



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **MODELLING MODE AND ROUTE CHOICES ON PUBLIC TRANSPORT SYSTEMS**

**SEBASTIÁN RAVEAU FELIÚ**

Thesis submitted to the Office of Research and Graduate Studies in  
partial fulfilment of the requirements for the Degree of Doctor in  
Engineering Sciences

Advisor:

**JUAN CARLOS MUÑOZ ABOGABIR**

Santiago de Chile, April 2014

© 2014, Sebastián Raveau Feliú



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# MODELLING MODE AND ROUTE CHOICES ON PUBLIC TRANSPORT SYSTEMS

**SEBASTIÁN RAVEAU FELIÚ**

Members of the Committee:

**JUAN CARLOS MUÑOZ ABOGABIR**

**JUAN DE DIOS ORTÚZAR SALAS**

**LEONARDO BASSO SOTZ**

**ANTONIO GSCHWENDER KRAUSE**

**CARLO GIACOMO PRATO**

**CRISTIÁN VIAL EDWARDS**

Thesis submitted to the Office of Research and Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor in Engineering Sciences

Santiago de Chile, April 2014

To all public transport users who,  
like me, everyday wait, travel,  
transfer and look at maps

## ACKNOWLEDGMENTS

When writing a doctoral thesis, one might think that the Acknowledgments Section would be the easiest part (after all, there are no models or estimations involved). But that is not true. This Section could easily be the longest one. Over the last five years I had the amazing opportunity of interacting and working with a so many wonderful people. That has been the most valuable part of my doctorate.

First and foremost, I would like to thank my advisor Juan Carlos Muñoz for undertaking this journey with me. It has been an honour and a delight working with him. There are so many things I am grateful for, professionally and personally: his guidance and ideas towards conducting a better research, his trust in my work, his capacity of always going one step further and asking “what happens if?”, his sense of humour and constant joy, and the opportunities he offered me, just to name a few.

I would like to thank Juan de Dios Ortúzar, not only for his contributions as a committee member, but mostly for his contribution in educating me as a researcher. It was with him that I began my career as a researcher, not so many years ago. His advice and constant availability have been fundamental and have moulded me.

I would like to thank Louis de Grange for his invaluable support, for his constant ideas for joint research, for the opportunities he provided, for his faith in me, and for coming up with such a fantastic concept as the “angular cost”. I am in debt with him.

I would like to thank Carlo Prato, whom I came to idolize during my doctorate. When I began this research, all I had as a guide me was a review article he wrote on route choice modelling. Five years later, the concepts from that review article have become my doctoral thesis. Not only I had the great honour of having Carlo in my committee, but I also had the opportunity of visiting him in Copenhagen and working closely with him.

I would like to thank Leonardo Basso and Antonio Gschwender, not only for their contributions as members of my committee, but also for their contributions in my academic formation. I began this research with no experience in public transport modelling and analysis, but fortunately I had the opportunity of taking courses with both of them in topics related.

I would like to thank Francisca Yáñez, Zhan Guo and Nigel Wilson for their contributions and support in different stages of my research. I am proud to have gotten publications with all of them as a result of joint collaborations.

I would like to thank Joan Walker for hosting me for a year as a visiting scholar researcher at UC Berkeley, where I had the opportunity of work and learn from her entire research group. I would also like to thank Stefano Scagnolari, Akshay Vij, André Carrel and Francesc Soriguera for their contributions to my research. The year I spent in Berkeley was the best year of my doctorate.

I would like to thank all my fellow graduate students, for their friendship, the good times and the support. Thanks to Felipe Delgado, Felipe Zúñiga, Tomás Pinochet, Felipe Herrera, Francisca Navarrete, Luis Aliaga, Philippe Burq, Pablo Mandujano, Homero Larrain, Julián Arellana, Christopher Bucknell, Daniel Hernández, and Felipe Ortíz.

I would like to thank all the people of my Department of Transport Engineering and Logistics, as they are (and forever will be) embedded in me. Thanks to Ignacia Torres for his constant support and constant help, this doctorate has certainly been nicer and cuter because of her. Thanks to Joaquín de Cea, Patricia Galilea, Juan Carlos Herrera, Luis Rizzi and Juan Enrique Coeymans, for being both professors and friends. Thanks to María Pía Vargas, Ana Lorena Olguín, Cyndi Romero, Juvenal Díaz and Katherine Silva for all the help they provided.

I would like to thank all the people that helped me so many times in so many things. Thanks to the FONDEF team: Alexandra Soto, Alfonso Salgado, Paula Iglesias, Carlos

Melo, Felipe González, Ignacio Abud and Sebastián Tamblay. Thanks to the DIPEII: Fernanda Kattan, Debbie Meza and Danisa Herrera. Thanks to my team of survey takers: Francisca Feliú (and her friends), David Rodríguez, Romina Concha, José Ignacio Benavente and Karina Cáceres.

I would like to thank the Chilean Council for Scientific and Technological Research (CONICYT) for their fundamental support through their Scholarship for Doctoral Studies and their Scholarship for Attendance to Conferences. I would also like to thank the BRT Centre of Excellence, the Centre for Sustainable Development (CEDEUS), the Millennium Institute on Complex Engineering Systems, the Chilean Fund for Scientific and Technological Development (FONDEF), the Chilean Fund for Scientific and Technological Research (FONDECYT), and Pontificia Universidad Católica de Chile for their support.

Lastly, I would like to thank myself. More precisely, I would like to thank the 2009 version of myself. Starting a doctorate seemed like a big and risky decision at the time, but looking back it was the best decision I could have made. I regret nothing. It has been a wonderful ride.

# CONTENTS

	Page
DEDICATION .....	ii
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES .....	x
LIST OF FIGURES.....	xii
RESUMEN.....	xiii
ABSTRACT .....	xv
1 INTRODUCTION .....	1
1.1 Objectives.....	2
1.2 Hypotheses .....	3
1.3 Methodology .....	3
1.4 Contents and Contribution .....	5
1.4.1 Chapter 2 – Mode Choice Modelling .....	5
1.4.2 Chapter 3 – Route Choice Modelling .....	6
1.4.3 Chapter 4 – Route Choice Comparison .....	6
1.4.4 Chapter 5 – Route Choice Strategies .....	7
1.4.5 Chapter 6 – Public Transport Planning.....	8
1.5 Scope .....	8
2 TRANSPORT PREFERENCES AND LATENT VARIABLES .....	9
Abstract.....	9
2.1 Introduction .....	10
2.2 Hybrid Discrete Choice Models.....	11
2.3 Application to the Santiago Panel .....	14
2.3.1 Structure of the Models.....	15
2.3.2 Estimation Results and Analysis.....	17
2.4 Conclusions .....	23
Acknowledgments .....	24
Publication History .....	24

	References.....	24
3	A TOPOLOGICAL ROUTE CHOICE MODEL FOR METRO .....	28
	Abstract.....	28
	3.1 Introduction .....	29
	3.2 Specification of the Model .....	32
	3.2.1 General Description of the Model.....	34
	3.2.2 Explanatory Variables.....	34
	3.3 Results and Analysis .....	38
	3.3.1 Perceptions and Valuations.....	41
	3.3.2 Forecasting Application .....	43
	3.3.3 Policy Applications.....	46
	3.4 Conclusions .....	47
	Acknowledgments .....	49
	Publication History .....	49
	References.....	49
4	ROUTE CHOICE MODELLING ON METRO NETWORKS .....	51
	Abstract.....	51
	4.1 Introduction .....	52
	4.2 Route Choice Modelling .....	53
	4.2.1 Discrete Choice Models and Correlation.....	53
	4.2.2 Metro Networks for Analysis .....	55
	4.2.3 Topology and Map Effects .....	56
	4.2.4 Set of Alternative Routes .....	58
	4.2.5 Model Description .....	60
	4.3 Results and Analysis .....	63
	4.3.1 Perceptions and Valuations.....	65
	4.3.2 Specification of Topological Variables .....	68
	4.3.3 Model Specification and Information Omission.....	69
	4.3.4 Implications on Transportation Planning.....	71
	4.4 Conclusions .....	72
	Acknowledgments .....	73
	Publication History .....	74
	References.....	74

5	ROUTE CHOICE STRATEGIES ON TRANSIT NETWORKS .....	77
	Abstract.....	77
	5.1 Introduction .....	78
	5.2 Route Choice Strategies .....	78
	5.2.1 Minimum Itineraries .....	80
	5.2.2 Minimum Routes .....	81
	5.2.3 Minimum Hyper-Routes .....	83
	5.2.4 Route Choice Strategies and Uncertainty .....	84
	5.3 Transit Travel Survey.....	88
	5.4 Modelling Route Choice Strategies.....	91
	5.4.1 Modelling Travellers' Strategies .....	92
	5.4.2 Modelling Travellers' Route Choice .....	93
	5.5 Conclusions .....	95
	Acknowledgments .....	96
	Publication History .....	96
	References.....	96
6	A PLANNING TOOL FOR PUBLIC TRANSPORT ANALYSIS .....	98
	Abstract.....	98
	6.1 Introduction .....	99
	6.2 Methodological Framework .....	100
	6.2.1 Modelling Travel Strategies.....	100
	6.2.2 Modelling Stop Choice .....	101
	6.2.3 Modelling Route Choice.....	102
	6.2.4 Travel Assignment.....	103
	6.3 Application to a Real Network.....	104
	6.3.1 Modelling Travellers' Behaviour.....	105
	6.3.2 Tactical Planning Tool.....	106
	6.3.3 Trip Planning Tool.....	110
	6.4 Conclusions .....	112
	Acknowledgments .....	113
	Publication History .....	113
	References.....	113
7	CONCLUSIONS .....	115

8	REFERENCES .....	119
---	------------------	-----

## LIST OF TABLES

	Page
Table II-1: Perception Survey .....	16
Table II-2: Structural Equations' Results – Accessibility/Comfort .....	18
Table II-3: Structural Equations' Results - Reliability.....	18
Table II-4: Structural Equations' Results - Safety .....	19
Table II-5: Measurement Equations' Results – Accessibility/Comfort .....	20
Table II-6: Measurement Equations' Results – Reliability.....	20
Table II-7: Measurement Equations' Results – Safety.....	20
Table II-8: Discrete Choice Model's Results .....	21
Table II-9: Attribute Valuations .....	22
Table III-1: O-D Pairs with More than One Reasonable Route Alternative .....	33
Table III-2: Estimates for Route Choice Models .....	39
Table III-3: Mean Square Error for Route Choice Models .....	41
Table III-4: Trip Assignment for Model Application .....	41
Table III-5: Transfer Valuations by Station Characteristics .....	43
Table III-6: Trip Assignment for Model Application .....	45
Table IV-1: Route Availability Statistics .....	59
Table IV-2: Parameters Estimates for London and Santiago .....	64
Table IV-3: Marginal Rates of Substitution - Morning Peak & Restrictive Purpose	66

Table IV-4: Transferring Valuations - Morning Peak & Restrictive Purpose .....	67
Table IV-5: Specification of Topological Variables .....	69
Table IV-6: Information Omission - Morning Peak & Restrictive Purpose .....	70
Table V-1: Frequencies of the Transit Lines for the Network Example .....	79
Table V-2: Results of the Route Choice Strategies Example.....	85
Table V-3: Results of the Behaviour Strategies Model.....	92
Table V-4: Probabilities of Considering Common Lines .....	93
Table V-5: Results of the Mode/Route Choice Model.....	94
Table V-6: Subjective Monetary Valuations and Marginal Rates of Substitution.....	95
Table VI-1: Probabilities of Considering Common Lines .....	105
Table VI-2: Mode/Route Choice Model Parameters.....	106
Table VI-3: Stop Choice Probabilities on Plaza de Maipú .....	110

## LIST OF FIGURES

	Page
Figure 2-1: MIMIC Model for the Santiago Panel.....	16
Figure 3-1: Santiago Metro Network Maps .....	32
Figure 3-2: Angular Cost Definition .....	37
Figure 3-3: Asymmetry of Users' Route Choices.....	38
Figure 3-4: Network for Model Application.....	44
Figure 3-5: Change in the Santiago Metro Network Map.....	46
Figure 4-1: London Underground and Santiago Metro Networks Topologies .....	57
Figure 5-1: Transit Network Example .....	79
Figure 5-2: Geographic Information of the Survey.....	89
Figure 6-1: Definition of Attractive Stops .....	102
Figure 6-2: Assignment Flows for Santiago .....	107
Figure 6-3: Load Profiles on Metro Line 1 .....	108
Figure 6-4: Load Profiles on Santa Rosa Corridor.....	109
Figure 6-5: Trip Planning Tool Interface .....	111

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

MODELACIÓN DE LA ELECCIÓN DE MODO Y RUTA  
EN UN SISTEMA DE TRANSPORTE PÚBLICO

Tesis enviada a la Dirección de Investigación y Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

SEBASTIÁN RAVEAU FELIÚ

RESUMEN

Entender y modelar las decisiones de usuarios de sistemas de transporte público, para analizar sus elecciones y predecir correctamente los flujos en la red, es un elemento fundamental de la planificación urbana. De esta manera, es necesario identificar, cuantificar y modelar los factores relevantes que influyen en el comportamiento y la toma de decisiones individuales. Los modelos tradicionales de demanda por transporte tienden a considerar únicamente como variables explicativas algunos atributos tangibles, obviando otros atributos intangibles que afectan las actitudes y percepciones de los viajeros. Así, existe la necesidad de mejorar y extender la especificación de modelos de demanda por transporte, incorporando una mayor variedad de variables explicativas.

Este estudio busca entender cómo los viajeros escogen sus modos y rutas al viajar en sistemas de transporte público, identificando los factores relevantes que son tomados en consideración y cuantificando el impacto que tienen diferentes características del sistema sobre las preferencias de los viajeros. En particular, se analiza el impacto de diferentes características socioeconómicas sobre la percepción de los viajeros respecto a variables como accesibilidad, comodidad, confiabilidad y seguridad. Adicionalmente, este estudio incorpora percepciones y preferencias respecto de una amplia gama de factores (tales como la ocupación de vehículos, los transbordos y la topología de la red) para mejorar la capacidad explicativa y predictiva de los modelos de demanda por transporte.

Adicionalmente, se analizan las diferentes estrategias de elección de ruta que los usuarios de transporte público pueden seguir. Estas estrategias están asociadas a la decisión de esperar una línea particular de transporte público, o abordar la primera línea (entre un conjunto definido de “líneas atractivas”) que arriba a la parada. Mientras los modelos tradicionales tienen a asumir que todos los viajeros siguen la misma estrategia, acá se establece una relación directa entre las características socioeconómicas de los individuos y las estrategias que siguen.

Los resultados indican que los modelos propuestos presentan un ajuste superior y superan a modelos tradicionales (que ignoran el efecto de las actitudes y percepciones) en términos de capacidad predictiva. Cuando se ignoran factores asociados al comportamiento, serios problemas de sesgo pueden presentarse en las utilidades marginales y las tasas marginales de sustitución (tales como los valores del tiempo) que se desprenden de los modelos.

Finalmente, se presenta un marco metodológico para la aplicación de los modelos propuestos en planificación de sistemas de transporte público. Se desarrollan dos herramientas en base a los modelos formulados: (i) una herramienta táctica de planificación (para ser usada por autoridades, planificadores y operadores), y (ii) una herramienta para planificar viajes (diseñada y orientada para interactuar con los usuarios). Así, los principales aportes de este estudio son extendidos para realizar aportes reales en sistemas de transporte público, y no ser simplemente una contribución teórica.

Miembros de la Comisión de Tesis Doctoral

Juan Carlos Muñoz Abogabir  
Juan de Dios Ortúzar Salas  
Leonardo Basso Schotz  
Antonio Gschwender Krause  
Carlo Giacomo Prato  
Cristián Vial Edwards

Santiago, Abril 2014

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

MODELLING MODE AND ROUTE CHOICES  
ON PUBLIC TRANSPORT SYSTEMS

Thesis submitted to the Office of Research and Graduate Studies in partial fulfilment of the requirements for the Degree of Doctor in Engineering Sciences by

SEBASTIÁN RAVEAU FELIÚ

ABSTRACT

Understanding and modelling travellers' decisions on public transport systems by correctly analysing their choices and being able to forecast flows on the network are essential elements in urban planning. This way, it is necessary to identify, quantify and model all the relevant factors influencing individual behaviour and decision-making. Traditional travel demand models tend to consider only a few tangible attributes as explanatory variables, obviating other –many times intangible - attributes that affect travellers' attitudes and perceptions. Thus, there is a need to enhance and extend the specification of travel demand models, by incorporating a larger variety of explanatory variables.

This study seeks to understand how travellers choose their modes and routes when travelling in a public transport system, identifying the relevant factors that are taken into account and quantifying the impact that different characteristics of the system have on the preferences of travellers. The impacts that different socio-economic characteristics have on travellers' perceptions of accessibility, comfort, reliability and safety are studied. Additionally, the study incorporates perceptions and preferences regarding a wide variety of factors (such as crowding, transferring and network topology) to enhance the explanatory and forecasting capabilities of travel demand models.

Additionally, this study analyses different route choice strategies that public transport users can follow. These are related to the decision of either waiting for a particular public transport line, or boarding the first line (among a defined set of “attractive lines”) that arrives to the stop. While traditional models tend to assume that all travellers follow the same strategy, this study establishes a direct relationship between the individuals’ socio-economic characteristics and the strategy they tend to follow.

Results show that when an effort to understand and accommodate behavioural factors is made, the resulting models provide superior goodness-of-fit and dominate traditional models (that ignore the effect of subjective attitudes and perceptions) in terms of forecasting ability. When behavioural factors are ignored, serious biases may arise in terms of marginal utilities and marginal rates of substitution (such as the value of time).

Lastly, this study presents an integrated methodological framework to apply the proposed choice models to public transport planning. Two planning tools are developed based on the models formulated: (i) a tactical planning tool (designed for authorities, planners and operators), and (ii) a trip planning tool (designed to orient and interact with the travellers). This way, the main findings of the study are extended to make actual improvements and contributions to a given public transport system, instead of just being a theoretical contribution.

Members of the Doctoral Thesis Committee:

Juan Carlos Muñoz Abogabir  
Juan de Dios Ortúzar Salas  
Leonardo Basso Schotz  
Antonio Gschwender Krause  
Carlo Giacomo Prato  
Cristián Vial Edwards

Santiago, April 2014

## 1 INTRODUCTION

Public transport investment is fundamental in the development of any growing city (Hensher, 2007) from social (provides mobility alternatives to the population), economic (public transport systems tend to present economies of scale in their operation) and sustainable perspectives. This is particularly important nowadays, due to the worldwide increasing tendency in the use of the car, as a result of the economic growth in developing countries and the opening international markets. The consequences of this growth can be seen daily on the streets: higher congestion levels, more traffic accidents, and higher emissions.

To effectively promote the use of public transport, it is fundamental to understand the decision-making process of public transport travellers, as well as their preferences and perceptions. Travel decisions are made at two levels: (i) choice of public transport mode within the system (bus, metro, combined modes, etc.) and (ii) choice of travel route (selection of public transport lines and transfer points along the way). These decisions are interrelated, and their order might depend on a particular public transport system's characteristics. Traditional mode and route choice models tend to consider tangible attributes only (i.e. those that are easily measured and understood, such as fare, travel time, and number of transfers) to explain travellers' decisions (Ortúzar and Willumsen, 2011). Nevertheless, it is well known that other attributes, many times intangible (i.e. without a measurement scale, such as safety and reliability), are also considered by individuals (Koppelman and Pas, 1980; Ben-Akiva *et al.*, 2002; Raveau *et al.*, 2010). These variables cannot be quantified directly and are highly subjective; therefore there is a need to study them to understand their effect on individual decisions.

The impact that different characteristics of the public transport system, such as infrastructure and services, have on travellers' choices of mode and route is an interesting subject. This allows evaluating and analysing the public transport system on a tactical level. Currently, from a behavioural point of view, there is a need for better

methodological tools to evaluate the effects that tactical changes could have on public transport demand (e.g. modifications to existing routes, implementation of bus corridors, and addition of express services to the system).

## **1.1 Objectives**

The general objective of this thesis is to understand how travellers choose their modes and routes when travelling in a public transport system, identifying the relevant factors that individuals take into account and quantifying the impact that different characteristics of the system (such as station infrastructure, available information and schematic maps) have on the preferences of the travellers.

