		RDT2	
Parameter	mean	s.e.	mean	s.e.	mean	s.e.
$\hat{\beta}_1$	-1.700	0.105	-1.839	0.119	-1.767	0.123
$\hat{\beta}_2$	-3.768	0.184	-4.079	0.218	-3.823	0.258
$\hat{\beta}_3$	-7.751	0.372	-8.231	0.409	-8.575	0.444
$\hat{\theta}_1$	-	-	-0.311	0.454	0.047	0.307
$\hat{\theta}_2$	-	-	-	-	-1.068	0.169
Deviance	6295.502	2.355	5183.121	180.100	4665.145	174.639
DIC	6298.300	-	21113.100	-	19855.800	-

choice distribution, which is possible when the alternatives are the same for a sufficient number of choice problems. The case of the departure time is different. The reason is that in the MDG experiment the choice problems happen at any time between 6:45 and 9:00 in the morning. Thus, the departure time periods vary for different choice problems. For example, one choice problem may have 10 departure time intervals of length five minutes from 7:00 to 7:50, while a second choice problem from 7:30 to 8:20. In a strict sense, the alternatives of these two choice problems are not the same, even for the same OD pair. Therefore, it is not possible to obtain the observed distribution for the departure time choices (nor for the joint route and departure time choice). In order to measure the error of the models in time, the absolute difference between the departure time of the different alternatives and the actual observed departure time is used as an error measure. This error is then weighted by the choice probability predicted by the model. Formally, for a given choice problem let $DT_{(j,t)}$ be the departure time of the alternative route j in the time interval t, DT^* the actual departure time choice, and $Pr(DT = t)$ the marginal departure time probability given by the fitted model. Then, the prediction error for a choice problem is given by

$$err_{DT} = (1/3) \sum_{j=1}^{3} \sum_{t} |DT_{(j,t)} - DT^*| \cdot Pr(DT = t). \tag{7.4}$$

The errors err_{DT}, obtained for each choice problem, are then averaged to obtain the average error of the model $\bar{err}_{DT}$. The errors for both the route and departure time marginals are shown in Table 7.2. Compared to the MNL model, there is an improvement in the fit of both the route and departure time marginals. Model RDT1 has the smallest error for the departure time choice, but model RDT2 has the smallest error for the route choice. Even though the improvements are small, this suggests that considering the correlations between consecutive intervals could increase the accuracy of the models. The error incurred by a model can be also assessed by comparing the difference between the observed and predicted *ELATE*. This is done by replacing DT by $ELATE$ in equation (7.4); denote this error by err_{ELATE} and the average error for all choice situations as $\bar{err}_{ELATE}$. The difference between err_{DT} and err_{ELATE} is that the first measures the error in departure time, while the second is a measure of the error at arrival time. The error $\bar{err}_{ELATE}$ is also included in Table 7.2. The results show the same patterns as in $\bar{err}_{DT}$, with models RDT1 and RDT2 having smaller errors than the MNL model. An interesting result is that $\bar{err}_{ELATE} < \bar{err}_{DT}$, implying that the three models are more accurate for this variable than for the departure time. Therefore, suggesting that they are minimising the error with respect to the objective arrival time, and not with respect to the departure time.

Table 7.2: Training error of the estimated joint route and departure time models.

	Route choice (%)			$\bar{err}_{DT}$ (mins.)			$\bar{err}_{ELATE}$ (mins.)		
OD pair	MNL	RDT1	RDT2	MNL	RDT1	RDT2	MNL	RDT1	RDT2
All	12.0	11.9	11.8	9.92	9.79	9.83	9.59	9.43	9.48
CC1	26.6	26.3	26.3	10.70	10.55	10.61	11.00	10.80	10.80
CC2	9.7	10.7	10.3	9.49	9.42	9.44	8.95	8.81	8.89
CC5	16.3	16.1	16.0	9.87	9.70	9.73	10.30	10.10	10.10
CC7	3.2	2.9	2.6	10.58	10.48	10.51	8.87	8.70	8.78
CC8	4.1	3.5	3.7	8.95	8.81	8.88	8.78	8.62	8.70

7.4 Conclusions

This chapter presented the results of a first effort to model the joint route and departure time choices of travellers. The data comes from an experiment, conducted with the MDG platform, in which participants were faced with simultaneous route and departure time choice problems. The exploratory analysis shows that participants chose routes in order to minimise their expected travel time, at the same time that they avoid early and late arrivals. The preferences for early and late arrivals are, as expected, asymmetrical: late arrivals are considered to have worse consequences than early arrivals. This result is confirmed by the estimated joint choice models. Participants exhibit heterogeneous risk profiles. On the one hand, risk-averse participants (24%) planned their trips to arrive at least 5 minutes before the objective arrival, and on the other hand, risk-prone participants (34%) planned their trips to arrive close to the objective arrival time and often late. The fit of the two models that account for correlation between time periods was compared against the fit of the MNL model. The specifications of the models studied here are simple: only three explanatory variables enter the model, and the coefficients are considered fixed in the population. The results show a slightly improvement of the models that consider correlated departure time intervals. The improvement is observed for both the departure time and route choice marginals. Therefore, suggesting that considering correlated time intervals improves the fit of the models, and thus their predictive accuracy. Nevertheless, the significance of the results in this chapter need to be tested. Even though the results presented here are partial, they are promising and should be further investigated. As future work, more complex specifications could be tested, for example, by including random coefficients that account for the repeated choices of participants and taste variation. Also, the sensibility of the models to the size of the departure time intervals should be assessed, to determine the impact of the subjective discretisation of the departure time intervals in the quality of the models. Another line of research could be adapting the BRCS model, introduced in Chapter 6, to include lagged errors.