As specific objectives, this thesis seeks to:

- 1) Identify and model the relevant factors that affect mode and route choices in a public transport system, focusing specially on those attributes that are intangible and are usually ignored in the literature.
- 2) Analyse and compare the decisions of public transport users in different contexts and networks, identifying the similarities and differences in their preferences.
- 3) Analyse the different criteria and strategies that individuals can follow when choosing their travel alternatives within a public transport system, studying the advantages of considering these various strategies in the models.
- 4) Explore a methodology to apply the proposed travel behaviour models in a framework allowing tactical policy evaluation of urban public transport system intervention.

## **1.2 Hypotheses**

The hypotheses that support this research are as follow:

- 1) Travellers take into account a wide variety of factors when choosing their mode and route in a public transport system, beyond traditionally considered factors (such as travel times and fare), and these must be included when modelling their behaviour.
- 2) Although the characteristics of a particular system can influence the perceptions and preferences of travellers, it is possible to propose a behavioural modelling framework that can be applied to any public transport system.
- 3) Different individuals within a public transport system can follow different criteria and strategies when choosing their travel alternatives. Capturing these differences can improve significantly the modelling results.
- 4) A proper formulation of mode and route choice models, including relevant behavioural elements, can be embedded in a framework that allows the analysis and evaluation of tactical changes in a public transport system.

## **1.3 Methodology**

The study seeks to answer a main research question: what are the relevant factors that affect mode and route choices within a public transport system? The answer to that question determines, at the end, the levels and structure of public transport demand. To achieve this objective, this research is based on two main tasks: data collection and mathematical modelling. An additional goal of the study is to implement the resulting models for forecasting and public transport evaluation.

This study focuses almost exclusively in analysing the travel decisions of public transport users, without analysing the decisions of travellers of other modes (most significantly, car users). Although the decision of choosing public transport modes over other alternatives is a significant one, this study only analyses the subsequent decisions once the travellers have decided to use the public transport system.

The mode and route choice data was obtained from revealed-preference surveys conducted in Santiago, Chile and London, United Kingdom. The data considers origin-destination information with different travel structures (e.g. modes and lines considered), different socio-economic characteristics of travellers, different perceptions and attitudes towards the alternatives, and different sets of non-chosen available alternatives. When sampling origins and destinations, four aspects were considered: (i) that trips between these O-D pairs were representative of the city's network and demand structure; (ii) that the travel alternatives (in terms of modes and routes) were diverse enough to capture the desired effects; (iii) that the size of the problem (i.e. the additional information required to complete the data bank, such as levels-of-service) remained manageable, and (iv) that demand was large enough to avoid potential sampling biases.

The modelling of mode and route choices was based on discrete choice models (i.e. the decision is made among a finite number of available transport alternatives) according to the Random Utility Theory approach (McFadden, 1974). This approach proposes that individuals consider and value the combination of attributes (both tangible and intangible) related to each alternative, choosing the option with the "highest utility level".

This thesis seeks to be a significant contribution in understanding and modelling the decisions that public transport users make when travelling, by analysing and quantifying behavioural factors that have been ignored so far in the literature. A more detailed analysis of these factors can enhance significantly the mathematical models used in forecasting and evaluation. Therefore, this study also seeks to provide a framework for the application of the proposed models to create useful public transport planning tools.

## **1.4 Contents and Contribution**

This document presents the main findings of the study, analysing what are the relevant factors that affect mode and route choices within a public transport system, to effectively understand how travellers make their decisions.

The document is organized in six additional chapters. Chapter 2 through Chapter 6 seek to answer each of the objectives of the study, as presented in Section 1.1. Chapter 7 provides the main conclusions of the study. It is important to notice that Chapters 2 through 6 are organized and formatted as individual articles (the main scientific results of the research). This way, each chapter is self-contained and can be read without the strict necessity of reading the remaining chapters. Although this facilitates reading the document, it has the inevitable drawback of having to provide some redundant contents among the different chapters (mainly in terms of methodology).

### **1.4.1 Chapter 2 – Mode Choice Modelling**

Chapter 2 (along with Chapter 3) deals with the first objective presented in Section 1.1. In this chapter, choice models are formulated to explain travellers' mode preferences. This is the only chapter where non-public modes are also considered in the analysis. Four different types of modes are analysed: (i) pure private transport modes; (ii) pure public transport modes; (iii) combined public transport modes, and (iv) combined private-public modes.

The main novelty of the study presented in Chapter 2 is the use of hybrid discrete choice models to understand travellers' decisions. Latent variables are constructed to capture intangible and unquantifiable factors that are taken into account by individuals, and are generally ignored in traditional models.

The study case corresponds to a five-day pseudo diary of morning commuters in Santiago, Chile. Traditional variables (such as monetary cost, travel time and transfers) are

complemented with latent variables to measure the travellers' perception of safety, accessibility, comfort and reliability.

### **1.4.2 Chapter 3 – Route Choice Modelling**

Chapter 3 also deals with the first objective presented in Section 1.1. This chapter presents a complete analysis of route choices within the metro network of Santiago. The study proposes a wide variety of factors that might influence the decisions and preferences of travellers, and values their impact on the decision-making process of individuals.

Among the considered variables, the main novelty lies in the inclusion of topological factors to explain how the spatial perceptions of travellers might affect their decisions. These topological factors are influenced by the design of the schematic maps used to provide information to travellers. Additionally, the transfer experience and comfort are analysed in detail.

When the proposed model is compared with a traditional model (that is limited in terms of the variables considered), satisfactory results are obtained in terms of goodness-of-fit, explanatory power and forecasting capability.

### **1.4.3 Chapter 4 – Route Choice Comparison**

Chapter 4 provides a behavioural comparison of route choices within the metro systems of Santiago and London. The comparison is made by defining a common specification for the choice models of both cities; this is not commonly done in the literature when comparisons are made. The study seeks to identify, quantify and contrast the different factors that travellers from both cities take into account when choosing their routes.

Five types of variables were found to be significant: (i) different time components (which are perceived differently); (ii) transfer-related variables; (iii) measures of occupancy and

comfort; (iv) topological factors, which mainly come from the schematic representation of the networks, and (v) socio-demographic information from the travellers.

An interesting result is that all the variables considered were significant in both cities, although they might be perceived differently (for example, travellers in Santiago prefer waiting to walking, while travellers in London prefer the opposite). Having found so similar results in such different networks, the proposed specification could represent a good approximation for analysing route choice decisions on other systems.

#### **1.4.4 Chapter 5 – Route Choice Strategies**

Chapter 5 deals with the different strategies that travellers can follow when selecting public transport lines. Particularly, it deals with the decision of either waiting for a particular line, or boarding the first one (among a defined set of “attractive lines”) that arrives to the stop. Results show that different individuals may follow different strategies, while in the literature it is traditionally assumed that all travellers behave in the same way.

The study presents an analysis of the strategy decisions of public transport users in Santiago, while also analysing the implications of strategy selection on route choices. Strategy selection is modelled as an endogenous socio-economic variable, which depends on the individual’s income, age and familiarity with the network.

Results show that there are significant differences in terms of choice of strategy among travellers, so the assumption of uniform strategy selection in classical models can lead to explanatory and forecasting errors. Additionally, there are significant differences between the marginal rates of substitution of different strategy choosers, i.e., those who consider a set of “attractive lines” have higher monetary valuations.

### **1.4.5 Chapter 6 – Public Transport Planning**

Chapter 6 presents a methodological framework for modelling travel decisions on public transport networks. The framework considers three levels of selection: (i) choice of travelling strategy; (ii) choice of stop (which also implies choice of mode), and (iii) choice of route.

The main result of the proposed framework is the implementation of two planning tools: (i) a tactical planning tool (designed for authorities, planners and operators), and (ii) a journey planning tool (designed for travellers). Both tools were implemented for the integrated public transport system of Santiago. The results of the tactical planning tool show a high fit between observed and modelled flows, correctly reproducing the aggregate travel patterns.

### **1.5 Scope**

The main limitation of this research was the limited availability of data. Unlike private transport analysis, there are few public transport databases from which route choice can be analysed. Nevertheless, in the near future this might change, as vehicle GPS data increasingly becomes more available. This way, it could be possible to apply the proposed mode/route choice models on other public transport systems. As the data used on this study only covers buses (in Santiago) and metro (in London and Santiago), other public transport modes remain to be studied.

There is a need to expand the results presented in this study, which mainly focus on the morning peak period, to different times of the day. This is particularly important for both planning tools, which must be capable of considering the entire day. This would require additional information regarding levels of service and demand patterns, to calibrate the corresponding models.

**2 TRANSPORT PREFERENCES AND LATENT VARIABLES****Sebastián Raveau**Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile**María Francisca Yáñez**Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile**Juan de Dios Ortúzar**Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile**Abstract**

Travel demand models typically use only objective modal attributes as explanatory variables. Nevertheless, it has been well known for many years that attitudes and perceptions also influence users' behaviour. The use of hybrid discrete choice models constitutes a good alternative to incorporate the effect of subjective factors through the construction of latent variables. This study applies the hybrid approach to a real urban case study, quantifying the improvements in explanatory capability and analysing the advantages of considering latent variables. Our results show that hybrid models are clearly superior to traditional models that ignore the effect of subjective attitudes and perceptions.

## 2.1 Introduction

Traditionally, discrete choice models have considered only objective/measurable attributes from the alternatives and socio-economic characteristics of the individuals as explanatory variables. The inclusion of subjective elements in discrete choice models has re-emerged as an analysis and discussion topic, after losing some of the importance that had made it an interesting subject in the late 1970s (Koppelman and Hauser, 1978; Prashker, 1979; Koppelman and Pas, 1980).

To capture the impact of subjective factors over the decision process, during the last decade a new breed of *hybrid choice* models have been developed; these allow including not only tangible attributes but also more intangible elements associated with users' perceptions and attitudes, through latent variables (McFadden, 1986; Ben Akiva *et al.*, 2002). Moreover, it has been shown that introducing latent variables helps to improve choice model fit (Ashok *et al.*, 2002; Raveau *et al.*, 2010).

The main objective of this study is to estimate hybrid choice models in a panel data context, with a particular emphasis on understanding how latent variables can influence the choice of public modes. Based on an application to data from the Santiago Panel (Yáñez *et al.*, 2010a), we provide empirical evidence that hybrid choice models can substantially improve traditional practice, which ignores the effect of subjective attitudes and perceptions.

The study is organized as follows. In Section 2.2, we summarise the background supporting our research, and present the theoretical formulation for estimation of hybrid discrete choice models including latent variables. In Section 2.3, we discuss the results of an application to a real urban case using observations from the five-day pseudo diary corresponding to the fourth wave of the Santiago Panel, and in Section 2.4 we present our main conclusions.

## 2.2 Hybrid Discrete Choice Models

Latent variables are a measure of psychological factors (such as attitudes and perceptions) that, although influencing individual behaviour, cannot be quantified in practice. This is because of their intangibility (as these factors do not have a measurement scale) or their intrinsic subjectivity (because different people may perceive them differently). Identification of latent variables requires supplementing a standard preference survey with questions that capture users' perceptions about some aspects of the alternatives (and choice context). The answers to these questions generate *perception indicators* that serve to identify and measure the latent variables.

In the recent literature, the most popular approach to include the effect of latent variables is using a MIMIC (Multiple Indicator Multiple Cause) model (Bollen, 1989), where the latent variables ( $\eta_{ilq}$ ) are explained by characteristics  $s_{iqr}$  from the users and from the alternatives through *structural equations* as (2.1). At the same time, the latent variables explain the perception indicators ( $y_{ipq}$ ), which are observed by the modeller from the survey, through measurement equations as (2.2):

$$\eta_{ilq} = \sum_r \alpha_{ilr} \cdot s_{iqr} + v_{ilq} \quad (2.1)$$

$$y_{ipq} = \sum_l \gamma_{ilp} \cdot \eta_{ilq} + \zeta_{ipq} \quad (2.2)$$

where the index  $i$  refers to an alternative,  $q$  to an individual,  $l$  to a latent variable,  $r$  to an explanatory variable and  $p$  to an indicator;  $\alpha_{ilr}$  and  $\gamma_{ilp}$  are parameters to be estimated, while  $v_{ilq}$  and  $\zeta_{ipq}$  are error terms with mean zero and a certain covariance matrix. As the  $\eta_{ilq}$  terms are unknown, both equations must be considered jointly in the parameter estimation process (Bolduc and Alvarez-Daziano, 2010).

Traditionally, in discrete choice modelling it is assumed that people are rational decision makers maximising their perceived utility  $U_{iq}$  over the alternatives; the modeller, who is an observer, defines a representative utility  $V_{iq}$  and (as he does not have perfect information) an error term  $\varepsilon_{iq}$  associated with each alternative (Ortúzar and Willumsen, 2011), such that:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad (2.3)$$

The representative utility  $V_{iq}$  is a function of the tangible attributes  $X_{ikq}$ , where  $k$  refers to a particular attribute (such as travel time, fare, or socio-economic characteristics of the individual); if latent variables are also included, a utility function such as (2.4) is postulated, where  $\theta_{ik}$  and  $\beta_{il}$  are parameters to be estimated associated with the tangible attributes and the latent variables, respectively.

$$V_{iq} = \sum_k \theta_{ik} \cdot X_{ikq} + \sum_l \beta_{il} \cdot \eta_{ilq} \quad (2.4)$$

Since the latent variables ( $\eta_{ilq}$ ) are unknown, the discrete choice model must be estimated with the MIMIC model structural (2.1) and measurement (2.2) equations. Finally, to characterise the individuals' decisions over their set of available alternatives (defined as  $A_q$ ), binary variables  $d_{iq}$ , that take values according to equation (2.5), are defined:

$$d_{iq} = \begin{cases} 1 & \text{if } U_{iq} \geq U_{jq}, \quad \forall j \in A_q \\ 0 & \text{in other case} \end{cases} \quad (2.5)$$

To estimate the MIMIC model, a subset of parameter values must be constrained. Generally, as many parameters as latent variables must be fixed, but more constraints could be necessary depending on the structure of the relationships. Stapleton (1978) recommends fixing as many measurement equations' parameters ( $\gamma_{ilp}$ ) as needed, but also shows that the model can be estimated if the variances of the structural equations ( $v_{ilq}$ ) are

fixed. Constraining the parameters of the measurement equations is a more popular approach (Ben-Akiva *et al.*, 2002; Vredin Johansson *et al.*, 2006) than constraining the variances of the structural equations (Raveau *et al.*, 2010). Based on simulations, Raveau *et al.* (2012) recommend constraining the variances  $v_{ilq}$ ; this is the approach followed in this study.

The traditional method for estimating the parameters is maximizing the likelihood of the probability of replicating the individual choices based on the representative utility proposed by the modeller, i.e.  $\Pr(d_{iq} | V_{iq})$ . From (2.4) this conditional probability can be expressed in terms of the variables and parameters of the discrete choice model as  $\Pr(d_{iq} | X_{ikq}, \eta_{ilq}, \theta_{ik}, \beta_{il})$ . However, as the latent variables are not observed, it is necessary to integrate over their whole variation range, conditioning them by their explanatory variables. Moreover, to estimate the model it is necessary to introduce the information provided by the perception indicators since otherwise the model would not be identifiable.

The indicators are not explanatory variables; instead, they are endogenous to the latent variables through (2.2). This implies that the choice probability used during estimation is given by (2.6), where  $f(\cdot)$  is the probability density function of the indicators and  $g(\cdot)$  is the probability density function of the latent variables (Ben-Akiva *et al.*, 2002). Once the functional form of the discrete choice model is defined, simulated maximum likelihood can be used for the estimation (Bolduc and Alvarez-Daziano, 2010).

$$\Pr(d_{iq}, y_{ipq} | X_{ikq}, s_{iqr}, \theta_{ik}, \beta_{il}, \alpha_{ilr}, \gamma_{ipq}) = \int_{\eta_{ilq}} \Pr(d_{iq} | X_{ikq}, \eta_{ilq}, \theta_{ik}, \beta_{il}) \cdot f(y_{ipq} | \eta_{ilq}, \gamma_{ipq}) \cdot g(\eta_{ilq} | s_{iqr}, \alpha_{ilr}) \cdot d\eta_{ilq} \quad (2.6)$$

### 2.3 Application to the Santiago Panel

The data for the study corresponds to the fourth wave of the Santiago Panel. This dataset has features of both the *short* and *long* survey panel approaches (Yáñez *et al.*, 2010a). It may be described as a five-day pseudo diary that has also five waves, one before and four after the implementation of Transantiago (Muñoz *et al.*, 2009), a radically new urban public transport system for Santiago de Chile. The panel was built between December 2006 and October 2008.

The characteristics of the trips are quite similar over these five days because the members of the Santiago Panel have a fairly static routine (88% of the sample selects the same mode each day). This behaviour is similar to what was found by Cherchi and Cirillo (2008) for the six-week data panel from Mobidrive (2000); choices tend to be more persistent for tours the main activity of which is work or study.

The initial sample consisted of 303 individuals who live in Santiago (Chile) and worked at one of the five campuses of the Pontificia Universidad Católica de Chile. By the fourth wave the sample had been reduced to 258 individuals (with five observations each), due to some respondents changing their workplaces. The first three waves of the panel did not consider the inclusion of latent variables, but to enhance the modelling results obtained after the first three waves (Yáñez and Ortúzar, 2010), on the fourth one respondents were specifically asked to evaluate different characteristics of the various modes (Raveau *et al.*, 2010). This way, latent variables were built based on both socio-economic variables and perception indicators from this fourth wave of the panel (Yáñez *et al.*, 2010b).

The need to incorporate data about users' perceptions to improve model fit was especially important here, as the panel incorporated a large change (a *shock*) generated by the introduction of Transantiago just prior to the second wave; this change (which due to the faulty initial implementation of the system created much stir in the media and population)

may have modified users' perceptions about the various modes available in the city and, as a consequence, their choices and behaviour in general.

In this study, 10 modes are considered for latent variable treatment: (i) car-driver, (ii) car-passenger, (iii) shared taxi, (iv) metro (underground), (v) bus, (vi) car-driver/metro, (vii) car-passenger/metro, (viii) shared taxi/metro, (ix) bus/metro, and (x) shared taxi/bus. For each mode, there was information available about trip times (walking, waiting and in-vehicle travel time), trip cost and number of transfers made. Regarding the users, the panel also gathered information about socio-economic variables, such as age, gender, educational level and income, among others.

### **2.3.1 Structure of the Models**

We considered three latent variables: (i) accessibility, (ii) reliability and (iii) comfort/safety; the effects of these variables were captured through seven perception indicators, based on the evaluation of various aspects of the pure modes, on a scale from 1 to 7. The perception indicators were consistent with those used in many previous studies (Nicolaidis, 1975; Prashker, 1979; Rundmo and Hale, 2003). Table II-1 describes the survey used to capture the perception indicators.

Four explanatory variables were included in the MIMIC model: educational level, with four categories (elementary school, high school, technical studies and college), age, gender and monthly income, divided into three categories (low: less than US\$680, medium: from US\$680 to US\$1,700 and high: from US\$1,700 onwards). As mentioned above, the original survey-panel design did not consider the measurement of attributes of the alternatives that could be related to the latent variables. The MIMIC model structural relationships were examined using factor analysis to guide the specification (Figure 2-1 illustrates the results of this process). The categorised variables must be represented by  $n - 1$  binary variables, leaving women, elementary school and low income as references.

Table II-1: Perception Survey

Using grades from 1 to 7, indicate your level of satisfaction about the following factors, in relation to the following modes (7 is very satisfactory and 1 least satisfactory)					
Factors	Bus	Metro	Shared Taxi	Car-Driver	Car-Passenger
A.-Possibility of calculating the waiting time prior to the trip					
B.- Possibility of calculating the travel time prior to the trip					
C.-Ease of access					
D.-Comfort during the trip					
E.-Safety regarding accidents					
F.-Safety regarding theft					
G.-Availability of suitable information					

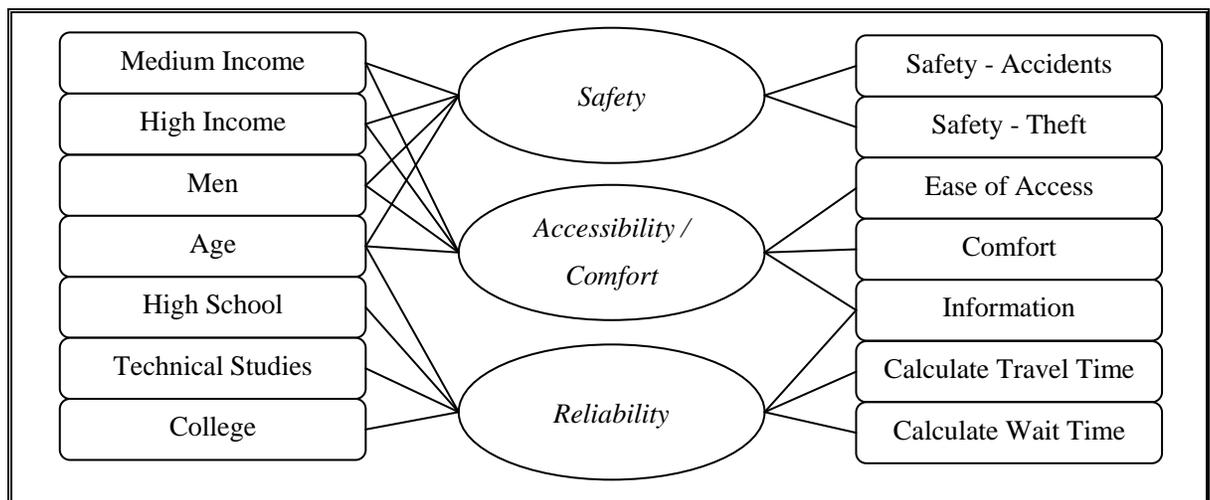


Figure 2-1: MIMIC Model for the Santiago Panel

Based on the MIMIC specification and the hypotheses regarding the indicators on the mode choice decision process, this model structure was expected (i.e. it is consistent with what was expected when the survey was designed), except for the fusion of accessibility and comfort in a unique new latent variable.

The representative utility function for the choice model includes the number of transfers during the trip as well as the different time variables measured; in the case of travel time, systematic taste variations according to the respondent's sex were found (Ortúzar and Willumsen, 2011). Travel cost was standardized by the individual's wage rate,  $w_q$ . Equation (2.7) shows the general form of the utility function corresponding to the best specification obtained among several formulations studied; it is important to mention that since the perception indicators were collected only for the pure modes, the combined modes do not possess latent variables. The model also includes a complete set of alternative specific constants, denoted by  $ASC_i$  (bus was taken as reference alternative).

$$\begin{aligned}
 V_{iq} = & \theta_{cost} \cdot Cost_{iq} / w_q + (\theta_{travel} + \theta_{men} \cdot Men_q) \cdot Ttravel_{iq} + \theta_{wait} \cdot Twait_{iq} \\
 & + \theta_{walk} \cdot Twalk_{iq} + \theta_{trans} \cdot Transfers_{iq} + \beta_{rel} \cdot Reliability_{iq} \\
 & + \beta_{acc\ com} \cdot AccessibilityComfort_{iq} + \beta_{saf} \cdot Safety_{iq} + ASC_i
 \end{aligned} \tag{2.7}$$

### 2.3.2 Estimation Results and Analysis

The hybrid model was estimated using the Python version of Biogeme (Bierlaire, 2003; Bierlaire and Fétiarison, 2009). Although all parameters are obtained simultaneously, the results are presented and analysed sequentially in three parts: (i) structural equations of the MIMIC model, (ii) measurement equations of the MIMIC model, and (iii) discrete choice model.

Table II-2 presents the results for the estimation of the structural equations related to accessibility/comfort, with their  $t$ -values. It can be seen that this latent variable is higher for men and that it decreases with the age. Therefore, men tend to find more accessible and comfortable the transport alternatives, as will younger travellers. In private transport modes, it increases with the income, while in public transport modes, it decreases with the income. This way, as expected, travellers with higher income will tend to prefer private transport modes as they usually provide a more comfortable level of service.

Table II-2: Structural Equations' Results – Accessibility/Comfort

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Age	-0.077 (-75.14)	-0.023 (-25.43)	-0.057 (-71.54)	-0.010 (-23.58)	-0.047 (-87.88)
Medium Income	1.373 (9.46)	0.172 (2.04)	-0.923 (-8.10)	-0.429 (-2.33)	-0.368 (-4.50)
High Income	2.790 (14.38)	0.971 (6.99)	-1.150 (-9.10)	-1.117 (-9.39)	-1.527 (-10.38)
Men	0.409 (8.19)	1.171 (28.51)	0.024 (1.78)	0.600 (21.01)	0.457 (13.09)

The results for the estimation of the structural equations related to reliability are presented in Table II-3. Based on the results, the perception of reliability tends to decrease with the age (i.e. older travellers will find the alternatives less reliable). The influence of the educational level is not the same in all cases: for the three car-based modes the reliability perception increases with the educational level; for the metro there are no significant differences; and for the bus the reliability perception decreases with the educational level.

Notably, the car-passenger option is always perceived as more reliable than the car-driver option, as all parameters are higher. Also, the most reliable mode is metro, while the bus has the lowest perception of reliability (given the variables' scales, the age parameter dominates the educational variables).

Table II-3: Structural Equations' Results - Reliability

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Age	-0.017 (-38.69)	-0.012 (-26.24)	-0.033 (-70.80)	-0.022 (-28.10)	-0.056 (-93.47)
High School	0.162 (10.17)	0.272 (-4.11)	0.065 (3.61)	1.035 (17.79)	1.196 (15.63)
Technical Studies	0.296 (8.43)	0.421 (-10.27)	0.228 (7.71)	1.048 (20.93)	0.817 (19.71)
College	0.435 (13.43)	0.508 (-11.16)	0.342 (-6.02)	1.028 (30.46)	0.328 (10.45)

Table II-4 presents the results for the estimation of the structural equations related to safety. As with the perception of accessibility/comfort, the perception of safety for all modes is higher for men and it decreases with the age. For the private modes, the perception of safety increases with the income, although the medium and high levels are similar. For the public modes, the effect of the income is negative, with significant difference between medium and high levels for the metro and bus.

The results for the public transport modes show that the perception of safety will always be higher for the metro than for the bus or shared taxi; between these two modes the relationship is not clear and will depend on the particular characteristics of the travellers; the same will happen between the private transport modes.

Table II-4: Structural Equations' Results - Safety

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Age	-0.102 (-92.16)	-0.091 (-88.10)	-0.098 (-78.11)	-0.044 (-74.93)	-0.033 (-74.31)
Medium Income	0.236 (10.61)	0.151 (9.42)	-0.237 (-18.17)	-0.042 (-11.88)	-0.314 (-15.23)
High Income	0.256 (8.85)	0.181 (5.20)	-0.220 (-17.03)	-0.276 (-12.02)	-0.434 (-15.75)
Gender	0.327 (28.55)	0.288 (33.55)	0.127 (39.58)	0.316 (73.00)	0.188 (13.89)

Table II-5 presents the estimates and *t*-values for the measurement equations of accessibility/comfort. As all indicators have the same scale, it can be seen that the most relevant indicator is the ease of access, followed by the comfort during the trip. Based on the results of the measurement equations of reliability, presented in Table II-6, the most relevant indicator is the possibility of calculating the travel time, followed by the possibility of calculating the waiting time. The results for the measurement equations of safety are presented in Table II-7, where it can be seen that for the private transport modes the most relevant indicator is related to accidents, while for the public modes the most relevant indicator is related to theft.

Table II-5: Measurement Equations' Results – Accessibility/Comfort

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Ease of Access	1.679 (98.97)	0.724 (73.46)	1.155 (102.79)	1.564 (82.73)	1.474 (67.87)
Comfort	0.206 (90.08)	0.348 (86.56)	0.747 (105.91)	0.963 (45.23)	1.420 (66.23)
Information	0.225 (69.03)	0.305 (65.18)	0.912 (75.28)	0.117 (27.40)	0.566 (51.13)

Table II-6: Measurement Equations' Results – Reliability

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Wait Time	1.539 (28.63)	1.604 (28.05)	2.309 (19.68)	1.560 (34.43)	2.379 (42.95)
Travel Time	1.627 (39.97)	1.760 (24.04)	1.252 (21.07)	1.966 (49.24)	2.746 (59.17)
Information	0.983 (32.09)	0.662 (18.48)	0.555 (13.01)	1.563 (34.86)	1.240 (29.45)

Table II-7: Measurement Equations' Results – Safety

	Car Driver	Car-Passenger	Shared Taxi	Metro	Bus
Accidents	1.475 (12.61)	1.658 (13.39)	1.593 (16.70)	1.203 (14.29)	1.194 (14.59)
Theft	1.304 (13.68)	1.640 (13.18)	1.722 (15.13)	1.945 (13.81)	1.768 (14.34)

The estimates of the discrete choice model are presented in Table II-8. Two models are presented: a base Multinomial Logit without latent variables, as well as the proposed hybrid Multinomial Logit. For models, the *t*-values of the parameters and the log-likelihoods are presented.

Table II-8: Discrete Choice Model's Results

<b>Parameter</b>	<b>Hybrid Model</b>		<b>Model without Latent Variables</b>	
$\theta_{cost}$	-0.032	(-7.32)	-0.027	(-8.13)
$\theta_{travel}$	-0.006	(-4.67)	-0.033	(-4.82)
$\theta_{men}$	-0.001	(-3.01)	0.030	(2.98)
$\theta_{wait}$	-0.015	(-1.69)	-0.009	(-0.53)
$\theta_{walk}$	-0.022	(-2.89)	-0.016	(-1.80)
$\theta_{trans}$	-1.102	(-8.21)	-1.110	(-8.20)
$\beta_{acc\ com}$	0.622	(3.79)	n. a.	
$\beta_{rel}$	0.441	(2.70)	n. a.	
$\beta_{saf}$	0.613	(1.87)	n. a.	
$ASC_{car-driver}$	0.733	(2.03)	1.220	(5.84)
$ASC_{car-passenger}$	-0.889	(-2.12)	-0.800	(-3.64)
$ASC_{shared\ taxi}$	-1.331	(-1.78)	-1.420	(-4.60)
$ASC_{metro}$	0.247	(0.81)	0.241	(1.56)
$ASC_{car-driver/metro}$	0.223	(0.51)	0.779	(2.65)
$ASC_{car-pass/metro}$	-0.882	(-2.22)	-0.309	(-1.28)
$ASC_{shared\ taxi/metro}$	-0.913	(-1.55)	-0.078	(-0.36)
$ASC_{bus/metro}$	0.342	(1.41)	0.608	(4.59)
$ASC_{shared\ taxi/bus}$	-1.005	(-3.68)	-0.473	(-1.68)
Observations	1,107		1,107	
Log-Likelihood MIMIC Model	-46,784.22		n. a.	
Log-Likelihood Choice Model	-1,099.21		-1,115.94	

All the parameters' signs are consistent with the microeconomic theory: the marginal utilities of the latent variables are positive; on the other hand, the time, cost and transfer represent a disutility for the travellers. Only on the hybrid model, all the parameters are statistically significant at least with a 90% confidence. The model without latent variables presents problems in the estimates of waiting time and travel time for women. Men tend to be more sensitive to the travel time than women.

The traditional model without latent variables has a smaller log-likelihood (and therefore, a worst goodness-of-fit), as it omits significant variables that help explain the decisions of the individuals. It also has alternative specific constants that are more statistically significant; this is expectable, as the model has less explanatory variables and therefore the reminding variables must adjust to explain the missing information.

To further analyse the results of the discrete choice model, Table II-9 presents the subjective values of the different time components (Gaudry *et al.*, 1989) and the monetary valuation of the transfers and the latent variables. It is important to consider that the latent variables do not possess a measurement scale. For all valuations presented, a mean wage rate of US\$5.50 per worked hour (obtained from the socio-economic characteristics of the respondents) was considered.

Table II-9: Attribute Valuations

<b>Parameter</b>	<b>Hybrid Model</b>	<b>Model without Latent Variables</b>
Value of Travel Time Men	1.26 [US\$/hour]	0.47 [US\$/hour]
Value of Travel Time Women	1.08 [US\$/hour]	6.74 [US\$/hour]
Value of Waiting Time	2.67 [US\$/hour]	1.85 [US\$/hour]
Value of Walking Time	3.90 [US\$/hour]	3.31 [US\$/hour]
Value of Transfers	3.23 [US\$/transfer]	3.82 [US\$/transfer]
Value of Accessibility/Comfort	1.82 [US\$/unit]	n. a.
Value of Reliability	1.29 [US\$/unit]	n. a.
Value of Safety	1.80 [US\$/unit]	n. a.

The model without latent variables results on biased estimators of the values of travel time: while for men the value is underestimated, for women it is extremely high. This is a clear sign of the problem of not considering all the relevant factors that travellers take into

account (in this case, the three latent variables). The values of walking time, waiting time and transferring are within a reasonable range when both models are compared. Based on the hybrid model's results, the biggest time disutility comes from walking (due to physical effort), then from waiting (due to uncertainty), and finally from travelling in-vehicle (where men have a higher value of time than women).

## 2.4 Conclusions

Based on the analysis of the application of hybrid discrete choice models to our urban multimode choice case, it is possible to affirm that hybrid models are clearly superior in fit to traditional discrete choice models that do not incorporate latent variables. In fact, differences in magnitude for certain parameters can be observed when latent variables are not considered, producing problems in the valuations obtained.

Our proposed model, estimated with real data from the Santiago Panel, reveals that latent variables, such as Accessibility/Comfort, Safety and Reliability have significant effects over the choice process. The introduction of latent variables allows us not only to improve model fit, but also to achieve better estimated parameters and value functions. Indeed, Table II-9 shows that the model without latent variables clearly underestimates the subjective value of time for men, and overestimates the subjective value of time for women. This can be attributed to the information omission.

Even though the improvement of adding latent variables seems to be clear in terms of estimation, we also identified an important gap in terms of hybrid model usage in forecasting. Traditionally, latent variable models include only socio-economic variables in the MIMIC structure. Consequently, they are not better than traditional discrete choice models in terms of their sensitivity to evaluate policies that modify the transport system. As certain policies could also influence individual perceptions, we recommend testing the possibility of including objective factors in MIMIC structure. Additionally, as the latent

variables do not have a measurement scale, their impact on the utility function (as well as the resulting valuations) is hard to interpret.

### **Acknowledgments**

This study was funded by the Chilean Fund for Scientific and Technological Research, FONDECYT; the Millennium Institute on Complex Engineering Systems; and the Chilean Council for Scientific and Technological Research, CONICYT.

### **Publication History**

The main results of this study were published in 2010 in *Transportation Research Record*, number 2156, under the title “Sequential and simultaneous estimation of hybrid discrete choice models: some new findings”, and in *Transportation Research Part A*, volume 44, under the title “Inclusion of latent variables in Mixed Logit models: modelling and forecasting”. The results covered in this study were presented at the *XIV Chilean Conference on Transportation Engineering* (Concepción, 2009), at the “2009 European Transport Conference” (Leiden, 2009), and at the *89<sup>th</sup> Annual Meeting of the Transportation Research Board* (Washington D.C., 2010).

### **References**

Ashok, K., Dillon, W. and Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, 39(1), 31-46.

Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H. S. Mahmassani (Ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam. 431-470.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. and Feterison, M. (2009). Estimation of discrete choice models: extending BIOGEME. *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland

Bolduc, D. and Alvarez-Daziano R. (2010). On estimation of hybrid choice models. In S. Hess and A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice*. Emerald Group Publishing, Bingley.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, Chichester.

Cherchi, E. and Cirillo, C. (2008). A modal mixed logit choice model on panel data: accounting for systematic and random heterogeneity in preferences and tastes. *86th Annual Meeting of the Transportation Research Board*. Washington D.C.

Gaudry, M.J.I., Jara-Díaz, S.R. and Ortúzar, J. de D. (1989). Value of time sensitivity to model specification. *Transportation Research Part B*, 23(2), 151-158.

Koppelman, F.S. and Hauser, J.R. (1978). Destination choice behaviour for non-grocery-shopping trips. *Transportation Research Record*, 673(1), 157-165.

Koppelman, F.S. and Pas, E.I. (1980). Travel-choice behaviour: models of perceptions, feelings, preference, and choice. *Transportation Research Record*, 765(1), 26-33.

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4), 275-297.

Mobidrive (2000). *Mobidrive questionnaires. Arbeitsberichte Verkehrs-und Raumplanung* 52. Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau, ETH Zürich (in German).

Muñoz, J.C., Ortúzar, J. de D. and Gschwender, A. (2009). Transantiago: the fall and rise of a radical public transport intervention. In W. Saleh and G. Sammer (eds.). *Travel Demand Management and Road User Pricing: Success, Failure and Feasibility*. Ashgate, Farnham.

Nicolaidis, G.C. (1975). Quantification of the comfort variable. *Transportation Research*, 9(1), 55-66.

Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Edition, John Wiley and Sons, Chichester.

Prashker, J.N. (1979). Mode choice models with perceived reliability measures. *Transportation Engineering Journal*, 105(3), 251-262.

Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D. and Ortúzar, J. de D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record*, 2156(1), 131–139.

Raveau, S., Yáñez, M.F. and Ortúzar, J. de D. (2012). Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research Part B*, 46(10), 1374-1383.

Rundmo, T. and Hale, A.R. (2003). Managers' attitudes towards safety and accident prevention. *Safety Science*, 41(7), 557-574.

Stapleton, D.C. (1978). Analyzing political participation data with a MIMIC model. *Sociological Methodology*, 9(1), 52-74.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. 2<sup>nd</sup> Edition, Cambridge University Press, Cambridge.

Vredin Johansson, M., Heldt, T. and Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40(6), 507-52.

Yáñez, M.F., Mansilla, P. and Ortúzar, J. de D. (2010a). The Santiago Panel: measuring the effects of implementing Transantiago. *Transportation*, 37(1), 125-149.

Yáñez, M.F. and Ortúzar, J. de D. (2010). Modelling choice in a changing environment: assessing the shock effects of a new transport system. In S. Hess and A.J. Daly (eds.), *Choice Modelling: the State-of-the-Art and the State-of-Practice*. Emerald, Bingley.

Yáñez, M.F., Raveau, S. and Ortúzar, J. de D. (2010b). Inclusion of latent variables in Mixed Logit models: modelling and forecasting. *Transportation Research Part A*, 44(9), 744-753.

### 3 A TOPOLOGICAL ROUTE CHOICE MODEL FOR METRO

**Sebastián Raveau**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

**Juan Carlos Muñoz**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

**Louis de Grange**

Department of Industrial Engineering  
Universidad Diego Portales

#### **Abstract**

This study presents a route choice model for public transport networks that incorporates variables related to network topology, complementing those found in traditional models based on service levels (travel time, cost, transfers, etc.) and users' socio-economic and demographic characteristics (income level, trip purpose, etc.). The topological variables represent concepts such as the directness of the chosen route and user knowledge of the network. For both of these factors, the necessary data is endogenous to the modelling process and can be quantified without the need for information-gathering beyond what is normally required for building route choice models. Other novel variables in the proposed formulation capture notions of user comfort such as vehicle occupancy rates and certain physical characteristics of network stations. We conclude that these new variables significantly improve the explanatory and predictive ability of existing route choice specifications.

### 3.1 Introduction

The purpose of this study is threefold: to advance our understanding of the behaviour of public transport users when choosing a route in a public transport network, to quantify the impacts of the underlying factors that influence their decisions, and to improve the statistical inference and predictive ability of route choice models.

The route choice factors normally included in these models relate to the service levels of the route alternatives (in-vehicle travel time, waiting time, access time, number of transfers, train passenger density, etc.) and the socio-economic and demographic characteristics of users (income level, purpose of trip, etc.) (Ortúzar and Willumsen, 2011).

However, there are good reasons to suspect the existence of other influences on the user's route choice process that existing models generally ignore. With this possibility in mind, we propose a new model based on three main hypotheses:

- 1) Public transport users tend to penalize routes that deviate from a direct path to the final destination along certain segments. In other words, between two routes with exactly the same trip time, users prefer the most direct one, whether in geographical or topological (schematic map) terms.
- 2) Public transport users tend to prefer routes that are better known or more heavily travelled.
- 3) Public transport users consider different factors when choosing travel routes, besides traditional variables such as travel time or fare. Some of these factors may include comfort, reliability or physical characteristics from the vehicles and stations.