General conclusions and research perspectives

In this thesis, the choice behaviour of travellers was empirically studied through in-laboratory computer experiments. The objective was to propose and estimate choice models that predict the choices of travellers at large-scale urban network level, for subsequent implementation in traffic simulators. To attain this objective, the models need to generalise the choices of travellers to the large variety of situations that are found in an urban network, maintaining, at the same time, their consistency with the actual behaviour of travellers. The models adopted to predict the choices of travellers come from the the mixed logit model family. The data to estimate the models comes from several in-laboratory computer experiments that comprise the choices of unimodal car trips. The experiments were carried out using the Mobility Decision Game (MDG), an internally developed platform that allows to observe the choices of participants in a variety of hypothetical scenarios in an urban transportation network. The majority of this thesis focused on travellers' route choices, Part I is dedicated to this subject. Part II contains only one chapter, devoted to the simultaneous route and departure time choices. The main results obtained in this thesis are summarised below, and some research perspectives are given at the end of this chapter.

In a city-scale network, trips are made in thousands of OD pairs connected by a large number of diverse routes. In the case concerning this thesis, the city of Lyon in France, the network has more than 96,096 OD pairs and more than 559,423 routes. From the point of view of the design of experiments, this implies that the number of scenarios must be reduced to a small set in which the choices of participants can be observed through experimentation. Furthermore, this small set must be representative of the scenarios found in the whole network, in the sense that the choices of travellers in any scenario can be predicted by a choice model estimated with the representative set. In Chapter 4, a methodology based on cluster analysis was proposed in order to find the representative set of OD configurations. In cluster analysis, elements are assigned to groups whose elements are similar between themselves, but dissimilar to the elements in the other groups. The mean element of each cluster (the centroid) can be selected to represent its group, and the set of all the cluster centroids to represent the whole network. This implies that a choice model estimated with data obtained for the representative

OD pairs will generalise to the rest of the OD pairs in the network. Moreover, the centroids are dissimilar from each other, allowing to identify the influence of each attribute in the choices of the participants. For the city of Lyon in France, 9 OD pairs and their three connecting routes were used as representative of 624,490 OD configurations. The routes of the nine representative OD configurations cover around 83% of the values of the attributes of the routes in the network. It was found that the models estimated with the representative set of OD and routes have, in general, higher predictive accuracy. On average, the accuracy is increased in around 2.5 percentage points. Moreover, the models estimated with the representative set showed no extreme errors for individual OD pairs. This implies that the models estimated with the representative set will show a relative better global prediction accuracy without incurring in large errors on individual OD-routes. These results demonstrate that the accuracy of a choice model for the whole network can be improved by estimating the choice models on representative OD pairs, and that the representative OD pairs and routes can be obtained with cluster analysis.

Chapter 5 and Chapter 6 are concerned with finding the route choice model that best predicts the choices of travellers. To this purpose, the influence that the travel time information and the route attributes have on the choices of travellers is first studied. The first main finding is that travellers evaluate relative rather than absolute differences in travel time. This means that a 5-minute difference in travel time weights differently for trips of 10 and 30 minutes. In the first case, the difference represents an increment of travel time of 50% with respect to the alternative, whilst in the second case the difference is of 15%. In traffic assignment, travellers are often treated as perfect rational with respect to travel time, assuming that their choices are such that they minimise their travel time. However, recent studies have shown that this is not the case. The results in Chapter 5 point in this direction, finding that participants chose the fastest route in 60% of the cases, and that only 10% of them chose always the fastest route. Thus, suggesting that the choices of travellers with respect to this variable are best explained by boundedly rational behaviour. In this regard, it was found that the participants have heterogeneous indifference bands, and that at least 70% of them would not consider routes 1.5 times slower than the fastest alternative. An estimate for the mean indifference band in the population is 31.3%, meaning that the average participant did not consider routes with travel time differences 1.3 times slower than the fastest alternative. These findings are the motivation of the bounded rational choice set generation mixed logit model (BRCS), developed in Chapter 6. This model contemplates (i) a boundedly rational behaviour to generate the choice set, discarding the alternatives with travel times above a threshold, and then (ii) a rational behaviour to choose one alternative from the choice set. The BRCS model jointly estimates the choice set and the route choice, allowing for the BRCS to implicitly infer the latent population's indifference band distribution. In terms of predictive accuracy, the BRCS model was superior to all the MXL model specifications tested in this thesis, as it is shown in the conclusions of Part I.

Finally, Chapter 7 presents partial results of a first effort to simultaneously model the route and departure time choices of participants in a MDG experiment. The approach was to account for the stimuli of the participants' choices in a single representation of the utility. The response departure time was discretised in intervals of five minutes and correlation between consecutive intervals considered. The results are promising. Compared to a MNL model without time-correlated alternatives (reference model), the proposed model shows slightly better fit for both the route and departure time. The models that were tested are simple and further research must be done to improve the quality of the models. Possible lines of investigation are mentioned below.

Research perspectives

The work in this thesis can be further extended in four directions: experimentation, models, validation, and implementation.

Experimentation. The empirical data collected in this thesis encompasses the choices of the participants who are mainly staff and students from the university. Hence, it is not representative of the whole population. The data could be complemented by performing new experiments for other segments of the population. The new experiments should be carried out using the nine representative OD pairs in order to guarantee the generalisation of the choice models to the whole network. Choice models estimated with the new data could be modified to treat differently the segments of the population. This could improve the choice models in terms of representativeness not only of the network, but the population. In its current development state, the MDG does not allow to study certain aspects of travellers' behaviour, particularly, learning from experience and the reaction to travel time information uncertainty. The learning behaviour cannot be observed due to the long duration between the beginning and the end of a trip (more than 10 minutes), which prevents to present choice problems sequentially to the participants and thus to observe how the consequences of the previous trip influence the current choice. Learning could be studied with the MDG by reducing the duration of the trips and observing the repeated choices over a same OD pair with unknown and variable traffic conditions. Thus, allowing to estimate the learning rate as a function of travel time variability and the influence of travel time uncertainty in their choices. The reaction of travellers to travel time information uncertainty could be studied by providing to the participants the distribution of the travel times over the alternative routes, instead of a point estimate.

Models. The route choice models in this thesis could be improved by defining more complex specifications of the systematic part of the utility. In particular, the overlapping of the alternatives. Even though the mixed logit models account for the correlations between alternatives, making this correlation explicit for overlapping routes could improve the predictions and the interpretation of the models. Regarding the BRCS model proposed in Chapter 6, different forms of the distribution of the indifference bands could be tested. Specially, a mixture of distributions (zero-inflated model) could help the model to better detect travellers that behave rationally. The BRCS model in this thesis was estimated considering only three alternative routes in the choice set. The questions on how the model can be estimated considering a large number of alternatives and if the model is still capable of finding the underlying distribution of the indifference bands can be addressed using synthetic data. In the last research chapter, the first results of an investigation of the simultaneous route and departure time choices were presented. A future step in this direction could be to consider separate utility functions for the route and departure time choices. This would require to enter different variables in each of the utilities, and to estimate the probabilities jointly. The BRCS model could be used for the probability of route choice, while a time-correlated MXL model for the departure time.

Validation. The MDG experiments belong to the SP methods. Thus, as discussed in the introduction to this thesis, it suffers from external validity issues that are not present in the RP methods. Alternative RP data sources, such as GPS traces, could be used as ground truth to compare the predictions of the estimated choice models. Therefore, obtaining a measure of the discrepancy between the two data sources. This discrepancy is an estimator of the *real* predictive error. Moreover, the data sources from the SP and RP methods could be merged to estimate choice models. An interesting question that could be addressed is if the choice models estimated here adapt well to other urban networks.

Implementation. The purpose of the choice models obtained in this thesis is their future implementation in traffic assignment algorithms at large-scale, specially in dynamic traffic assignment. The simulations based on these algorithms can be used to test different traffic control strategies and transport network planning. The implementation can be done considering the mean behaviour found for the participants (the mean of the random coefficients estimated in this work), or by simulating the individuals' coefficients based on their distribution. The choice models can be implemented in two ways: (i) a unique model for all the network, and (ii) one model per group of OD pairs. The models estimated in this thesis are of the first kind. However, the choice models could be estimated for each cluster of OD pairs, and implemented in the traffic assignment algorithms per cluster. The implementation should take into account the heterogeneity and the correlation of travellers' preferences, captured in the estimated models by the joint distribution of the coefficients. The findings in Chapter 5, related to size and heterogeneity of the indifference bands could be used to reduce the number of alternatives considered by the travellers. Sensibility analysis of this parameter could help determine the trade-off between the efficiency of the assignment algorithm and the incurred error. The traffic assignment could also be validated against observed traffic states coming, for example, from GPS traces.

References

Abdel-Aty M.A., Kitamura R., and Jovanis P.P., 1997. Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. *Transportation Research Part C: Emerging Technologies*, 5(1):39–50. ISSN 0968090X. doi: 10.1016/S0968-090X(96)00023-X.

Adamowicz W., Boxall P., Williams M., and Louviere J., 1998. Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics*, 80(1):64–75. doi: 10.2307/3180269.

Adamowicz W., Louviere J., and Williams M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26(3):271 – 292. ISSN 0095-0696. doi: https://doi.org/10.1006/jeem.1994.1017.