To test the first hypothesis our model defines and incorporates a novel explanatory variable called angular cost to represent the underlying effects of route segment deviation on user route choice in a public transport network. The function expressing this variable is such that the more indirect a given route is, the higher will be the value of its angular cost variable. If the parameter of this variable is found to be statistically significant, the first hypothesis is confirmed. Various definitions and specifications of route directness and angular cost penalties may be found in Montello (1991), Tversky (1992), Turner (2001), and Conroy-Dalton (2003).

The angular cost of a route is obtained directly from a map of the public transport network. Here, however, an important issue arises, for the network maps published and displayed by public transport systems typically incorporate design considerations that may distort the true geographical coordinates or spatial relationships of the network's stations. This implies that different map designs will generate different angular cost values and as a result may vary in their effects on user behaviour. By including the angular cost variable we are therefore also attempting to explain how travel decisions are affected by the way alternative routes are visually presented to the user.

To test the second hypothesis we need a variable that indicates how knowledgeable users are of a given route. Since this factor cannot be determined exactly, we will employ a variable measuring the use levels of the different routes as a proxy. Thus, our model is an endogenous formulation in which the probability of a route choice depends in part on how heavily used its constituent segments are.

To test the third hypothesis our model includes different variables related with level of service that are not usually included in route choice modelling. These variables are related to comfort on the vehicles, facilities on the stations such as escalators and level of usage, among others. The inclusion of this kind of variables allows the analysis of how the environment's characteristics impact the way alternative routes are perceived and, ultimately, the choices.

In the literature there are multiple route choice models based on users' socio-economic and demographic characteristics and their perceptions of route attributes (Daganzo and Sheffi, 1977; Ramming, 2001; Prashker and Bekhor, 2004) such that user behaviour is determined by perceived costs. However, the attributes included in these models are all tangible and easily justified as significant factors in a rational individual decision-making process. Furthermore, the data on user characteristics are normally gathered through user surveys or through network measurements so that the models are completely exogenous. The result is that the choices they generate display considerable variability, a reflection of the inadequacy of the data incorporated by the modeller on the factors affecting users' route decisions. Needless to say, any additional data that could help explain this variability would be very welcome.

A possible source of such data is suggested by the well-established fact that public transport user decision-making is affected by psychological considerations such as aesthetics, comfort and travel-time reliability (see Papinski *et al.*, 2009). However, there are major difficulties inherent in integrating these factors into route choice modelling that stem from (i) their subjectivity, given that each user perceives them differently; and (ii) their intangibility, since there is no scale for measuring them. Although some progress has been made in accommodating these phenomena in mode choice models (Ben-Akiva *et al.*, 2002; Raveau *et al.*, 2010) their applications in route choice contexts remain limited (Prato *et al.*, 2012).

In the present study we aim to improve the explanatory and predictive abilities of route choice models by incorporating variables that capture angular cost and route knowledge level. Our proposed model is a Multinomial Logit formulation that explicitly adds these novel explanatory variables to the ones traditionally found in existing designs. It is calibrated with data from the Santiago Metro system and will be compared to a base model that does not include angular cost, network knowledge or variables related to intangible factors of the level of service. Our conclusions will be that the added variables significantly improve the ability to explain and predict route choice processes in public transport networks.

### 3.2 Specification of the Model

The Santiago Metro system has a number of origin-destination pairs that can be travelled by more than one feasible route, as is evident from the map of the network (see Figure 3-1). To test our route choice model, therefore, we empirically analyse how system users travelling these pairs choose the transfer point or points between their origin and destination stations. The analysis is based on trips made during the morning (7am-9am) and evening (6pm-8pm) peak hours. During these periods users take approximately 700,000 trips, 44% of which include transfers.

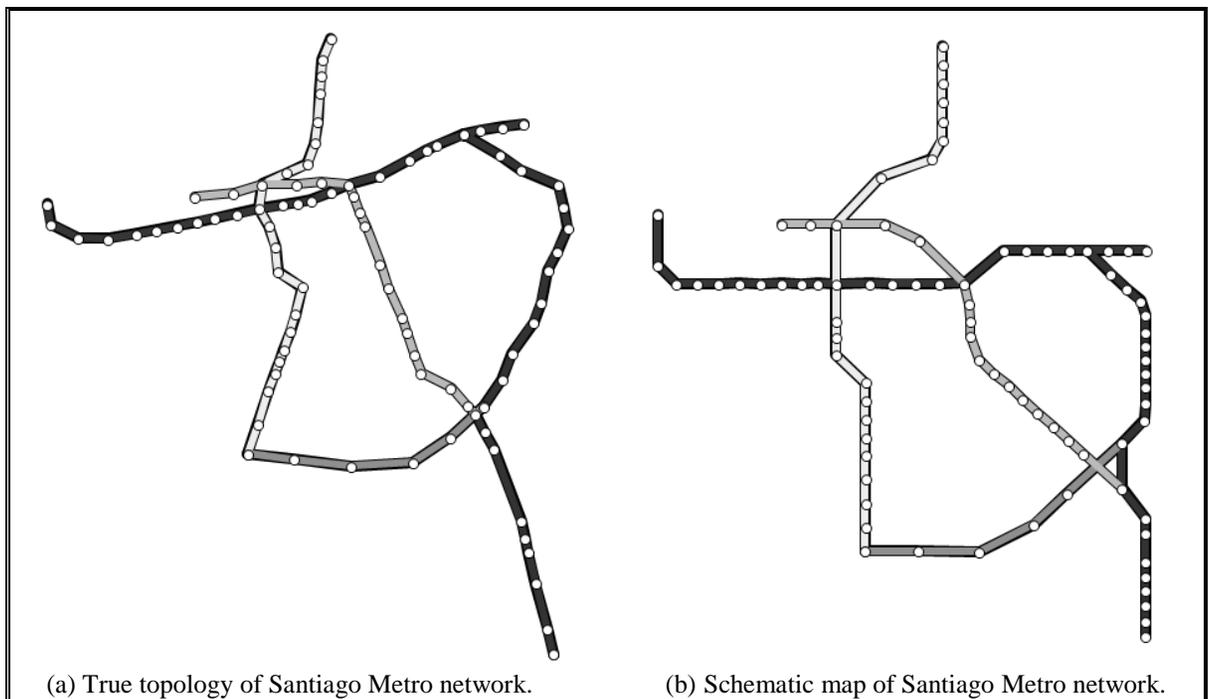


Figure 3-1: Santiago Metro Network Maps

Trip data were obtained from an origin-destination survey conducted at Metro stations in October 2008. Information was gathered on the peak-hour trips of 92,800 individuals, or about 12% of all users. Only those whose origin-destination pairs could have been taken by more than one route are included in our analysis. There were 16,029 such cases, amounting to about 40% of the individuals surveyed who transferred at some point on their trip.

Two different depictions of the Metro network are given in Figure 3-1. The design on the left shows the true topology of the various stations and lines (geo-referenced data) while the one on the right, which is the map actually displayed in the stations and consulted by users, contains serious distortions both in the actual location of the stations and the distances between them. This means that users' angular and geometric perceptions also diverge from the real ones, which may induce them to choose routes that are not the optimal or lowest cost ones. Though the extent of this distortion has increased with the considerable growth of the Santiago system over the last decade, it is still a long way from the complexity and high density of the networks in the major cities of Europe and the United States where the angular and geometric distortions of the schematic maps are much more significant. The scale of the potential impact in these systems lends particular significance to the study of how graphical presentations of network information effect user behaviour.

In 2008, when the survey was conducted, the Santiago Metro network had 5 lines and 85 stations, 7 of which were transfer points. Of the 7,140 possible origin-destination station pairs, 4,985 (70%) required transferring between lines. For 1,365 of these there is more than one route that is "reasonable," by which we mean a route that was taken by at least one user in our survey database. The origin-destination pairs with more than one reasonable route are broken down in Table III-1 by the number of such routes. Although in the majority of cases the route choice is between just two reasonable alternatives, there are also some pairs for which three and even four different options were chosen by some of the individuals surveyed. In denser transport networks, both the quantity and percentage of origin-destination pairs with alternative route possibilities would no doubt rise.

Table III-1: O-D Pairs with More than One Reasonable Route Alternative

<b>Observed route alternatives</b>	<b>% of O-D pairs</b>	<b>% of trips</b>
2	97 %	93 %
3	3 %	7 %
4	< 1 %	< 1 %

### 3.2.1 General Description of the Model

As just noted, our route choice model is a Multinomial Logit design in which the user utility function contains the standard explanatory variables representing network service levels plus additional ones for angular cost and route knowledge levels. Since the socio-economic and socio-demographic variables usually included are not relevant for the purposes of this study, they have been left out of our specification. Nor have we incorporated ticket prices since the Santiago Metro uses a flat fare system. The variables are discussed in detail below.

The base model used for comparison purposes is a restricted or nested formulation containing only traditional service level variables (that is, without angular cost, route knowledge and variables related to intangible factors of the level of service). The estimation of the parameters is performed using the maximum likelihood method.

### 3.2.2 Explanatory Variables

The standard explanatory variables in our model are **in-vehicle travel time** from origin to destination, total **waiting time** at origin and transfer stations, and the **number of transfers** for each route alternative as an indicator of additional (walking) time between platforms and the disutility of changing train lines. No fare variable is included since under the Metro's flat-fare policy, user monetary cost is the same regardless of route length or transfers and thus has no effect on route choice.

To strengthen the explanation of line changes provided by the number of transfers reported, three explanatory variables were added to explicitly represent line-change characteristics: **walking time** between different train lines, whether or not platforms are accessed by **escalators**, and whether or not a given transfer involves an **ascending level change**. Data for these variables were collected via a field survey of transfer stations.

Also, since detailed information was available on the load profiles of each Metro line for different times of day, we were able to include an explanatory variable indicating the **average occupancy rate** for the different routes. This variable represents the degree of passenger crowding on train carriages, and indirectly also captures train capacity restrictions and the probability a user is unable to board the first-arriving train due to overcrowding and has to await the next one (a frequent occurrence on the Santiago Metro at peak hours). The variable itself is the average occupancy level (i.e., the ratio of load to capacity) of the arcs constituting the alternative routes, and is weighted by the route distances. By definition the rate can vary between 0 (for a train travelling empty along all arcs of a route) and 1 (for a train travelling fully loaded along all arcs).

Two more variables reflecting factors related to extreme occupancy rates at origin and transfer stations are also integrated into the model. If the rate is greater than or equal to 85%, not only is comfort reduced due to the high passenger density but there is a chance users will be **unable to board** the first train. If, on the other hand, the rate is less than or equal to 15%, chances are that users can **ride seated**. Thus, for the stations where passengers board, the two percentages are the occupancy rate thresholds at which user satisfaction changes. These particular values were chosen based on two criteria: (i) they reflect reasonable ranges given the various factors involved; (ii) they provide the best fit of the model to the data. Note also that the “unable to board” and “ride seated” variables are both dummies.

As regards network or route knowledge levels, since the Santiago Metro was developed in various stages over the last few decades (the oldest line being inaugurated in 1975 while the newest one opened in 2006), familiarity with the different route possibilities varies considerably and users tend to prefer the ones they know best. In addition, the more a given line is used the more likely it is to be better known than the others. To incorporate this effect in the proposed model we defined a **network knowledge** proxy variable that is simply the average passenger volume on each route during peak hours.

As regards the network topology factor, we assume users prefer the most direct routes from origin to destination. To reflect this preference, and to facilitate our evaluation of the effects of network map distortions of true route geography on route choice, we define the **angular cost** variable for each route as a penalty indicator that must satisfy the following three characteristics:

- 1) The penalty is at a minimum for route segments that head directly toward the final trip destination (i.e., the angle with respect to the destination is  $0^\circ$ ), and at a maximum for segments heading in the opposite direction (angle =  $180^\circ$ ).
- 2) Marginal changes at the extremes ( $0^\circ$  and  $180^\circ$ ) generate small penalty variations whereas a large variation is produced at the point where a route segment turns toward or away from the destination (this occurs when the angle is  $90^\circ$ ).
- 3) The penalty is angularly symmetric in the sense that if, for example, the angle of a segment is  $15^\circ$ , the penalty is the same as for an angle of  $345^\circ$ .

A penalty function that has all three of these characteristics is  $\sin(\theta/2)$ , where  $\theta$  is the angle formed by a straight line from the initial point of a segment to the final trip destination and another straight line from the initial point of the segment itself. An example of the estimation of angular cost is given in Figure 3-2. The penalty for a segment is weighted by its length to reflect how far the deviation distances users from their destination. In the case of the two Santiago Metro network maps, note that the angular costs for the routes as depicted in the true topology version (Figure 3-1a) will be different than those for the Metro schematic map (Figure 3-1b). This factor can therefore be tested separately for the two designs.

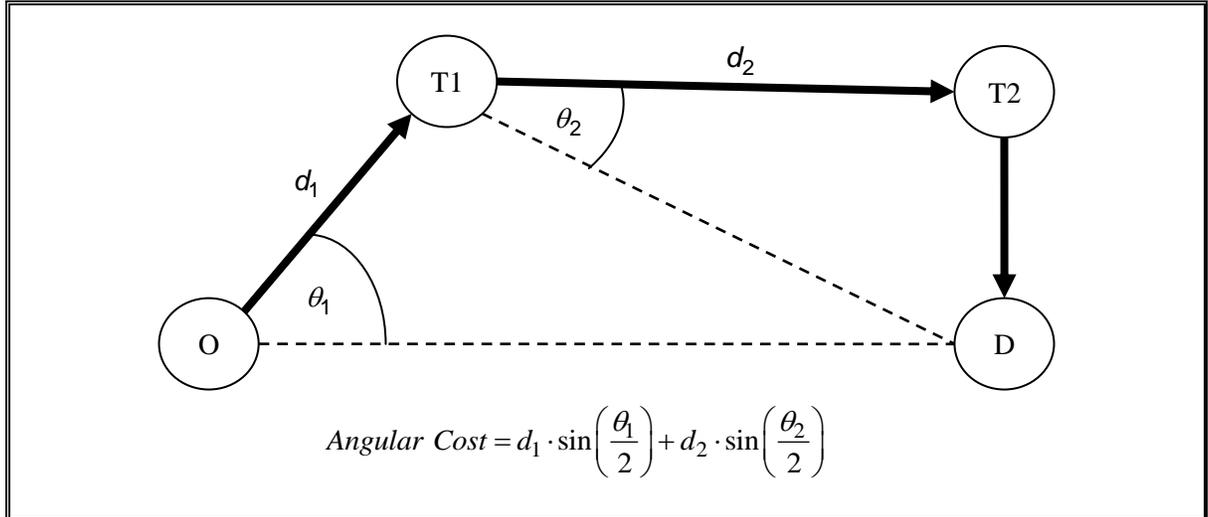


Figure 3-2: Angular Cost Definition

One of the characteristics of our angular cost specification is that it is not topologically symmetric. In other words, the angular cost of travelling from O to D is not the same as for travelling from D to O. This is consistent with the structure of route choices observed in the Santiago Metro. A dispersion graph of the proportions of route choices in both directions for all O-D pairs with at least 50 observed trips in our survey sample is set out in Figure 3-3. For each route between two stations that has more than one possible route, the graph shows the proportion of outbound (O-D) trips versus that of inbound (D-O) trips. If user route choices were symmetrical, the points on the graph would tend to be concentrated around the 45° diagonal. But as can be seen, this is not the case. The absence of topological symmetry is thus consistent with actual user route choices and is therefore not a shortcoming of the proposed specification.

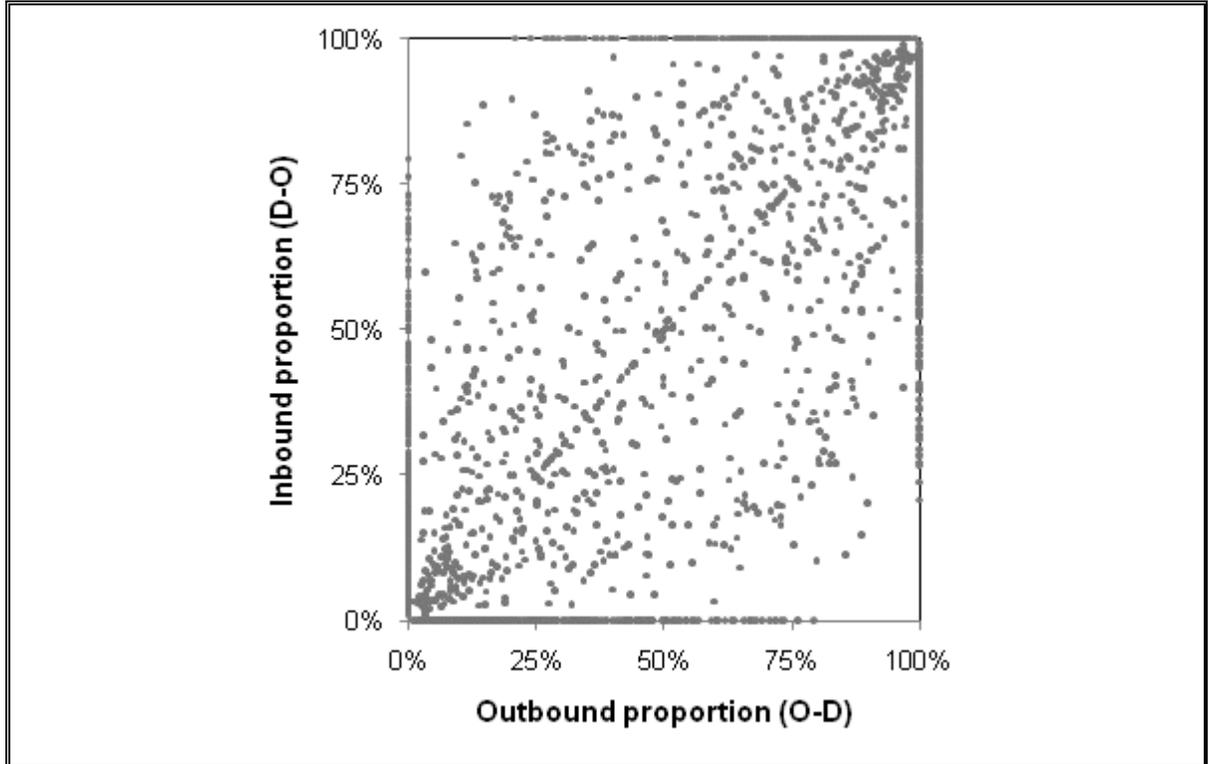


Figure 3-3: Asymmetry of Users' Route Choices

Finally, and also in relation to network topology and the angular cost concept, we incorporated two dummy variables first defined by Dial (1971) that capture another relevant aspect of route geometry. The first variable identifies routes with a transfer station (if any, otherwise the destination) that is closer to the origin than the station immediately before it, creating the impression of **turning back to the origin**, while the second one identifies routes that have a transfer station further from the destination than the station immediately before it, giving the impression of **turning away from the destination**.

### 3.3 Results and Analysis

Based on the data gathered on route choice in the Santiago Metro network we estimated both our proposed Multinomial Logit model, built around the complete set of explanatory variables just described, and a base model (which only considers the standard explanatory variables) for comparison purposes. The sample consisted in route choices of 28,961 users

on peak hour, over 1,365 different OD pairs. The number of route alternatives varied from 2 to 4, as presented in Table III-1. The results of the estimation are summarized in Table II-2. Note that two versions of the proposed model were solved, one based on angular cost for a true topological map of the Metro and the other for the distorted schematic map.

Table III-2: Estimates for Route Choice Models

<b>Variable</b>	<b>Base model</b>	<b>Proposed Model (true distances)</b>	<b>Proposed Model (map distances)</b>
In-vehicle travel time	- 0.144 ( - 46.61 )	- 0.119 ( - 27.94 )	- 0.117 ( - 27.29 )
Waiting time	- 0.203 ( - 3.68 )	- 0.111 ( - 3.17 )	- 0.121 ( - 3.17 )
Walking time	-	- 0.240 ( - 7.14 )	- 0.229 ( - 6.87 )
Number of transfers	- 1.220 ( - 17.06 )	- 0.449 ( - 3.95 )	- 0.420 ( - 3.69 )
Ascending level change	-	- 0.426 ( - 7.89 )	- 0.432 ( - 8.02 )
No escalators	-	- 0.438 ( - 11.95 )	- 0.457 ( - 12.34 )
Occupancy rate	-	- 1.430 ( - 3.53 )	- 1.250 ( - 3.05 )
Unable to board	-	- 0.426 ( - 6.00 )	- 0.413 ( - 6.06 )
Ride seated	-	0.126 ( 2.16 )	0.106 ( 1.81 )
Network knowledge	-	0.031 ( 7.14 )	0.030 ( 6.81 )
Angular cost	-	- 0.024 ( - 7.14 )	- 0.038 ( - 6.75 )
Turn back to origin	-	- 0.516 ( - 10.96 )	- 0.503 ( - 10.65 )
Turn away from destination	-	- 0.505 ( - 11.43 )	- 0.512 ( - 11.59 )
Sample size	28,961	28,961	28,961
Log-likelihood	- 7,416	- 7,142	- 7,136

As may be observed, all of the parameter estimates for the proposed models have the right sign and are statistically significant at the 95% level. Of particular interest is the result for the network knowledge variable, measured by a route's average passenger volume, whose effect is positive. This being so we could easily specify an endogenous or fixed-point route choice model in which the probability of choosing a particular route depends on the number of users choosing it. Such a formulation would be needed for use in making predictions.