Adler J.L. and McNally M.G., 1994. In-laboratory experiments to investigate driver behavior under advanced traveler information systems. *Transportation Research Part C*, 2(3):149–164. ISSN 0968090X. doi: 10.1016/0968-090X(94)90006-X.

Ashford J.R. and Sowden R.R., 1970. Multi-variate probit analysis. *Biometrics*, 26(3):535–546. ISSN 0006341X, 15410420.

Avineri E. and Prashker J.N., 2005. Sensitivity to travel time variability: Travelers learning perspective. *Transportation Research Part C: Emerging Technologies*, 13(2):157–183. ISSN 0968090X. doi: 10.1016/j.trc.2005.04.006.

Baillon J.B. and Cominetti R., 2008. Markovian traffic equilibrium. *Mathematical Programming*, 111(1-2):33–56. ISSN 00255610. doi: 10.1007/s10107-006-0076-2.

Bajwa S., Bekhor S., Kuwahara M., and Chung E., 2009. DISCRETE CHOICE MODELING OF COMBINED MODE AND DEPARTURE TIME. 8602. doi: 10.1080/18128600808685681.

Balcombe K., Chalak A., and Fraser I., 2009. Model selection for the mixed logit with bayesian estimation. *Journal of Environmental Economics and Management*, 57(2):226 – 237. ISSN 0095-0696. doi: https://doi.org/10.1016/j.jeem.2008.06.001.

Barber D., 2011. *Bayesian Reasoning and Machine Learning*. October. Cambridge University Press, Cambridge. ISBN 9780511804779. doi: 10.1017/CBO9780511804779.

Bekhor S., 2016. Link-Nested Logit Model of Route Choice. (January 1998). doi: 10.3141/1645-17.

Bekhor S., Ben-Akiva M.E., and Ramming M.S., 2006. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1):235–247. ISSN 1572-9338. doi: 10.1007/s10479-006-0009-8.

Ben-akiva M., 1987. Incorporating Random Constraints in Discrete. (2):91–102.

Ben-Akiva M. and Bierlaire M., 1999. *Discrete Choice Methods and their Applications to Short Term Travel Decisions*, pages 5–33. Springer US, Boston, MA. ISBN 978-1-4615-5203-1. doi: 10.1007/978-1-4615-5203-1_2.

Ben-Akiva M., Bolduc D., and Walker J., 2001. Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. *Department of Civil ...*.

Ben-Akiva M. and De Palma A., 1991. Dynamic network models and driver information systems. *Transportation Research Part A: General*, 25(5):251–266. ISSN 01912607. doi: 10.1016/0191-2607(91)90142-D.

Ben-Akiva M., McFadden D., Gärling T., Gopinath D., Walker J., Bolduc D., Börsch-Supan A., Delquié P., Larichev O., Morikawa T., Polydoropoulou A., and Rao V., 1999. Extended Framework for Modeling Choice Behavior. *Marketing Letters*, 10(3):187–203. ISSN 09230645. doi: 10.1023/A:1008046730291.

Ben-Akiva M. and Morikawa T., 1990. Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6):485 – 495. ISSN 0191-2607. doi: https://doi.org/10.1016/0191-2607(90)90037-7.

Ben-Akiva M., Morikawa T., and Shiroishi F., 1992. Analysis of the reliability of preference ranking data. *Journal of Business Research*, 24(2):149 – 164. ISSN 0148-2963. doi: https://doi.org/10.1016/0148-2963(92)90058-J.

Ben-Akiva M.E., 1973. *Structure of passenger travel demand models*. Ph.D. thesis, Massachusetts Institute of Technology.

Ben-Elia E. and Avineri E., 2015. Response to Travel Information: A Behavioural Review. *Transport Reviews*, 35(3):352–377. ISSN 14645327. doi: 10.1080/01441647.2015.1015471.

Ben-Elia E. and Shiftan Y., 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice*, 44(4):249–264. ISSN 09658564. doi: 10.1016/j.tra.2010.01.007.

Bernoulli D., 1954. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36. ISSN 00129682, 14680262.

Bhat C.R., 1998. Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological*, 32(6):361–371. ISSN 01912615. doi: 10.1016/S0191-2615(98)00004-6.

Bhat C.R. and Castelar S., 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the san francisco bay area. *Transportation Research Part B: Methodological*, 36(7):593 – 616. ISSN 0191-2615. doi: https://doi.org/10.1016/S0191-2615(01)00020-0.

Bhat C.R. and Steed J.L., 2002. A continuous-time model of departure time choice for urban shopping trips.

Bierlaire M. and Frejinger E., 2008. Route choice modeling with network-free data. *Transportation Research Part C: Emerging Technologies*, 16(2):187–198. ISSN 0968090X. doi: 10.1016/j.trc.2007.07.007.

Bifulco G.N., Di Pace R., and Viti F., 2014. Evaluating the effects of information reliability on travellers' route choice. *European Transport Research Review*, 6(1):61–70. doi: 10.1007/s12544-013-0110-4.