Regarding angular cost and the variables on turning back to the origin or away from the destination, the results are more statistically significant when the proposed model is constructed using the distorted schematic map distances (Figure 3-1b) as opposed to the true topological distances (Figure 3-1a). This strongly indicates the significance of the way in which route information is presented to users. To the extent the schematic map distortions induce users to choose certain routes over others, this phenomenon could be exploited to modify the use of certain network lines or route segments.

Another point demonstrated by these results is that the traditional explanatory variables travel time, waiting time and number of transfers, although certainly important (travel time is the most statistically significant), do not fully explain the decision-making process. Indeed, the absence of the information represented by the proposed new variables may bias the parameter estimates of the traditional ones (note the large difference in the waiting time and number of transfers parameters).

Also confirmed by the results in Table III-2 are the two hypotheses stated in the Introduction: users perceive an additional cost for routes that have segments which turn away from the destination, and tend to choose the routes they are most familiar with or are heavily used by other users.

In light of the foregoing, the inclusion of non-traditional variables is clearly needed. A particular advantage of adding factors relating to network topology such as those suggested

here is that the necessary information is readily obtained at low cost as part of the usual data-gathering process for building route choice models.

As regards goodness-of-fit, the proposed model, with 10 more explanatory variables than the base model, delivers a significant improvement over the latter of more than 200 points in log-likelihood. This leads us to reject the null hypothesis that the two specifications are equivalent ( $2 \times 280 = 560 > \chi^2 = 23.2$  at 99% confidence with 10 degrees of freedom). The mean square error (see Table III-3) indicates that the proposed model is also superior to the base model in terms of predictive ability. The prediction error for both models increases as the number of alternative routes grows, but so do the differences between them.

Table III-3: Mean Square Error for Route Choice Models

<b>Origin-destination pairs</b>	<b>Base model</b>	<b>Proposed model</b>
<b>All</b>	13.3	10.9
<b>With 2 routes</b>	7.9	6.7
<b>With 3 routes</b>	20.0	10.7
<b>With 4 routes</b>	27.8	15.9

### 3.3.1 Perceptions and Valuations

Although our model does not produce monetary valuations for the various attributes (e.g., the subjective value of time) given that no cost variable was included due to the Metro's flat-fare system, marginal rates of substitution between the various time factors can be derived using the in-vehicle travel time as the baseline. The values so obtained are presented in Table III-4.

Table III-4: Trip Assignment for Model Application

<b>Variable</b>	<b>Base model</b>	<b>Proposed model</b>
<b>In-vehicle travel time</b>	1.00 (base)	1.00 (base)
<b>Waiting time</b>	1.41	1.07
<b>Walking time</b>	-	1.79

Waiting time in the base model is 41% more highly valued than in-vehicle travel time whereas in the proposed model, the two variables are similarly valued statistically. In the case of walking time, not included in the base model, the proposed model values it 79% more than in-vehicle travel time. Also notable is that the waiting time value from the base model is close to the average of the walking time and waiting time values from the proposed model. The explanation for this last result may be that in the base model the waiting time value captures part of the walking time effect, reflecting the fact that the two variables are correlated since they are both directly related to, and simultaneously increased by, transferring. As a result, the base model parameters are biased due to the absence of significant variables that are incorporated in the proposed model.

As for transfers, the base model can only generate a single value of 8.5 travel minutes per transfer. In the proposed model, however, the components associated with this factor are disaggregated and values vary depending on the transfer station. The valuation for walking time between line platforms was given in Table III-4 while the valuations for the specific station characteristics are indicated in Table III-5. The valuation the different transfer cases is obtained by considering all the parameters related to each scenario (i.e. the base parameter associated to transfer and the marginal parameters associated to the possibilities of boarding or getting a seat, the possibility of using escalators and the possibility of descend/ascend when transferring). These figures can be interpreted as the additional travel time users will accept to avoid a transfer. Due to its omission of significant information the base model delivers an average transfer value that does not resemble any of the 12 possible cases valued by the proposed model.

Table III-5: Transfer Valuations by Station Characteristics

Transfer Characteristic		Without escalators	With escalators
<b>Chance of riding seated (occupancy <math>\leq</math> 15%)</b>	<b>Descending</b>	2.7 min. of travel	6.6 min. of travel
	<b>Ascending</b>	6.4 min. of travel	10.3 min. of travel
<b>Intermediate (15% &lt; occupancy &lt; 85%)</b>	<b>Descending</b>	3.6 min. of travel	7.5 min. of travel
	<b>Ascending</b>	7.3 min. of travel	11.2 min. of travel
<b>Chance of being unable to board (occupancy <math>\geq</math> 85%)</b>	<b>Descending</b>	7.9 min. of travel	11.2 min. of travel
	<b>Ascending</b>	11.0 min. of travel	14.9 min. of travel

Finally, many public transport systems experience serious problems of congestion during peak periods and the Santiago Metro is no exception, exhibiting high load factors during the hours of heaviest use over much of the network. As has been demonstrated, the way in which map-based network information is presented to users can induce them to change their route choice behaviour. Such changes can be positive, tending to decongest overloaded lines, or negative, overloading them even further. The design of graphic information displays is thus an important issue that must be handled carefully in order to ensure the effects on user behaviour are the desirable ones.

### 3.3.2 Forecasting Application

To quantify the improvement in predictive ability brought about by the new variables in the proposed model, we analysed the route choices between a group of 11 stations on a single line of the Santiago Metro and another group of 9 stations on a different line (this is, a set of 198 O-D pairs). According to our data, trips taken between the two groups during peak hours numbered some 7,300 and followed 4 different route patterns (see Figure 3-4).

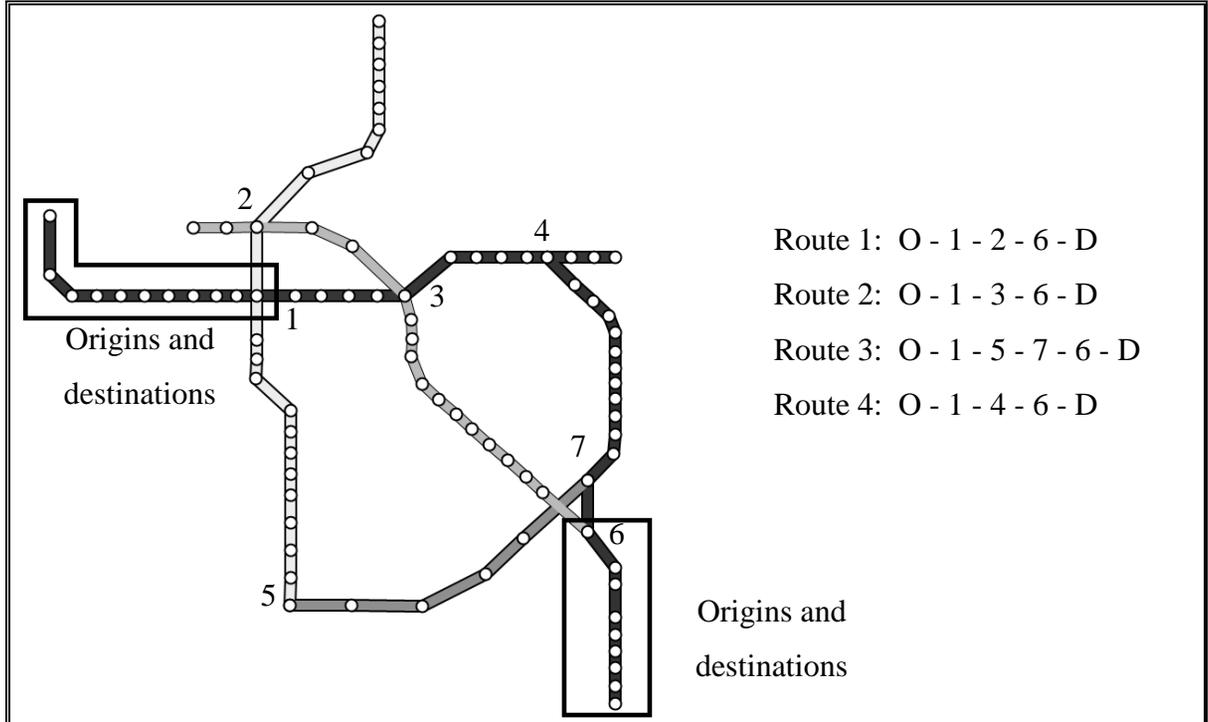


Figure 3-4: Network for Model Application

This example is particularly suitable for our purposes because the 4 routes use all 7 transfer stations in the Metro network and can be clearly described in topological terms, which are given below:

- 1) Route 2 is the most direct route and has no segment that turns back to the origin or away from the destination.
- 2) Route 1 has the highest angular cost on certain segments, especially between transfer stations 1 and 2. However, the segment between transfer station 2 and the destination is relatively direct.

- 3) Routes 3 and 4 are less direct and require transferring at stations far from the destination. However, they have lower occupancy rates, particularly Route 3, and Route 4 is the fastest one to arrive at the destination line (but not necessarily the fastest one to arrive at the destination station).
- 4) Routes 1 and 2 both pass through transfer stations 1 and 3. On Route 2 the section between those two stations is direct but the occupancy rate is the highest of any part of the network. On Route 1 that leg is made through transfer station 2, but the occupancy rate is lower and there are fewer intermediate stations (i.e., lower travel time).

The trip assignment results for our application example are summarized in Table III-6, showing (i) the observed route choices, (ii) the route choices predicted by the base model, (iii) the route choices predicted by the proposed model, and (iv) the minimum-time assignments (an all-or-nothing loading in terms of travel, waiting and walking times). Note that the minimum-time assignments do not assign the entire flow to a single route. This occurs because train frequencies during the peak periods for the different routes vary and no single route is always the fastest for all of the O-D pairs.

Table III-6: Trip Assignment for Model Application

<b>Assignment</b>	<b>Route 1</b>	<b>Route 2</b>	<b>Route 3</b>	<b>Route 4</b>
<b>Observed trips</b>	99	5,362	30	1,854
<b>Minimum time</b>	668	6,676	0	0
<b>Base model</b>	361	3,884	33	3,067
<b>Proposed model</b>	270	4,446	25	2,603

As can be seen in Table III-6, Route 2, the most direct route, is generally the fastest and therefore the most frequently used. Nevertheless, the trips assigned to Route 2 in the proposed model are fewer than those assigned considering minimum time due to a high occupancy rate, which is perceived as a negative attribute. Route 1, though faster than Route 4, is used considerably less frequently, because the segment between transfer

stations 1 and 2 has a high angular penalty and turns away from the destination. The proposed model generates an assignment that is closer to the observed trips than that of the base model, thus demonstrating the proposed model's superior predictive ability.

### 3.3.3 Policy Applications

The results of the study have been used by the Santiago Metro authorities to evaluate and change their design of the schematic map, in order to avoid inducing the travellers to make suboptimal decisions. In particular, the biggest distortion of the Santiago Metro map was in its central part (see Figure 3-1), which is the most congested part of the network. The new map used now by Metro reduced that distortion, as can be seen in Figure 3-5. Although the change may seem minor (after all, the general distortion of the map is small), it may have local impacts on the most congested part of the network, and it is a good example of the real impacts and potential that behavioural modelling can have on public policies.

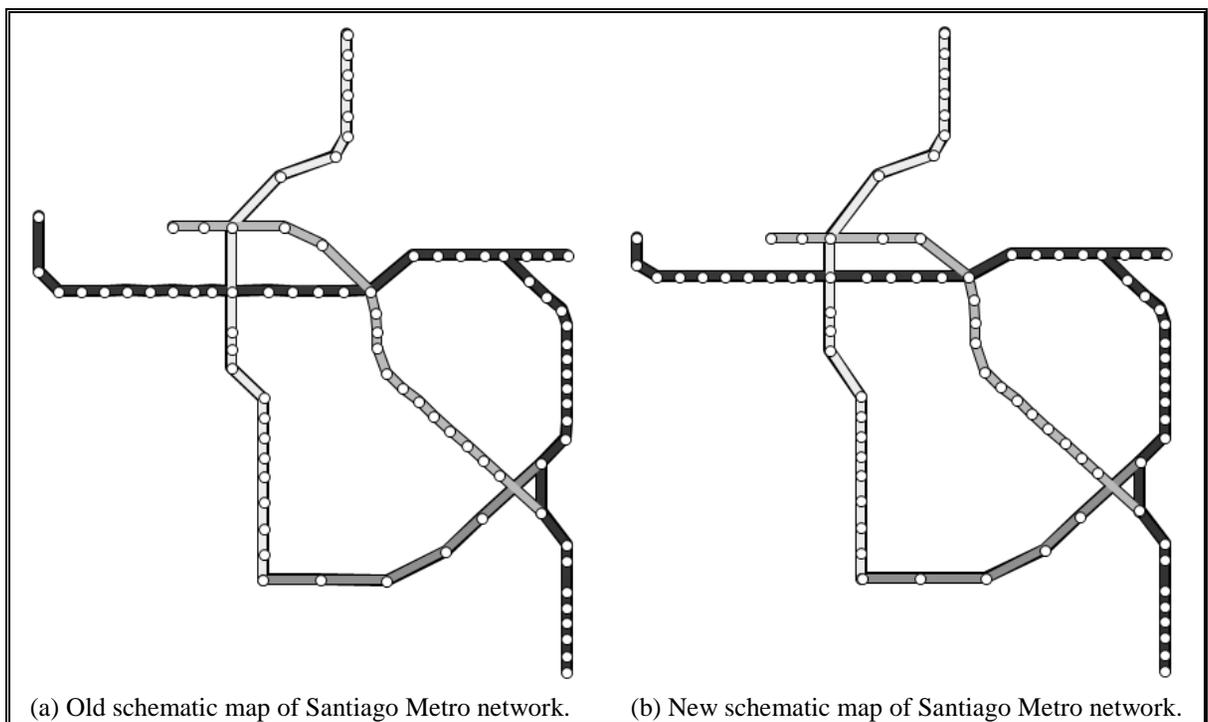


Figure 3-5: Change in the Santiago Metro Network Map

The modelling approach presented in this study has also been used by the Santiago Metro authorities to forecast and evaluate the extended network for 2018, where two additional lines will be in operation. Different aspects have been analysed, based on the forecast of the route choices for the future network: (i) load profiles of every line have been obtained, identifying potential capacity problems, (ii) transfers in every station (by direction) have been predicted, to study potential improvements in terms of accessibility and mobility, (iii) the average platform interactions (i.e. number of boarding and alighting passengers coinciding) have been studied, to guarantee comfort and safety, and (iv) potential interventions to the network (such as restricting movements, promoting transfers and lines, or implementing a fare scheme within the network) have been evaluated. Additionally, behavioural aspects from this study have been used to design the future schematic map, with the two additional lines.

### **3.4 Conclusions**

Route choice modelling is traditionally based only on tangible explanatory variables. However, the perceptions of transport users regarding available route alternatives are such that they do not always choose what the modeller would consider as the “lowest cost” option. In this study we have demonstrated and quantified the influence of non-traditional factors that also impact individual user decision-making. Specifying certain aspects of the trip environment (e.g., passenger densities and the physical characteristics of Metro stations) is also shown to improve the explanation of route choices.

Among the non-traditional variables incorporated in our proposed model are a number that relate to the topology of the public transport network, representing the angular cost (i.e., directness) of the various route alternatives and their geometry (whether a segment turns back to the origin or away from the destination). One of the main advantages of adding these factors is the low cost of collecting the corresponding data. Other novel variables included in our specification express user knowledge of the network’s route alternatives.

The results generated by our model lead us to conclude that when the non-traditional variables are absent the estimated parameters are biased. This is manifested in inaccurate marginal rates of substitution and route choice models with insufficient explanatory and predictive abilities. It is therefore highly recommendable that field surveys be undertaken of networks' physical characteristics (e.g., certain aspects of Metro stations) and operating features (e.g., user information) in order to obtain the data necessary for specifying these proposed new variables in route choice formulations that will produce more accurate estimators and better predictions.

Finally, we have shown that the way in which network information aimed at users of a public transport system is presented can influence their decision-making. Distortions in the representation of a network's geographical relationships can induce trip assignments that reduce service levels for all users. At the same time, these distortions can be exploited to induce optimal trip assignments and thereby making the best use of transport system capacity.

In judging our findings it should also be recalled that the Santiago Metro is a small network with relatively few route alternatives for any given origin-destination pair compared to the systems of major cities such as London, New York or Madrid. The results in Table III-3 show that as the number of alternative routes grows, the proposed model improves significantly compared to a standard route choice model without the additional variables introduced in this study. The superior predictive ability of our formulation should therefore be even more apparent with denser networks. The distortion of the true spatial relationships by the Santiago network's schematic map, though by no means negligible, is not overly significant. For large and highly complex networks with many route alternatives, however, the distortions in these schematic maps would take on considerable importance. Our findings could assist in orienting the design considerations for the maps these systems publish in order to induce socially optimal behaviour in their users.

A natural extension of the study presented here would be to use the proposed model for making predictions. This would require solving the fixed-point problem given that the

variables relating to network knowledge and occupancy rates are endogenous to the modelling process.

### **Acknowledgments**

This research was supported by the Across Latitudes and Cultures – Bus Rapid Transit Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF), Project ACT-32 Anillos Tecnológicos of Conicyt and Fondecyt Iniciación 11100024.

### **Publication History**

The basis of this article was published in 2011 in *Transportation Research Part A*, volume 45, under the title “A topological route choice model for metro”. The results covered in this article were presented at the *XVI Panamerican Conference on Transportation Engineering, Traffic and Logistics* (Lisbon, 2010), at the *19<sup>th</sup> Triennial Conference of IFORS* (Melbourne, 2011), and at the *XV Chilean Conference on Transportation Engineering* (Santiago, 2011).

### **References**

Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H. S. Mahmassani (Ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam. 431-470.

Conroy-Dalton, R. (2003). The secret is to follow your nose: route path selection and angularity. *Environment and Behavior*, 35(1), 107–131.

Dial, R.B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5, 83-113.

- Montello, D.R. (1991). Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior*, 23(1), 47–69.
- Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Edition, John Wiley and Sons, Chichester.
- Papinski, D., Scott, D.M. and Goherty, 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*, 12(4), 347-358.
- Prato, C.G., Bekhor, S. and Pronello, C. (2012). Latent variables and route choice behaviour. *Transportation*, 39(2), 299-319.
- Prashker J.N. and Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, 24(4), 437-463.
- Ramming, M.S. (2001). *Network Knowledge and Route Choice*. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D. and Ortúzar, J. de D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record*, 2156(1), 131–139.
- Turner, A. (2001). Angular analysis. *Proceedings of the 3rd International Symposium on Space Syntax*, Georgia Institute of Technology, Atlanta, 30.1–30.11.
- Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23(2), 131–138.

## 4 ROUTE CHOICE MODELLING ON METRO NETWORKS

**Sebastián Raveau**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

**Zhan Guo**

Robert F. Wagner Graduate School of Public Service,  
Rudin Center for Transportation Policy and Management  
New York University

**Juan Carlos Muñoz**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

**Nigel H.M. Wilson**

Department of Civil and Environmental Engineering  
Massachusetts Institute of Technology

### **Abstract**

Understanding travellers' behaviour is key element in transportation planning. This article presents a route choice model for metro networks that considers different time components as well as variables related to the transferring experience, train crowding, network topology and socio-demographic characteristics. The route choice model is applied to the London Underground and Santiago Metro networks, to make a comparison of the decision-making process of the users on both cities. As all the variables are statistically significant, it is possible to affirm that public transport users take into account a wide variety of elements when choosing routes. While in London the travellers prefer to spend time walking, in Santiago is preferable to spend time waiting. Santiago Metro users are more willing to travel in crowded trains than London Underground users. Both user groups have a similar dispreference to transfers after controlling for the time spent on transfer, but different attitudes to ascending and descending transfers. Topological factors presented on a distorted Metro map are more important than actual topology to passengers' route choice decisions.

## 4.1 Introduction

Understanding how public transport users make their travel decisions and being able to predict their behaviour is essential in transportation planning. The purpose of this study is to advance our understanding of the behaviour of public transport users when choosing a route in a Metro network and to quantify the impacts of the underlying explanatory variables that influence their decisions.

Route choice models have been explored and developed for private transport networks (Bovy and Stern, 1990; Ramming, 2001; Prato, 2009), but not much work has been done in public transport networks (Hunt, 1990; Bovy and Hoogendoorn-Lanser, 2005). The route choice variables normally included in traditional route choice models (either on private transport or public transport networks) limit to some basic service level attributes of the alternative routes, such as travel time and fare (Ortúzar and Willumsen, 2011). However, other variables, related to both the level of service and the travellers' perceptions, influence the user's route choice process but are generally ignored in traditional modelling. This study presents a route choice analysis on Metro networks, incorporating variables related to the different times involved (travel, waiting and walking times), trains and stations usage, transfer environment, network topology and socio-demographic information from the travellers

This study also conducts an empirical analysis to compare route choice decision-making on the London Underground system and the Santiago Metro system, using the same modelling approach and specification. Even though behavioural comparisons can be made between studies found on the literature (mainly between results such as values of time and demand elasticities), these can generally be made based on models with different specification and context. One of the main objectives of this study is to develop a common framework to analyse and compare the preferences of travellers from both networks, and provide general transportation planning insights from the comparison.

The rest of the study is organized as follows. In Section 4.2 we address the modelling approach and present the modelling variables considered. In Section 4.3 we present and discuss the route choice results for the London Underground and Santiago Metro networks. Finally, in Section 4.4 we present our main conclusions.

## 4.2 Route Choice Modelling

In this study we seek to identify and quantify the different aspects of travelling that are taken into account by public transport users from two cities (London and Santiago), particularly when choosing their travel routes. For this, it is essential to understand and model their decision-making process, based on the different characteristics of the alternative routes.

### 4.2.1 Discrete Choice Models and Correlation

We consider a random utility model, where it is assumed that each traveller  $q$  chooses a route  $i$  among a set  $A(q)$  of available alternatives in order to obtain the maximum possible utility level  $U_{iq}$ . It is also assumed that the modeller, who is just an observer without perfect information regarding the decision-making process, is only able to define a representative utility level  $V_{iq}$ . Thus, it is necessary to associate an error term  $\varepsilon_{iq}$  to each alternative (McFadden, 1974), typically as shown in (4.1).

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{4.1}$$

The representative utility level  $V_{iq}$  is a function of different attributes  $X_{ikq}$  related to the routes and the travellers (e.g. travel times, transfer characteristics and travellers' perceptions). Generally,  $V_{iq}$  is assumed to be a linear function of the attributes, as shown in (4.2), where  $\theta_{ik}$  are parameters to be estimated.

$$V_{iq} = \sum_k \theta_{ik} \cdot X_{ikq} \quad (4.2)$$

To characterize the individual decisions, binary variables  $d_{iq}$  that take values according to (4.3) are also needed. These binary variables correspond to the actual decisions made by the travellers.

$$d_{iq} = \begin{cases} 1 & \text{if } U_{iq} \geq U_{jq}, \quad \forall j \in A(q) \\ 0 & \text{in other case} \end{cases} \quad (4.3)$$

If the random terms  $\varepsilon_{iq}$  are assumed to be i.i.d. Gumbel, a Multinomial Logit (MNL) model is obtained, from which is possible to obtain an analytical expression for the choice probabilities  $P_{iq}$ , given by (4.4).

$$P_{iq} = \frac{\exp(V_{iq})}{\sum_{j \in A(q)} \exp(V_{jq})} \quad (4.4)$$

One of the limitations of the MNL model is that it does not consider correlation between alternatives. This may be particularly serious when modelling route choices, as strong correlation between the alternative routes may arise due to overlapping. To address this issue, different models have been proposed. We consider a C-Logit model (Cascetta *et al.*, 1996), which includes a “commonality factor”  $CF_i$  in the MNL utility function to capture correlation between alternatives, as shown in (4.5). The inclusion of the commonality factor helps in correcting the models’ fit and predictions by lowering the probabilities of choosing similar alternatives.

$$P_{iq} = \frac{\exp(V_{iq} + \beta \cdot CF_i)}{\sum_{j \in A(q)} \exp(V_{jq} + \beta \cdot CF_j)} \quad (4.5)$$

$\beta$  is a negative parameter that captures the travellers perception towards correlated alternatives (e.g. if  $\beta$  equals zero, all routes are considered independent; if  $\beta$  is large in magnitude, independent routes will tend to be chosen over correlated routes). The commonality factor  $CF_i$  can be defined in many ways (different specifications may be found in Prato, 2009). The specification used in this study is shown in (4.6), where  $l_a$  is the length of link  $a$ ,  $L_i$  is the length of route  $i$ , and  $\delta_{aj}$  is equal to 1 if link  $a$  belongs to route  $j$  or 0 otherwise.

$$CF_i = \ln \sum_{a \in i} \left( \frac{l_a}{L_i} \cdot \sum_{j \in A(q)} \delta_{aj} \right) \quad (4.6)$$

#### 4.2.2 Metro Networks for Analysis

We conduct an empirical analysis to compare route choice decision-making on the London Underground system (based on the work of Guo, 2011) and the Santiago Metro system (based on the work of Raveau *et al.*, 2011). Both systems have high ridership (3 million daily trips in London Underground and 2.3 million daily trips in Santiago Metro) but differ in size (London Underground has 402 Km of length, while Santiago Metro only 103 Km) and complexity. Based on their characteristics, the results from the comparison could be generalized to public transport systems in general. Also, as both systems conduct similar route choice surveys for planning purposes, we can specify and estimate the same utility function on both networks to compare the behaviour characteristics of the respective travellers.

The London's trip database is based on data collected by Transport for London (TfL) from 1998 to 2005. During the period of analysis, the London Underground network had 11 lines, 255 stations and 72 transfer stations. The database consists of 17,073 individual trips along the day between 2,127 OD pairs (3% of the total OD pairs of the network) with two or more alternative routes. The details of the database can be found in Guo and Wilson (2011).

The Santiago's trip database was obtained from an origin-destination survey conducted at Metro stations in October 2008. When the survey was conducted the network consisted of 5 Metro lines, 85 stations and 7 transfer stations. The database obtained from the survey consists of 28,961 individual trips along the day between the 1,365 different OD pairs (19% of the total OD pairs of the network) which have two or more alternative routes to travel (i.e. those OD pairs where there is a route choice decision). A specific analysis of the peak-hours travels can be found in Raveau *et al.* (2011).

### **4.2.3 Topology and Map Effects**

One key element to be analysed is the influence of the network's topology on the route choice, as travellers might tend to prefer those alternatives that seem more direct between the origin and destination stations (aside from other variables such as travel time). To test this hypothesis we define an angular cost (Raveau *et al.*, 2011) to measure how direct/indirect a given route is.

On most public transport systems in the world the network is presented to the users through schematic maps, where all the relevant information for travelling (different lines, stations, transfer points, etc) is included. To enhance the understanding and make the maps easier to read, the public transport network tends to be distorted in the maps (Ovenden, 2007). As public transport maps usually do not include service information (such as travel time or crowding levels), their distortion might affect the decisions made by the travellers.

Figure 4-1 shows the true geographic (panel *a*) and schematic map (panel *b*) depictions of the London underground and Santiago Metro networks. On both cases the map displayed in the stations and consulted (or at least seen) by users on a day-to-day basis contains serious distortions, both in the relative location of the stations and in the distances between them. This way, users' angular and geometric perceptions may also diverge from the geographical reality, which may induce them to choose routes that are not the best for them.

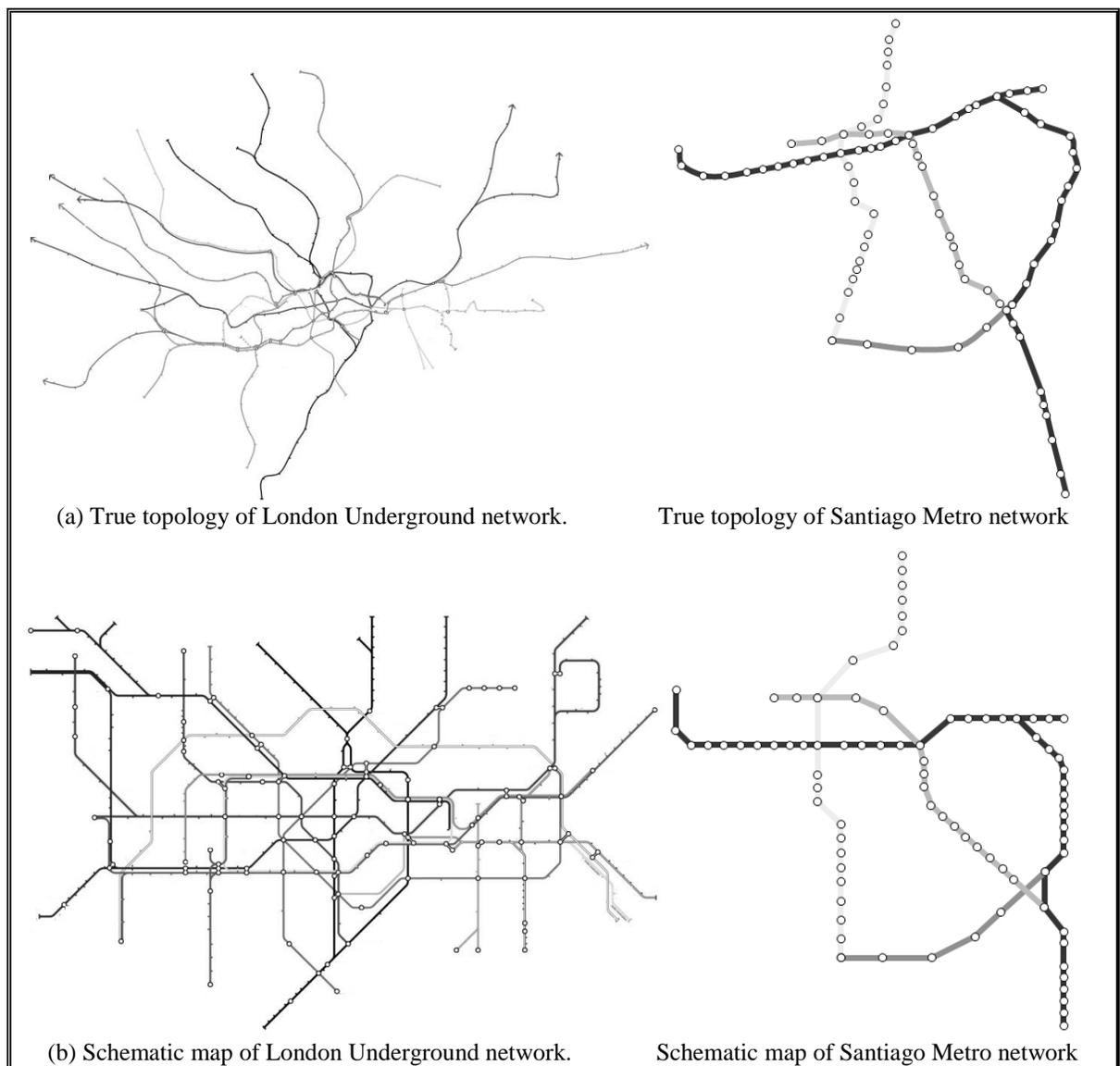


Figure 4-1: London Underground and Santiago Metro Networks Topologies

Although the Santiago Metro network has grown considerably over the last decade, and will continue growing, it is still far away of being as complex and dense as the London Underground network. Also, the distortion is less severe (the correlation between the true and map distances in Santiago Metro is 94%, while in London Underground is only 22%). This way, the topology and map impacts on the travel decisions should be more significant in London.

#### **4.2.4 Set of Alternative Routes**

When dealing with probabilistic route choice models (such as a C-Logit model) it is fundamental to explicitly define an appropriate set  $A(q)$  of available alternatives. This is not trivial, as there is no sure way of knowing which routes were considered by the travellers but were not chosen. For this, many approaches (both deterministic and stochastic) have been developed for constructing sets of alternative routes, trying to replicate the mental selection process undertaken by travellers (for a complete review on these approaches see Prato, 2009).

For the Santiago Metro study case the set of alternative routes for each OD pair was generated based on all the routes chosen by the travellers of that OD pair. This was possible due to the large ratio between trips and OD pairs (21.2 trips per OD pair, in average), so in every one of the 1,365 considered OD pair there were at least two routes used. This way, no methodological approach for generating alternative routes was necessary, and the set obtained comes directly from the observed traveller's behaviour. This procedure is not common in the literature (in most surveys the number of trips per OD pair tends to be low), but it can be consider an ideal scenario, as no exogenous assumptions must be made regarding how travellers select alternative routes. Nevertheless, if choice set generation approaches were to be applied, the resulting choice set would be very similar to the one obtained by observing the travellers choices, as in this case the network is not particularly big or dense.

On the London Underground study case the ratio between trips and OD pairs is lower (7.7 trips per OD pair, in average) and in most OD pairs there is only one chosen route. Therefore, a methodological approach for obtaining a set of available alternatives is needed. A labelling approach (Ben-Akiva *et al.*, 1984) was used to generate the set of available alternatives, by obtaining the shortest path for different definitions of “link costs”. In this case, different combinations of weights for the time components (travelling, waiting, walking and dwelling), transfer penalties and reliability measures were considered. Details about the application of the method and its results (such as efficiency, effectiveness and coverage) can be found in Guo and Wilson (2011). A minimum number of two and up to six different routes were obtained for each of the 2,127 OD pairs. The main advantage of using a labelling approach lies on its behavioural assumptions, as different route selection strategies can be addressed (e.g. some travellers might consider the fastest route, while others might want a route that is good in terms of comfort) and every route is optimal in some sense.

The distribution of OD pairs and trips according to the number of available routes is presented in Table IV-1. As the London Underground network is denser, there are more alternative routes for travelling than in the Santiago Metro network. It can be seen as well that in both cities the OD pairs distribution and the trips distribution is very similar: no particular OD structure (in terms of available alternatives) tends to concentrate more trips than another.

Table IV-1: Route Availability Statistics

Available Routes	Santiago Metro		London Underground	
	OD pairs	Trips	OD pairs	Trips
2	1,322 (96.9%)	26,950 (93.1%)	1,161 (54.6%)	9,589 (56.2%)
3	37 (2.7%)	1,892 (6.5%)	694 (32.6%)	5,559 (32.6%)
4	6 (0.4%)	119 (0.4%)	202 (9.5%)	1,479 (8.7%)
5	0 (0.0%)	0 (0.0%)	58 (2.7%)	348 (2.0%)
6	0 (0.0%)	0 (0.0%)	12 (0.6%)	98 (0.6%)
Total	1,365 (100.0%)	28,961 (100.0%)	2,127 (100.0%)	17,073 (100.0%)

#### 4.2.5 Model Description

In the literature there are multiple route choice models (mainly MNL models or extensions of it) based on users' socio-economic characteristics and route attributes (Ramming, 2001; Prashker and Bekhor, 2004; Liu *et al.*, 2010). However, the attributes included in these models are all tangible and quite limited to travel time components, fare and transfers. This study presents a more complete route choice analysis on Metro networks, incorporating variables related to the different times involved (travel, waiting and walking times), trains and stations usage, transfer environment, network topology and socio-demographics. The additional variables, included to improve the ability to explain and predict route choice decisions, are discussed below. Some of these additional variables were considered separately by Guo (2011) and Raveau *et al.* (2011); in this study we consider their joint effect when applied to two different metro networks. Thus, the main contributions of this paper when compared with the work previously conducted by Guo (2011) and Raveau *et al.* (2011), aside from the behavioural comparison of both cities (something rarely done in travel demand literature), are the inclusion of socio-demographic variables and route correlation. It's important to take into account that no fare variable was included because all alternative routes between a given OD pair have the same monetary cost.

The most traditional and important variable used to explain route choice behaviour is the travel time. Users tend to look for the fastest way of getting from their origin to their destination, and the travel time is the main criterion to discard unattractive (i.e. slow) alternatives. We consider three different time components: the **in-vehicle time**, the **waiting time** at the origin station and all subsequent transfer stations, and the **walking time** when transferring (as the origins and destinations are fixed, we obviate access and egress times). These different time components are considered separately to address their different perception and importance in the travellers' decision-making process.

Regarding the transferring experience, we firstly consider the total **number of transfers** of each alternative route; as the actual transferring time is captured by the walking and

waiting time variables, this variable solely captures the displeasure of having to transfer. To further understand the transferring valuation, we differentiate between possible types of transfers. In terms of stations layout, the transfers can either be made between **ascending levels** (i.e. going up), **even levels** (usually walking across the platform) and **descending levels** (i.e. going down). In terms of stations infrastructure, the transfers can be **assisted** (made completely using escalator and/or lift), **semi-assisted** (made partially using escalator and/or lift and partially on foot), and **non-assisted** (made completely on foot).

To address the level of comfort and crowding experienced by the public transport users during their trip, the **mean occupancy** along the route was included in the models. This variable is defined as the distance-weighted ratio between passengers load and train capacity in peak hour (even though not all trips are made in peak hour, this specification presented the best goodness-of-fit). By definition the rate can vary between 0 (trains travelling empty along all arcs of the route) and 1 (trains travelling fully loaded along all arcs of the route).

Two additional variables related to train usage and extreme crowding levels were included in the model. We distinguish those transfer stations where there is the **possibility of getting a seat**, depending on the occupancy of the trains leaving those stations. On London Underground this happens when the occupancy is 20% or less, while on Santiago Metro this happens when the occupancy is 15% or less (these percentages represent the percentage of the capacity that corresponds to seats). On the other hand, we distinguish those transfers where there is the **possibility of not boarding** the first train, and having to wait for the next train. We observe that on London Underground this happens when the occupancy is 70% or more, while on Santiago Metro this happens when the occupancy is 85% or more. Based on these thresholds, it may seem that Santiago Metro's users are more willing to board crowded trains, instead of simply waiting for the next one.

To deal with the topology's effect on the route choices of the travellers, we include different topological factors. Following Guo (2011), we include the **distance** and **number of stations** between the origin and the destination along the different routes. Due to

variable dwelling times, spacing between stations and distortion of the schematic maps (on which the topological variables are based), these variables are not highly correlated with the travel time components and can be included in the models without problems. We also include an **angular cost** to measure how direct a certain route is. The angular cost is defined accordingly with Raveau *et al.* (2011) as shown by (4.7), where  $s$  represents a leg of the route,  $d_s$  is the distance of the leg  $s$ , and  $\theta_s$  is the angle formed between the destination station, the first station of leg  $s$  and the last station of leg  $s$ . Travellers will tend to choose the routes that appear to be more direct between their origin and destination.

$$\text{Angular Cost} = \sum_s d_s \cdot \sin\left(\frac{\theta_s}{2}\right) \quad (4.7)$$

Also related to the topology's effect, two additional variables were considered. We say that a route **turns back** if it has a transfer station (or the destination, if it is the last leg) that is closer to the origin than the transfer station immediately before it. We say that a route **turns away** if it has a transfer station that is further from the destination than the transfer station (or the origin, if it is the first leg) immediately before it. These variables are an application of Dial's definition of reasonable routes (Dial, 1971), and capture the travellers reluctance to choose those routes that appear to be unreasonable in terms of geometry.

Finally, the socio-demographic information included in both databases consists of the **purpose of the trip** (we have restrictive purposes like going to work and going to study, and non-restrictive purposes like going home and going shopping), the **gender** and **age** of the traveller, **fare type** (full fare or discount pass), and **time of the day** the trip is made. All these variables were also considered for formulating interactions with the rest of the explanatory variables previously described.