Bogers E.a.I., 2005. *Joint modeling of ATIS, habit and learning impacts on route choice by laboratory simulator experiments*. Ph.D. thesis, Delft University of Technology.

Bogers E.A.I., Viti F., and Hoogendoorn S.P., 2005. Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments. *Transportation Research Record*, 1926(1):189–197. doi: 10.1177/0361198105192600122.

Bogers E.A.I., Viti F., Hoogendoorn S.P., and Zuylen H.J.V., 2006. Valuation of different types of travel time reliability in route choice: Large-scale laboratory experiment. *Transportation Research Record*, 1985(1):162–170. doi: 10.1177/0361198106198500118.

Bonsall P., 1992. The influence of route guidance advice on route choice in urban networks. *Transportation*, 19(1):1–23. ISSN 00494488. doi: 10.1007/BF01130771.

Bovy P.H. and Stern E., 1990. *Route Choice: Wayfinding in Transport Networks*. Kluwer Academic Publishers. ISBN 9789400906334.

Bradley M. and Daly A., 1991. Estimation of logit choice models using mixed stated preference and revealed preference information. *Paper presented to the 6th International Conference on Travel Behavior, Quebec.*

Brockmann D., Hufnagel L., and Geisel T., 2006. The scaling laws of human travel. *Nature*, 439(7075):462–465. ISSN 14764687. doi: 10.1038/nature04292.

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):315 – 338. ISSN 0191-2615. doi: https://doi.org/10.1016/S0191-2615(99)00031-4.

Cascetta E., Nuzzolo A., Russo F., and Vitetta A., 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In J.B. Lesort, editor, *Transportation and traffic theory : proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 697–711. Pergamon, 1996., Lyon, France.

Chen P.S.t. and Mahmassani H.S., 1993. Dynamic interactive simulator for studying commuter behavior under real-time traffic information supply strategies. *Transportation Research Record*, 1413(November 1993):12–21.

Daganzo C., 1979. *Multinomial probit : the theory and its application to demand forecasting.* Academic Press, New York, NY, USA.

Daganzo C.F. and Sheffi Y., 1977. On stochastic models of traffic assignment. *Transportation Science*, 11(3):253–274. doi: 10.1287/trsc.11.3.253.

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):245 – 268. ISSN 1366-5545. doi: https://doi.org/10.1016/S1366-5545(02)00037-6.

De Moraes Ramos G., Daamen W., and Hoogendoorn S., 2013. Modelling travellers' heterogeneous route choice behaviour as prospect maximizers. *Journal of Choice Modelling*, 6:17–33. ISSN 17555345. doi: 10.1016/j.jocm.2013.04.002.

de Moraes Ramos G., Mai T., Daamen W., Frejinger E., and Hoogendoorn S.P., 2020. Route choice behaviour and travel information in a congested network: Static and dynamic recursive models. *Transportation Research Part C: Emerging Technologies*, 114(February):681–693. ISSN 0968090X. doi: 10.1016/j.trc.2020.02.014.

de Palma A. and Picard N., 2005. Route choice decision under travel time uncertainty. *Transportation Research Part A: Policy and Practice*, 39(4 SPEC. ISSS.):295–324. ISSN 09658564. doi: 10.1016/j.tra.2004.10.001.

Di X. and Liu H.X., 2016. Boundedly rational route choice behavior : A review of models and methodologies. *Transportation Research Part B*, 85:142–179. ISSN 0191-2615. doi:

10.1016/j.trb.2016.01.002.

Earnhart D., 2002. Combining revealed and stated data to examine housing decisions using discrete choice analysis. *Journal of Urban Economics*, 51(1):143 – 169. ISSN 0094-1190. doi: https://doi.org/10.1006/juec.2001.2241.

Fosgerau M., Frejinger E., and Karlstrom A., 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56. ISSN 01912615. doi: 10.1016/j.trb.2013.07.012.

Frejinger E., 2008. *Route choice analysis: data, models, algorithms and applications*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne. doi: 10.5075/epfl-thesis-4009.

Frejinger E. and Bierlaire M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, 41(3):363–378. ISSN 01912615. doi: 10.1016/j.trb.2006.06.003.

Frejinger E., Bierlaire M., and Ben-Akiva M., 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994. ISSN 01912615. doi: 10.1016/j.trb.2009.03.001.

Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., and Rubin D.B., 2014. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL, third edition.

Golob T.F. and Regan A.C., 2002. Trucking industry adoption of information technology: a multivariate discrete choice model. *Transportation Research Part C: Emerging Technologies*, 10(3):205 – 228. ISSN 0968-090X. doi: https://doi.org/10.1016/S0968-090X(02)00006-2.

González M.C., Hidalgo C.A., and Barabási A.L., 2008. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782. ISSN 14764687. doi: 10.1038/nature06958.

Gumbel E.J., 1954. *Statistical Theory of Extreme Values and Some Practical Applications: A Series of Lectures*, volume 33. US Government Printing Office, Washington.