Information regarding the **frequency of the journey** (e.g. daily, weekly, monthly, first time ever) was available only for London and **income level** was available only for Santiago; these two variables were excluded from the modelling for comparison purposes

It is important to notice that the income level, a traditionally important variable in transport modelling, was not relevant anyway when modelling the decisions in Santiago Metro, as no monetary cost is included as an explanatory variable (therefore, there is no income effect in the route choices).

### 4.3 Results and Analysis

Based on the route choice data from both networks, C-Logit models were estimated to understand the travellers' decision-making process. Other models that consider route correlation were also estimated (in particular, the Path Size Logit Model, the Cross Nested Logit, and the Paired Combinatorial Logit), but the C-Logit resulted on the best goodness-of-fit for both networks. When specifying the model, all topological variables were defined based on the schematic maps of the networks, rather than based on their true topology, as they resulted on better results (this issue will be further treated on Section 4.3.2). The estimated parameters, their  $t$ -values and goodness-of-fit indicators for the model are presented in Table IV-2. For the explanatory variables that interact with socio-demographic variables, the base parameter and marginal parameters are presented; this way, for example, the parameter of in-vehicle time for a trip on the morning peak period with a restrictive purpose in Santiago would be  $-0.074 - 0.014 - 0.025 = -0.113$ .

All the parameters presented in Table IV-2 have the expected signs and are statistically significant at 95% confidence level. For the station layout, the descending transfer parameter was fixed in zero as the base category; for the station infrastructure, the assisted transfer was defined as base category. It can be seen that an even transfer is preferred over changing levels; if a change in levels must be made, users prefer to descend rather than ascend, as ascending seems to be mentally associated with a greater effort. In terms of infrastructure, as the grade of assistance increases, the transfer experience improves. The parameter of the commonality factor, related to route correlation due to overlapping, has a similar statistical significance on both cities. This means that, at some degree, travellers in London and Santiago tend to perceive overlapping routes as non-independent alternatives.

Table IV-2: Parameters Estimates for London and Santiago

Attribute	London Underground		Santiago Metro	
	Parameter	<i>t</i> -value	Parameter	<i>t</i> -value
In-Vehicle Time	- 0.121	- 9.20	- 0.074	- 6.30
+ <i>Morning Peak</i>	- 0.084	- 5.10	- 0.014	- 2.52
+ <i>Afternoon Peak</i>	n. a. <sup>(1)</sup>	n. a.	- 0.009	- 2.62
+ <i>Restrictive Purpose</i>	- 0.042	- 2.42	- 0.025	- 5.93
Waiting Time	- 0.269	- 14.21	- 0.083	- 3.62
+ <i>Morning Peak</i>	- 0.208	- 5.94	- 0.094	- 2.60
Walking Time	- 0.299	- 9.32	- 0.210	- 2.34
+ <i>Women</i>	- 0.048	- 2.06	- 0.074	- 2.67
Number of Transfers	- 1.321	- 4.14	- 0.662	- 4.19
Ascending Transfers	- 0.206	- 2.53	- 0.308	- 2.60
Even Transfers	0.613	3.82	n. a. <sup>(2)</sup>	n. a.
Descending Transfers	0.000 <sup>(3)</sup>	n. a.	0.000 <sup>(3)</sup>	n. a.
Assisted Transfers	0.000 <sup>(3)</sup>	n. a.	0.000 <sup>(3)</sup>	n. a.
Semi-Assisted Transfers	- 0.271	- 5.30	n. a. <sup>(2)</sup>	n. a.
Non-Assisted Transfers	- 0.398	- 6.33	- 0.182	- 5.11
Mean Occupancy	- 2.898	- 3.25	- 0.935	- 5.10
Getting a Seat	0.117	2.22	0.105	3.68
Not Boarding	- 0.502	- 6.23	- 0.358	- 2.29
Angular Cost	- 0.088	- 3.89	- 0.029	- 3.84
+ <i>Restrictive Purpose</i>	0.049	3.79	0.011	2.70
Map Distance	- 0.364	- 5.43	- 0.278	- 4.83
Number of Stations	- 0.424	- 5.07	- 0.168	- 3.62
Turning Back	- 0.650	- 8.85	- 0.142	- 8.10
Turning Away	- 0.943	- 7.77	- 0.231	- 8.87
Commonality Factor	- 0.396	- 3.74	- 0.541	- 3.41
Sample Size	17,073		28,961	
Log-Likelihood	- 6,690		- 12,881	
Corrected $\rho^2$	0.567		0.383	

<sup>(1)</sup>: This parameter is not statistically significant.

<sup>(2)</sup>: These kinds of transfers do not exist in Santiago Metro.

<sup>(3)</sup>: These parameters were fixed as base categories for the layout and infrastructure variables.

For both cities, the socio-demographic are most relevant when interacted with the travel time components. There is also a significant interaction between the angular cost and the purpose of the trip (those travelling to work or study are more likely to be commuters, who are less influenced by the map). Other interactions, particularly with the variables related to transferring and crowding, are not statistically significant in the presence of time

interactions. On both cities the socio-demographic variables of age and fare type (which are correlated, as fare discount are available for students and elders) did not provide any significant interaction with the explanatory variables, which is somehow surprising given the rich and broad trip sample used for both cities. The interaction of afternoon peak time and in-vehicle travel time was significant only for Santiago. It is worth noticing the surprising similarity of the explanatory and socio-demographic variables that are relevant when understanding the route choices in London and Santiago.

#### **4.3.1 Perceptions and Valuations**

The parameters obtained for the London Underground and Santiago Metro are not directly comparable between each other, as the MNL-based models (such as the C-Logit) have different scales from their Gumbel errors variances (Train, 2009), but marginal rates of substitution can be derived from these parameters and compared for both networks. These marginal rates are presented in Table IV-3. As the interaction between socio-demographic variables and explanatory variables provide multiple in-vehicle time parameters (4 in London and 6 in Santiago), we are only presenting and analysing the results for morning peak trips with restrictive purpose; the results for all other trips can be obtained analogously. For this direct comparison, the angular cost and the map distance are excluded as they don't have a measurement scale.

On both cities the waiting and walking times are more valuable than the in-vehicle travel time, but while in London the value of the waiting time is greater than the value of the walking time, in Santiago the relationship is the opposite. When choosing transferring alternatives, Santiago Metro users will be more willing to wait and London Underground users will be more willing to walk, in part because Santiago Metro has fewer transfer options and their environment is generally simpler, while London Underground presents many transfer alternatives and the connection between platforms could be longer and more complex due to the incremental addition of lines over the time.

Table IV-3: Marginal Rates of Substitution - Morning Peak &amp; Restrictive Purpose

<b>Attribute</b>	<b>London Underground</b>	<b>Santiago Metro</b>
1 min of waiting time	1.93 min in-vehicle	1.57 min in-vehicle
1 min of men walking time	1.21 min in-vehicle	1.86 min in-vehicle
1 min of women walking time	1.40 min in-vehicle	2.51 min in-vehicle
1 base transfer <sup>(2)</sup>	5.35 min in-vehicle	5.86 min in-vehicle
1 ascending transfer	0.83 min in-vehicle	2.73 min in-vehicle
1 even transfer	2.48 min in-vehicle <sup>(1)</sup>	n.a. <sup>(3)</sup>
1 non-assisted transfer	1.61 min in-vehicle	1.61 min in-vehicle
1 semi-assisted transfer	1.10 min in-vehicle	n.a. <sup>(3)</sup>
1 % of occupancy	0.12 min in-vehicle	0.08 min in-vehicle
Possibility of seating	0.47 min in-vehicle <sup>(1)</sup>	0.93 min in-vehicle <sup>(1)</sup>
Possibility of not boarding	2.03 min in-vehicle	3.17 min in-vehicle
1 station along the way	1.72 min in-vehicle	1.49 min in-vehicle
Turning back	2.63 min in-vehicle	1.26 min in-vehicle
Turning away	3.82 min in-vehicle	2.04 min in-vehicle

<sup>(1)</sup>: Absolute value, these variables represent a gain in utility.

<sup>(2)</sup>: Based on the model specification, this base transfer corresponds to an assisted descending transfer where there is no possibility of either getting a seat or not being able to board.

<sup>(3)</sup>: These kinds of transfers do not exist in Santiago Metro.

Even though in both cities the valuation of a base transfer (this is, an assisted descending transfer where there is no possibility of either getting a seat or not being able to board) might seem very similar, it's necessary to take into account the marginal valuations of the layout and infrastructure variables. As in London Underground the most used parts of the network are underground, a descending change of levels will tend to have associated a corresponding ascending change of levels; this way the valuation of ascending transfers is lower in London Underground than in Santiago Metro. For example, while London Underground is completely underground in the city centre of London, in the city centre of Santiago almost 30% of the Santiago Metro network is overground.

To further analyse the users' valuation of transfer experiences, Table IV-4 presents the marginal rates of substitution for all possible transfer types (once again for morning peak trips with restrictive purpose). There is a high variability in London transfer's valuations (between 2.39 and 9.83 minutes of in-vehicle travel time) and Santiago transfer's valuations (between 4.93 and 13.36 minutes of in-vehicle travel time). This way, the

necessity of distinguishing by station layout, infrastructure and occupancy is clear; otherwise the heterogeneity in preferences would not be captured. Weighting each valuation for the number of transfer of each type made on a regular day, the average transfer penalty in London Underground is 7.0 minutes and in Santiago Metro is 10.2 minutes. It seems that, in general, London Underground users are more willing to transfer than Santiago Metro users. As the London Underground network is bigger and denser, there are more transfer possibilities and Londoners are more used to transferring: 69% of the trips in London involve at least one transfer, while in Santiago is the 47% of the trips.

Also from Table IV-3 it can be seen that occupancy is more valued in London, as the London Underground users are less willing to travel in crowded trains than Santiago Metro users. This also explains why Londoners are more sensitive towards not boarding a train due to crowding, consistent with the lower threshold obtained when defining the possibility of not boarding. As an example of this, TfL defines a “crowded train” as 4 passengers per m<sup>2</sup>, while Santiago Metro defines a “crowded train” as 6 passengers per m<sup>2</sup>. On the other hand, extreme occupancies (getting a seat or not being able to board) are more valued in Santiago, probably because in London the occupancies are lower.

Table IV-4: Transferring Valuations - Morning Peak & Restrictive Purpose

<b>London Underground</b>				
<b>Characteristics</b>		<b>Getting a Seat</b>	<b>Intermediate</b>	<b>Not Boarding</b>
<b>Ascending</b>	<b>Assisted</b>	5.71 min	6.18 min	8.21 min
<b>Ascending</b>	<b>Semi-assisted</b>	6.84 min	7.28 min	9.31 min
<b>Ascending</b>	<b>Non-assisted</b>	7.32 min	7.79 min	9.83 min
<b>Even</b>		2.39 min	2.87 min	4.90 min
<b>Descending</b>	<b>Assisted</b>	4.87 min	5.35 min	7.38 min
<b>Descending</b>	<b>Semi-assisted</b>	5.97 min	6.45 min	8.48 min
<b>Descending</b>	<b>Non-assisted</b>	6.49 min	6.96 min	8.99 min
<b>Santiago Metro</b>				
<b>Characteristics</b>		<b>Getting a Seat</b>	<b>Intermediate</b>	<b>Not Boarding</b>
<b>Ascending</b>	<b>Assisted</b>	7.66 min	8.58 min	11.75 min
<b>Ascending</b>	<b>Non-assisted</b>	9.27 min	10.19 min	13.36 min
<b>Descending</b>	<b>Assisted</b>	4.93 min	5.86 min	9.03 min
<b>Descending</b>	<b>Non-assisted</b>	6.54 min	7.47 min	10.64 min

Avoiding one station along the way (this is, taking a route with one station less between the origin and the destination) is valued similarly on both cities. To further understand this marginal rate, the mean travel time between two consecutive stations (considering dwelling time) in London Underground and Santiago Metro are 1.8 minutes and 1.6 minutes, respectively. The similarity between these values and the marginal rates obtained indicates that the public transport users have a good notion of the operation times of the networks.

Londoners are much more sensitive to taking routes that seem unreasonable in terms of their geometry, either because they turn back to the origin or turn away from the destination. This could be due to the high distortion of the London Underground map. Another possible explanation is that the London Underground network generates more unreasonable alternative routes due to its complexity and density, which in reality are barely considered by the users.

#### **4.3.2 Specification of Topological Variables**

As mentioned before, the topological variables (in particular, the angular cost, the distance, and the impressions of turning back and turning away) can be specified either using the true topology of the networks or their schematic maps. Additionally to the model specified using the schematic maps (whose complete results were presented in Table IV-2), the alternative model specified using the true topologies was also estimated. Table IV-5 summarizes some of the main differences between both models, by presenting the *t*-values of the topological variables (all other control variables are also included but we focus on these variables, as these are the ones that change in specification) and goodness-of-fit indicators. It's important to take into account that in the new model, specified with the true topologies, all parameters have the correct sign and that the parameters of the rest of the variables (those omitted in Table IV-5) are similar to the parameters obtained for the model specified using the schematic map.

Table IV-5: Specification of Topological Variables

Attribute	London Underground		Santiago Metro	
	True <i>t</i> -values	Map <i>t</i> -value	True <i>t</i> -values	Map <i>t</i> -value
Angular Cost	- 1.62	- 3.89	- 2.12	- 3.84
+ <i>Restrictive Purpose</i>	1.34	3.79	2.03	2.70
Distance	- 3.64	- 5.43	- 2.81	- 4.83
Turning Back	- 4.02	- 5.07	- 3.97	- 3.62
Turning Away	- 4.30	- 8.85	- 4.12	- 8.10
Sample Size	17,073		28,961	
Log-Likelihood	- 6,716	- 6,690	- 12,957	- 12,881
Corrected $\rho^2$	0.565	0.567	0.379	0.383

It can be seen from Table IV-5 that all the topological variables have less statistically significant parameters when the true topologies are used to specify them. In particular, the angular cost turns out to be not statistically significant at 90% confidence level for the London Underground. Better goodness-of-fit is obtained when the topological variables are specified using the schematic maps information. This reassures the idea that public transport users take into account schematic maps (with which are more familiarized) rather than the actual topology of the networks. This strongly indicates the relevance of the way the information is provided to users, as any map distortion could induce them to make inadequate decisions.

### 4.3.3 Model Specification and Information Omission

To analyse the importance of a correct and complete model specification, five additional models with different amount of variables were estimated: one without the occupancy-related variables, one without the transfer-related variables, one without the topology-related variables, one without the socio-demographic variables, and without the four types of variables (this is, considering only the time-related variables). Even though all models have correct parameters' signs, some serious problems in terms of bias can arise. Table IV-6 presents the marginal rates of substitution between time components and the global goodness-of-fit indicator for all models.

Table IV-6: Information Omission - Morning Peak &amp; Restrictive Purpose

<b>Model</b>	<b>Time</b>	<b>London Underground</b>	<b>Santiago Metro</b>
Complete Model	Waiting	1.93 min in-vehicle	1.57 min in-vehicle
	Men's walking	1.21 min in-vehicle	1.86 min in-vehicle
	Women's walking	1.40 min in-vehicle	2.51 min in-vehicle
	Corrected $\rho^2$	0.567	0.383
Model without Occupancy variables	Waiting	1.80 min in-vehicle	2.38 min in-vehicle
	Men's walking	1.29 min in-vehicle	1.90 min in-vehicle
	Women's walking	1.43 min in-vehicle	2.68 min in-vehicle
	Corrected $\rho^2$	0.558	0.373
Model without Transfer variables	Waiting	2.57 min in-vehicle	3.82 min in-vehicle
	Men's walking	1.98 min in-vehicle	2.50 min in-vehicle
	Women's walking	2.31 min in-vehicle	3.11 min in-vehicle
	Corrected $\rho^2$	0.535	0.369
Model without Topological variables	Waiting	0.61 min in-vehicle	2.30 min in-vehicle
	Men's walking	0.46 min in-vehicle	2.53 min in-vehicle
	Women's walking	0.53 min in-vehicle	3.77 min in-vehicle
	Corrected $\rho^2$	0.520	0.357
Model without Socio-Demographic variables	Waiting	1.66 min in-vehicle	1.46 min in-vehicle
	Men's walking	1.17 min in-vehicle	1.62 min in-vehicle
	Women's walking	1.17 min in-vehicle	1.62 min in-vehicle
	Corrected $\rho^2$	0.566	0.382
Model without Occupancy, Transfer, Topological and Socio-Demographic variables	Waiting	0.80 min in-vehicle	4.48 min in-vehicle
	Men's walking	0.73 min in-vehicle	2.64 min in-vehicle
	Women's walking	0.73 min in-vehicle	2.64 min in-vehicle
	Corrected $\rho^2$	0.483	0.334

It can be seen that, when omitting information by excluding explanatory variables, there is bias in the marginal rates of substitution obtained. The incomplete models tend to overestimate the valuations of waiting and walking times; this is expected when, for example, transfer variables are omitted, as these two time components (directly related with the transfer experience) must capture the disutility of transferring. When the topological variables are excluded from the London model, underestimations of the marginal rates of substitution are obtained, while in Santiago the marginal rates are overestimated. One possible explanation might be the topological variables tend to guide Londoners to choose transfer alternatives that involve longer walking and waiting time, so when these variables are excluded from model, it would appear that Londoners have a

lower valuation on transfer walking and waiting. The opposite happens in Santiago, where the topological variables tend to guide passengers to choose transfer stations with shorter walking and waiting time.

In terms of goodness-of-fit, the worst results (aside from when all four subsets of variables are excluded) are obtained when the five topological variables are omitted from the model, rather than when the five transfer variables or the three occupancy variables are omitted. This happens in both cities, and reassures the importance of the topology in the travellers' decision-making process despite the very different network structure and complexity in the two systems.

#### **4.3.4 Implications on Transportation Planning**

Methodological comparisons between travel behaviour in different cities and contexts using the same approach are rarely done, mainly because limitations in terms of data access. Comparisons are generally between models with different specifications, different data collection methods, or different choice contexts. This way, the comparison of the route choice preferences in the London Underground and Santiago Metro networks can provide some findings about route choice preferences that might be extendable to any metro network, and also provide some insights for transportation planning.

One of the main findings of the study is the (somehow surprising) similitude of the results obtained for both cities, particularly in terms of which are the relevant variables when explaining the decisions of the travellers. Users of both systems consider different time components, transfer-related characteristics, crowding variables, and topological variables (which come from the respective schematic maps). Even the socio-demographic variables that interact (and do not interact) with the explanatory variables were found to be almost the same.

Having found so similar results in so different networks, we can argue that the explanatory variables included in this study can provide satisfactory results when analysing and modelling route choice behaviour in any metro network, and can be a good starting point when dealing with route choice in multi-modal public transport networks.

It is important to notice that, although the relevant variables were found to be very similar in both networks, the specific preferences and valuations were generally different within a reasonable and expected range in both cities. This way, none of the particular results or valuations should be transferred and used in other context; with the appropriate data, the specification proposed could be used to obtain specific attribute valuations.

Finally, from a social planning point of view, it has been shown and confirmed that travellers take into account a wide variety of attributes when choosing their routes, and that their preferences can vary depending on their gender, the time of the day or the purpose of the trip. Travellers do not only care about travel times and number of transfers, but also care about crowding, transfer characteristics and topological factors. These results should be considered by the authorities and the planners, as many of the factors included in this study are not generally included in traditional route choice models. Particularly interesting is the impact of the schematic maps on the travellers' decisions, especially of those who are not familiar to the network (i.e. are not commuters).

#### **4.4 Conclusions**

Route choice modelling variables are traditionally limited to some tangible factors such as time and fare that, although relevant, fail to accommodate different aspects of traveller's behaviour. In this study we specify and estimate route choice models for metro networks that consider four different kinds of variables: travel time components, transfer experience characteristics, occupancy and comfort indicators, and network topology variables. All these variables result significant for understanding for travellers' behaviour. This reassures

the idea that public transport users take into account a wide variety of attributes when choosing routes.

An incomplete model specification can result on biased results, especially attributes valuations, which could have a great repercussion when valuating transportation projects (such as network expansions or operational changes). Also related to the specification, the topological variables have a greater explanatory power when specified based on the networks' schematic maps rather than their true topology. Although distortions in these schematic maps can induce route choice decisions that reduce utility for some users, they could also be exploited by planners to induce optimal route choices to make better use of public transport system capacity (Jankowski *et al.*, 2001). This is particularly relevant in London Underground, where the map distortion is extremely big (probably bigger than any other public transportation system in the world). Schematic maps can also be complemented with information regarding level of service, to enable a more informed decision by the travellers (Hochmair, 2009).