Haab T.C. and Hicks R.L., 1997. Accounting for choice set endogeneity in random utility models of recreation demand. *Journal of Environmental Economics and Management*, 34(2):127–147. ISSN 00950696. doi: 10.1006/jeem.1997.1009.

Habib K.M.N., 2013. A joint discrete-continuous model considering budget constraint for the continuous part: application in joint mode and departure time choice modelling. *Transportmetrica A: Transport Science*, 9(2):149–177. doi: 10.1080/18128602.2010.551525.

Hadjidimitriou S.N., Dell'Amico M., Cantelmo G., and Viti F., 2015. Assessing the consistency between observed and modelled route choices through gps data. In *2015*

International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pages 216–222. ISSN null. doi: 10.1109/MTITS.2015.7223259.

Hastie T., Tibshirani R., and Friedman J., 2009. *The Elements of Statistical Learning*, volume 1 of *Springer Series in Statistics*. Springer New York, New York, NY, 2 edition. ISBN 978-0-387-84857-0. doi: 10.1007/b94608.

Hensher D.A. and Bradley M., 1993. Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters*, 4(2):139–151. ISSN 1573-059X. doi: 10.1007/BF00994072.

Horowitz J.L. and Louviere J.J., 1995. What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, 12(1):39–54. ISSN 01678116. doi: 10.1016/0167-8116(95)00004-L.

Hosmer D.W. and Lemesbow S., 1980. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, 9(10):1043–1069. doi: 10.1080/03610928008827941.

Hougaard P., 1999. Fundamentals of survival data. *Biometrics*, 55(1):13–22. doi: 10.1111/j.0006-341X.1999.00013.x.

Huber J. and Train K., 2001. On the similarity of classical and bayesian estimates of individual mean partworths. *Marketing Letters*, 12(3):259–269.

Iida Y., Akiyama T., and Uchida T., 1992. Experimental analysis of dynamic route choice behavior. *Transp. Res. B*, 26(1):17–32. ISSN 0191-2615. doi: 10.1016/0191-2615(92)90017-Q.

Institut national de la statistique et des études économiques, 2018. Découpage infracommunal: Table d'appartenance géographique des iris. [Online; accessed 13-May-2019].

Kahneman D. and Tversky A., 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica: Journal of the Econometric Society*, 47(3):263–291. ISSN 0012-9682. doi: 10.1111/j.1536-7150.2011.00774.x.

Kaplan S. and Prato C.G., 2012. Closing the gap between behavior and models in route choice: The role of spatiotemporal constraints and latent traits in choice set formation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(1):9–24. ISSN 13698478. doi: 10.1016/j.trf.2011.11.001.

Koutsopoulos H.N., Lotan T., and Yang Q., 1994. A driving simulator and its application for modeling route choice in the presence of information. *Transportation Research Part C*, 2(2):91–107. ISSN 0968090X. doi: 10.1016/0968-090X(94)90002-7.

Koutsopoulos H.N., Polydoropoulou A., and Ben-Akiva M., 1995. Travel simulators for data collection on driver behavior in the presence of information. *Transportation Research Part C*, 3(3):143–159. ISSN 0968090X. doi: 10.1016/0968-090X(95)00003-2.

Kroes E.P. and Sheldon R.J., 1988. Stated preference methods, an introduction.

Krug J., Burianne A., Bécarie C., and Leclercq L., 2019. Fine-positioning of trip starts and ends during travel-demand at the large urban-scale. *Journal of Transport Geography*, Manuscript submitted for publication.

Laval J.A. and Leclercq L., 2008. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B: Methodological*, 42(6):511–522. ISSN 01912615. doi: 10.1016/j.trb.2007.10.004.

Laval J.A. and Leclercq L., 2010. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4519–4541. ISSN 1364503X. doi: 10.1098/rsta.2010.0138.

Leclercq L., 2007. Hybrid approaches to the solutions of the "Lighthill-Whitham-Richards" model. *Transportation Research Part B: Methodological*, 41(7):701–709. ISSN 01912615. doi: 10.1016/j.trb.2006.11.004.

Levin D.A. and Peres Y., 2017. *Markov Chains and Mixing Times*. American Mathematical Society, 2 edition. ISBN 978-1-4704-4232-3.

Liu Y.H. and Mahmassani H.S., 1998. Dynamic aspects of commuter decisions under advanced traveler information systems: Modeling framework and experimental results. *Transportation Research Record*, 1645(1):111–119. doi: 10.3141/1645-14.

Lotan T., 1997. Effects of familiarity on route choice behavior in the presence of information. *Transportation Research Part C: Emerging Technologies*, 5(3-4):225–243. ISSN 0968090X. doi: 10.1016/S0968-090X(96)00028-9.

Luce R.D., 1959. *Individual choice behavior*. John Wiley.

Mahmassani H.S. and Chang G.L., 1987. On boundedly rational user equilibrium in transportation systems. *Transportation Science*, 21(2):89–99. doi: 10.1287/trsc.21.2.89.

Mahmassani H.S. and Jayakrishnan R., 1991. System performance and user response under real-time information in a congested traffic corridor. *Transportation Research Part A: General*, 25(5):293–307. ISSN 01912607. doi: 10.1016/0191-2607(91)90145-G.

Mahmassani H.S. and Liu Y.H., 1999. Dynamics of commuting decision behaviour under Advanced Traveller Information Systems. *Transportation Research Part C: Emerging Technologies*, 7(2-3):91–107. ISSN 0968090X. doi: 10.1016/S0968-090X(99)00014-5.

MANSKI C. AND MCFADDEN D., 1981. *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press. ISBN 9780262131599.

MANSKI C.F., 1977. The structure of random utility models. *Theory and Decision*, 8(3):229–254.

MARCHETTI C., 1994. Anthropological invariants in travel behavior. *Technological Forecasting and Social Change*, 47(1):75–88. ISSN 00401625. doi: 10.1016/0040-1625(94)90041-8.

MARSCHAK J., 1959. Binary choice constraints on random utility indicators. Cowles Foundation Discussion Papers 74, Cowles Foundation for Research in Economics, Yale University.

MCFADDEN D., 1972. Conditional logit analysis of qualitative choice behavior. *In: Zarembka, P. (Ed.), Frontiers of Econometrics*.

MCFADDEN D., 1978. Modeling the Choice of Residential Location. *Transportation Research Record*, (673):72–77.

MCFADDEN D. AND TRAIN K., 2000. Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470. ISSN 08837252, 10991255.

MCFADDEN D.L., 1984. Chapter 24 econometric analysis of qualitative response models. volume 2 of *Handbook of Econometrics*, pages 1395 – 1457. Elsevier. doi: https://doi.org/10.1016/S1573-4412(84)02016-X.

NEUMANN J.V. AND MORGENSTERN O., 1944. *Theory of Games and Economic Behavior*. Princeton University Press. ISBN 0691119937.

NURUL HABIB K.M., 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):639–653. ISSN 09658564. doi: 10.1016/j.tra.2009.05.001.

PAPINSKI D., SCOTT D.M., AND DOHERTY S.T., 2009. Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based gps. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(4):347 – 358. ISSN 1369-8478. doi: https://doi.org/10.1016/j.trf.2009.04.001.

PENG C., JIN X., WONG K.C., SHI M., AND LIÒ P., 2012. Collective human mobility pattern from taxi trips in urban area. *PLOS ONE*, 7(4):1–8. doi: 10.1371/journal.pone.0034487.

PLUMMER M., 2003. JAGS : A program for analysis of Bayesian graphical models using Gibbs sampling JAGS : Just Another Gibbs Sampler. *DSC 2003 Working Papers (Draft Versions)*, pages 1–8.

PRATO C.G., BEKHOR S., AND PRONELLO C., 2012. Latent variables and route choice behavior. *Transportation*, 39(2):299–319. ISSN 1572-9435. doi: 10.1007/s11116-011-9344-y.

R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramming M.S., 2002. *Network knowledge and route choice*. Ph.D. thesis, Massachusetts Institute of Technology.

Ramos G.D.M., 2015. *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*. Ph.D. thesis, Delft University of Technology.

Ramos G.M., Frejinger E., and Royal K.T.H., 2012. A revealed preference study on route choices in a congested network with real-time information. *International Conference on Travel Behaviour Research*, (July).

Rapoport A., Gisches E.J., Daniel T., and Lindsey R., 2014a. Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment. *Transportation Research Part B: Methodological*, 68:154–172. ISSN 01912615. doi: 10.1016/j.trb.2014.06.007.

Rapoport A., Gisches E.J., Daniel T., and Lindsey R., 2014b. Pre-trip information and route-choice decisions with stochastic travel conditions: Theory. *Transportation Research Part B: Methodological*, 68:154–172. ISSN 01912615. doi: 10.1016/j.trb.2014.06.007.

Regier D.A., Ryan M., Phimister E., and Marra C.A., 2009. Bayesian and classical estimation of mixed logit: An application to genetic testing. *Journal of Health Economics*, 28(3):598 – 610. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2008.11.003.

Savage L.J., 1954. *The foundations of statistics*. John Wiley & Sons, Inc.

Selten R., Chmura T., Pitz T., Kube S., and Schreckenberg M., 2007. Commuters route choice behaviour. *Games and Economic Behavior*, 58(2):394–406. ISSN 08998256. doi: 10.1016/j.geb.2006.03.012.

Sheffi Y., 1985. *Urban transportation networks*. Prentice-Hall, Inc. ISBN 0139397299. doi: 10.1016/0191-2607(86)90023-3.

Sheffi Y., Hall R., and Daganzo C., 1982. On the estimation of the multinomial probit model. *Transportation Research Part A: General*. ISSN 01912607. doi: 10.1016/0191-2607(82)90071-1.

Simon H.A., 1957. *Models of man social and rational : mathematical essays on rational human behavior in a social setting*. Wiley New York.

Small K., 1987. A discrete choice model for ordered alternatives. *Econometrica*, 55(2):409–24.

Song C., Qu Z., Blumm N., and Barabási A.L., 2010. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021. ISSN 0036-8075. doi: 10.1126/science.1177170.

Srinivasan K. and Mahmassani H., 2000. Modeling Inertia and Compliance Mechanisms in Route Choice Behavior Under Real-Time Information. *Transportation Research Record: Journal of the Transportation Research Board*, 1725(January):45–53. ISSN 0361-1981. doi: 10.3141/1725-07.

Swait J., 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological*, 35(7):643–666. ISSN 01912615. doi: 10.1016/S0191-2615(00)00029-1.

Tachet R., Sagarra O., Santi P., Resta G., Szell M., Strogatz S.H., and Ratti C., 2017. Scaling law of urban ride sharing. *Scientific Reports*, 7(October 2016):1–6. ISSN 20452322. doi: 10.1038/srep42868.

Tawfik A.M., Rakha H.A., and Miller S.D., 2010. An experimental exploration of route choice: Identifying drivers choices and choice patterns, and capturing network evolution. In *13th International IEEE Conference on Intelligent Transportation Systems*, pages 1005–1012. ISSN 2153-0009. doi: 10.1109/ITSC.2010.5625253.

Taylor M.A., 2013. Travel through time: the story of research on travel time reliability. *Transportmetrica B: Transport Dynamics*, 1(3):174–194. ISSN 2168-0566. doi: 10.1080/21680566.2013.859107.

Thomas T. and Tutert S., 2010. Route choice behavior based on license plate observations in the dutch city of enschede. In *Seventh Triennial Symposium on Transportation Analysis: Tristan VII, Tristan*.

Thurstone L.L., 1927. A law of comparative judgment. *Psychological Review*, 34:273–286.

Train K., 2001. A Comparison of hierarchical bayes and MLE for mixed logit. Technical report.

Train K.E., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press. ISBN 9780511753930. doi: 10.1017/CBO9780511753930.

U.N., 2019. World urbanization prospects 2018: Highlights. Technical report, United Nations, Department of Economic and Social Affairs, Population Division (ST/ESA/SER.A/421).

van Essen M., Thomas T., van Berkum E., and Chorus C., 2016. From user equilibrium to system optimum: a literature review on the role of travel information, bounded rationality and non-selfish behaviour at the network and individual levels. *Transport Reviews*, 36(4):527–548. ISSN 0144-1647. doi: 10.1080/01441647.2015.1125399.

Vovsha P., 1997. Application of cross-nested logit model to mode choice in tel aviv, israel, metropolitan area. *Transportation Research Record*, 1607(1):6–15.

Vreeswijk J., Landman R., van Berkum E., Heygi A., Hoogendoorn S., and van Arem B., 2015. Improving the road network performance with dynamic route guidance by considering the indifference band of road users. *IET Intelligent Transport Systems*, 9(10):1–10. ISSN 1751-956X. doi: 10.1049/iet-its.2014.0258.

Vreeswijk J., Rakha H., Berkum E.V., and Arem B.V., 2014. Analysis of inertial choice behaviour based expected and experienced savings from a real-world route choice experiment. pages 1–21.

Walker J. and Ben-Akiva M., 2002. Generalized random utility model. *Mathematical Social Sciences*, 43(3):303 – 343. ISSN 0165-4896. doi: https://doi.org/10.1016/S0165-4896(02)00023-9. Random Utility Theory and Probabilistic measurement theory.

Wang P., Hunter T., Bayen A.M., Schechtner K., and González M.C., 2012. Understanding road usage patterns in urban areas. *Scientific Reports*, 2. ISSN 20452322. doi: 10.1038/srep01001.

Wardrop J.G., 1952. Road paper. some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362. doi: 10.1680/ipeds.1952.11259.

Watling D.P., Rasmussen T.K., Prato C.G., and Nielsen O.A., 2018. Stochastic user equilibrium with a bounded choice model. *Transportation Research Part B: Methodological*, 114:254 – 280. ISSN 0191-2615. doi: https://doi.org/10.1016/j.trb.2018.05.004.

Wilcoxon F., 1945. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83. ISSN 00994987.

Williams H.C.W.L., 1977. On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A: Economy and Space*, 9(3):285–344.

Yildirimoglu M. and Kahraman O., 2018a. Investigating the empirical existence of equilibrium conditions. In *Transportation Research Board 97th Annual Meeting, Washington D.C.*

Yildirimoglu M. and Kahraman O., 2018b. Searching for empirical evidence on traffic equilibrium. *PLoS ONE*, 13(5):1–16. ISSN 19326203. doi: 10.1371/journal.pone.0196997.

Zhao K., Musolesi M., Hui P., Rao W., and Tarkoma S., 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific Reports*, 5:1–7. ISSN 20452322. doi: 10.1038/srep09136.

Zhu S. and Levinson D., 2015. Do people use the shortest path? An empirical test of wardrop's first principle. *PLoS ONE*, 10(8):1–18. ISSN 19326203. doi: 10.1371/journal.pone.0134322.

ZIMMERMANN M. AND FREJINGER E., 2019. A tutorial on recursive models for analyzing and predicting path choice behavior.