As the route choice model was estimated for the London Underground and Santiago Metro networks, a comparison between traveller's behaviour in both cities was possible. London Underground users seem to be more willing to transfer than Santiago Metro users, as they are more used to required transfers when travelling. When transferring in London, users will tend to prefer to spend more walking than waiting, while in Santiago users will tend to prefer to waiting time over walking time. In terms of occupancy and comfort, Londoners are less willing to travel in crowded trains and will tend to stop boarding when the trains reach a lower threshold than in Santiago; on the other hand, they care less about getting a seat than Santiago Metro users.

### **Acknowledgments**

We wish to thank the Chilean Fund for Promotion of Scientific and Technological Development, FONDEF (Project D10I1049), the Chilean Fund for Scientific and

Technological Research, FONDECYT (Project 1110720), the VREF Centre of Excellence on BRT, Transport for London and Metro de Santiago.

### **Publication History**

The basis of this article was accepted for publication in 2014 on *Transportation Research Part A*, under the title “A behavioural comparison of route choice on metro networks: time, transfers, crowding, topology and socio-demographics”. The results covered in this article were presented at the *13th International Conference on Travel Behaviour Research* (Toronto, 2012), at the *12th International Conference on Advanced Systems for Public Transport* (Santiago, 2012), at the *1st European Symposium on Quantitative Methods in Transportation Systems* (Lausanne, 2012), and at the *14th EURO Working Group on Transportation* (Paris, 2012).

### **References**

- Ben-Akiva, M.E., Bergman, M.J., Daly, A.J. and Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. In J. Volmuller and R. Hamerslag (eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*. VNU Science Press, Utrecht, 299-330.
- Bovy, P.H.L. and Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. *Transportation*, 32(4), 341-368.
- Bovy, P.H.L. and Stern, E. (1990). *Route Choice. Wayfinding in Transport Networks*. Kluwer Academic Publishers, Norwell, MA.
- 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: Lesort, J.B. (ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Pergamon, Lyon, 697-711.

Dial, R.B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5, 83-113.

Guo, Z. (2011). Mind the map! The impact of transit maps on path choice in public transit. *Transportation Research Part A*, 45(7), 625-639.

Guo, Z. and Wilson, N.H.M. (2011). Assessing the cost of transfer inconvenience in public transport systems: a case study of the London Underground. *Transportation Research Part A*, 45(2), 91-104.

Hochmair, H. (2009). The influence of map design on route choice from public transportation maps in urban areas. *The Cartographic Journal*, 46(3), 242-256.

Hunt, J.D. (1990). A logit model of public transport route choice. *ITE Journal*, 60(12), 26-30.

Jankowski, P., Andrienko, N. and Andrienko, G. (2001). Map-centred exploratory approach to multiple criteria spatial decision making. *International Journal of Geographical Information Science*, 15(2), 101-127.

Liu, Y., Bunker, J. and Ferreira, L. (2010). Transit users' route-choice modelling in transit assignment: a review. *Transport Reviews*, 30(6), 753-769.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers of Econometrics*. Academic Press, New York, 105-142.

Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Edition, John Wiley and Sons, Chichester.

Ovenden, M. (2007). *Transit Maps of the World*. Penguin Group, New York.

Prato, C.G. (2009). Route choice modelling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), 65-100.

Prashker J.N. and Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, 24(4), 437-463.

Ramming, M.S. (2001). *Network Knowledge and Route Choice*. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Raveau, S., Muñoz, J.C. and de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A*, 45(2), 138-147.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. 2<sup>nd</sup> Edition, Cambridge University Press, Cambridge.

## 5 ROUTE CHOICE STRATEGIES ON TRANSIT NETWORKS

**Sebastián Raveau**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

**Juan Carlos Muñoz**

Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile

### **Abstract**

Understanding and modelling mode and route decisions on a transit system is a key element in urban transportation planning. For doing this it is necessary to identify and model the individuals' behaviour and the route choice strategies they follow. Traditional models propose three route choice strategies: Minimum Itineraries (fixed stations and fixed single-lines), Minimum Routes (fixed stations and variable multiple-lines), and Minimum Hyper-Routes (variable stations and variable multiple-lines). Nevertheless, there is no empiric evidence to support or reject the use of a particular strategy as how travellers actually behave. In this study we analyse, through an econometric approach, the actual strategies followed by Santiago transit travellers, and propose a modelling approach that deals with the differences in behaviour among individuals. Our database suggests that 67% of the transit travellers follow the basic Minimum Itineraries approach, without considering common lines. Socio-economic characteristics of the individuals are significant to understand the route choice strategies followed by the transit travellers, enhancing the explanatory capability of mode/route choice models. The proposed approach should be tested in different transit systems.

## 5.1 Introduction

Traditional route choice and trip assignment models, applied either on private or public transportation networks, assume rational travellers maximizing their utility level (or alternatively minimizing their cost) associated to the different travel alternatives. Particularly in transit networks in which the time the next vehicle will arrive to the station is uncertain (i.e. frequency based systems without timetables), travellers can reduce their expected total travel time by following different route choice strategies.

The different route choice strategies that can be applied by the travellers differ in complexity and require different levels of information from the system. In literature, it is usual to assume that all travellers are capable of considering high-complexity strategies (which might require developed analytical capacities). Similarly, in literature it is usually assumed that all travellers have perfect information regarding the levels-of-service of all available alternatives. As expected, these assumptions might not be true for a considerable proportion of the travellers.

In this study we contrast the theoretical assumptions regarding route choice strategies, traditionally used for modelling route choices in a transit network, with the decisions actually made by transit travellers. In Section 5.2 we present the methodological background of the different route choice strategies of transit travellers; in Section 5.3 we describe the database used in the study; in Section 5.4 we present the main results of the route choice strategies analysis; and finally in Section 5.5 we present the main conclusions of the study.

## 5.2 Route Choice Strategies

In the literature it is possible to identify three route choice strategies: (i) Minimum Itineraries, (ii) Minimum Routes, and (iii) Minimum Hyper-Routes. These strategies differ in the complexity of the route choice decision-making process. To exemplify the different

route choice strategies, Figure 5-1 presents a simple transit network with a single origin-destination pair, defined by Spiess and Florian (1989). The network has four transit lines, whose travel times between stations are as shown. Additionally we assume that the lines' headways follow an Exponential distribution (so the expected waiting time at each station equals the inverse of the sum of all the frequencies of the lines that serve it, and is independent of how much time has passed since the last bus of each line visited the station) and that capacity constraints are never active. The operational frequencies of each line are also shown in Table V-1.

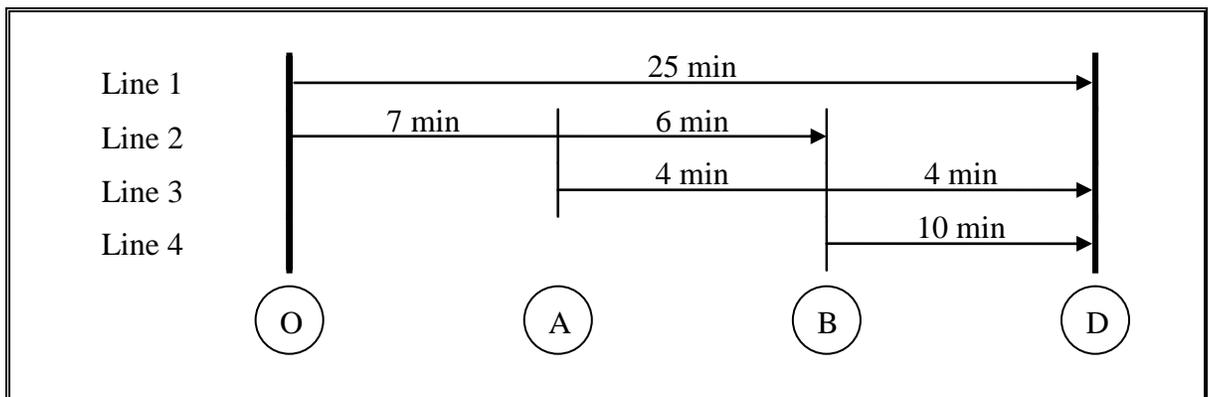


Figure 5-1: Transit Network Example

Table V-1: Frequencies of the Transit Lines for the Network Example

Line	Frequency	Expected Waiting Time
1 / O-D	10 buses per hour	6 minutes
2 / O-A-B	10 buses per hour	6 minutes
3 / A-B-D	4 buses per hour	15 minutes
4 / B-D	20 buses per hour	3 minutes

A route choice strategy can be seen as the set of rules for choosing the transit lines and transfer stations, in order to travel from an origin to a destination. For example, on the transit network depicted in Figure 5-1, two possible strategies are: (i) “take Line 2 to

station A, and then take Line 3 to the destination” or (ii) “take the Line 2 to station B, and then take whichever line arrives first (between Line 3 and Line 4) to the destination”.

### 5.2.1 Minimum Itineraries

This route choice strategy represents the simplest behaviour on a transit network. An itinerary is a sequence of transfer stations and transit lines that join an origin-destination pair. The choice of the minimum itinerary consists on finding the itinerary that has the shortest expected total travel time. In our example transit network (Figure 5-1) there are five reasonable itineraries (i.e. that do not consider boarding the same line in two consecutive trip legs or travelling back to the origin) for travelling between the origin and destination stations. These five itineraries and their associated average travel times are described below.

Itinerary 1: **O – Line 1 – D**

In-vehicle time: 25 min  
 Waiting time: 6 min  
 Total time: 31 min

Itinerary 2: **O – Line 2 – A – Line 3 – D**

In-vehicle time: 7 min + 8 min = 15 min  
 Waiting time: 6 min + 15 min = 21 min  
 Total time: 36 min

Itinerary 3: **O – Line 2 – A – Line 3 – B – Line 4 – D**

In-vehicle time: 7 min + 4 min + 10 min = 21 min  
 Waiting time: 6 min + 15 min + 3 min = 24 min  
 Total time: 45 min

Itinerary 4: **O – Line 2 – B – Line 3 – D**

In-vehicle time:	13 min + 4 min = 17 min
Waiting time:	6 min + 15 min = 21 min
Total time:	38 min

Itinerary 5: **O – Line 2 – B – Line 4 – D**

In-vehicle time:	13 min + 10 min = 23 min
Waiting time:	6 min + 3 min = 9 min
Total time:	32 min

It can be seen that itinerary 1 is the minimum itinerary, as it has the least total expected travel time. Based on this route choice strategy, all individuals travelling from the origin station to the destination station should choose the direct Line 1. The remaining three lines would not be chosen by any transit traveller.

### 5.2.2 Minimum Routes

In this route choice strategy, all passengers travelling in the same origin-destination pair visit the same sequence of transfer stations. However, in this case they select a set of “attractive lines” or “common lines” between each pair of successive stations (Chriqui and Robillard, 1975), in order to reduce the expected waiting times (and therefore reduce the total travel time between origin and destination). Unlike the Minimum Itineraries strategy, between each pair of stations along the way transit travellers will board the first attractive line that arrives, instead of waiting for a specific line. The probability of boarding each of the attractive lines is proportional to its frequency.

On our example transit network (Figure 5-1) we can analyse if it is convenient to define a set of common lines between stations A and B (boarding the first line that passes between Line 2 and Line 3) and between stations B and D (boarding the first line that passes between Line 3 and Line 4). Between A and B the best alternative without common lines is

Line 2 (6 minutes waiting + 6 minutes in-vehicle = 12 minutes of travel time). If we consider boarding the first vehicle of either line, the expected in-vehicle time would be  $(6 \cdot 1/6 + 4 \cdot 1/15) / (1/6 + 1/15) \approx 5.43$  minutes and the expected waiting time would be  $1 / (1/6 + 1/15) \approx 4.29$  minutes. This way, the total travel time would be reduced to 9.72 minutes. Therefore, it is convenient to consider common lines between A and B (both lines are attractive). Similarly, between B and D both Line 3 and Line 4 are attractive, resulting on a total travel time of 11.5 minutes (2.5 min. waiting + 9 min. in-vehicle).

In our example transit network (Figure 5-1) there are 4 different routes between the origin and destination stations. The Minimum Route strategy does not guarantee that a line will not be boarded two consecutive times along the way. For example, Line 2 could be boarded between O and A, to then board the first line between Line 2 (for a second time) and Line 3. The four routes and their associated times are described below.

Route 1:     **O – Line 1 – D**

In-vehicle time:     25 min

Waiting time: 6 min

Total time:             31 min

Route 2:     **O – Line 2 – A – Line 3 – D**

In-vehicle time:     7 min + 8 min = 15 min

Waiting time:        6 min + 15 min = 21 min

Total time:             36 min

Route 3:     **O – Line 2 – A – Line 2 / Line 3 – B – Line 3 / Line 4 – D**

In-vehicle time:     7 min + 5.43 min + 9 min = 21.43 min

Waiting time:        6 min + 4.29 min + 2.5 min = 12.79 min

Total time:             34.22 min

Route 4:	<b>O – Line 2 – B – Line 3 / Line 4 – D</b>
In-vehicle time:	13 min + 9 min = 22 min
Waiting time:	6 min + 2.5 min = 8.5 min
Total time:	30.5 min

It can be seen that Route 4 has the least total travel time, so it would be selected as the Minimum Route. Its total travel time (30.5 minutes) is slightly shorter than the total travel time of the Minimum Itinerary (31 minutes). Based on the Minimum Routes strategy, all transit travellers will board Line 2 between the origin and B, and then will transfer to the first line between Line 3 and Line 4 to reach the destination. Since Line 3's frequency is five times larger than Line 4's frequency, we expect five times as many passengers in Line 3 than in Line 4 on this segment. In this case Line 1 would not be used.

### 5.2.3 Minimum Hyper-Routes

This route choice strategy, also called Optimal Strategies, is the one that guarantees the least total travel time between an origin-destination pair. Proposed by Spiess and Florian (1989), extends the concept of common lines proposed by Chriqui and Robillard (1975) to paths that do not follow the same sequence of transfer stations between the origin and destination. For example, on the transit network of Figure 5-1 an individual could choose at the origin the first line that arrives between Line 1 and Line 2, even though both lines take the traveller to different stations. This way, hyper-paths (Nguyen and Pallottino, 1988) are generated, which differ from the alternatives obtained under the Minimum Routes approach. The probability of boarding each line that is part of a hyper-path strategy is proportional to its frequency.

On our example transit network (Figure 5-1) there are 29 different strategies when travelling between the origin and destinations stations; to be brief, we only present the Minimum Hyper-Route (the one that transit travellers behaving according to this strategy would choose). Unlike the Minimum Routes strategy, the Minimum Hyper-Routes

approach guarantees that a given line will not be boarded two consecutive times when travelling.

Minimum Hyper-Route:  $\{ \mathbf{O} - \text{Line 1} - \mathbf{D} \} / \{ \mathbf{O} - \text{Line 2} - \mathbf{B} - \text{Line 3} / \text{Line 4} - \mathbf{D} \}$

$$\text{In-vehicle time: } \frac{25 \cdot 1/6 + 22 \cdot 1/6}{(1/6 + 1/6)} = 23.5 \text{ min}$$

$$\text{Waiting Time: } \frac{1 + 2.5 \cdot 1/6}{(1/6 + 1/6)} = 4.25 \text{ min}$$

$$\text{Total time: } 27.75 \text{ min}$$

The Minimum Hyper-Route consists on boarding at the origin the first line between Line 1 and Line 2: if Line 1 is boarded, it takes the transit traveller to the destination; if Line 2 is boarded, it takes the transit traveller to station B. At station B the first line between Line 3 and Line 4 is boarded to the destination. It is interesting to note that in this case the Minimum Hyper-Route is composed of the Minimum Itinerary (Itinerary 1 presented on Section 5.2.1 and Route 1 presented on Section 5.2.2) and the Minimum Route (Route 4 presented on Section 5.2.2).

The total travel time of the Minimum Hyper-Route (27.75 minutes) is shorter than the total travel time of the Minimum Route (30.5 minutes). In this case, based on the Minimum Hyper-Routes approach, all lines will be used in at least one link when travelling between the origin and destination stations.

#### 5.2.4 Route Choice Strategies and Uncertainty

Based on the application of the three route choice strategies to the example shown in Figure 5-1, Table V-2 summarizes the most relevant results to be analysed. As expected, when transit travellers increase the complexity of their route choice strategies, better total travel times can be achieved. The main difference in the application of different route choice strategies lies on the resulting flows in the line links. For example, while according

to the Minimum Itineraries approach all trips are made on Line 1, according to the Minimum Routes approach Line 1 would not be used (even though the difference on the total travel time is merely 30 seconds). This is a general problem of all-or-nothing assignment methods, such as the three route choice strategies described.

Table V-2: Results of the Route Choice Strategies Example

<b>Result</b>	<b>Minimum Itineraries</b>	<b>Minimum Routes</b>	<b>Minimum Strategies</b>
Flow on Line 1 O-D	100 %	0 %	50 %
Flow on Line 2 O-A	0 %	100 %	50 %
Flow on Line 2 A-B	0 %	100 %	50 %
Flow on Line 3 A-B	0 %	0 %	0 %
Flow on Line 3 B-D	0 %	16.7 %	8.3 %
Flow on Line 4 B-D	0 %	83.3 %	41.7 %
In-vehicle Time	25 min	22 min	23.5 min
Waiting Time	6 min	8.5 min	4.25 min
<b>Total Time</b>	<b>31 min</b>	<b>30.5 min</b>	<b>27.75 min</b>

Looking at the significant differences in terms of trip flows, it is clear that following one route choice strategy or another when modelling transit traveller's behaviour is a sensitive and relevant issue. In the literature it is usually assumed that all transit travellers, being rational and wanting to minimize their total travel time, behave according to more complex approaches, like Minimum Routes and Minimum Hyper-Routes. For this reason, methodological improvements related to congestion and capacity (de Cea and Fernández, 1993), randomness on the levels-of-service (Lam *et al.*, 1999), reliability (Shimamoto *et al.*, 2010) or inter-modality (Ziliaskopoulos, and Wardell, 2000) traditionally consist on extensions to the Minimum Routes and Minimum Hyper-Routes approach. Nevertheless, there is not sufficient empiric evidence to support or reject the assumption that all travellers behave according to any particular approach. Even more, it is likely that in any transit system there are individuals that behave according to all three route choice strategies.

Another relevant aspect of the three route choice strategies presented is the fact that they are deterministic. In all three cases it is assumed that transit travellers have perfect information, that they all perceive the attributes identically and that they all behave in the same way. For example, the Minimum Itineraries approach assumes that all individuals will choose Itinerary 1, even when Itinerary 5 takes barely one minute longer (i.e. only 3% more). The resulting flows are, as a result, very unstable: if Line 4 increases its frequency so the total travel time of Itinerary 5 decreases to 30.9 minutes, all transit travellers will change their choice (this way Line 1 will shift from carry all individuals to not being used). The Minimum Routes and Minimum Hyper-Routes approaches have the same problem.

To solve the instability of the deterministic approaches, it is possible to propose stochastic models where each alternative (either an itinerary, a route or a strategy, depending on the approach) has a probability of being chosen by the transit travellers. We propose a random utility approach to model the transit travellers' behaviour, where each individual  $q$  chooses an alternative  $i$  from a choice set  $A(q)$ , in order to achieve the maximum utility level  $U_{iq}$ . It is also assumed that the modeller, who is only an observer without perfect information regarding the decision-making process, is only capable of observe a representative systematic utility level  $V_{iq}$ . This way, it is necessary to consider error components  $\varepsilon_{iq}$  on each alternative (McFadden, 1974), typically defined according to (5.1).

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad (5.1)$$

The systematic utility  $V_{iq}$  is a function of the different attributes  $X_{ikq}$  related to the alternatives and the individuals (e.g. travel time, transfers, and socio-economic characteristics). Generally, it is assumed that  $V_{iq}$  is a linear function of the attributes, as shown in (5.2), where  $\theta_k$  are parameters that must be estimated.

$$V_{iq} = \sum_k \theta_k \cdot X_{ikq} \quad (5.2)$$

To characterize the travellers' decisions, it is necessary to define binary variables  $d_{iq}$  according to (5.3). These binary variables capture the actual decisions made by the transit travellers. According to the random utility approach, they should choose the alternative with the highest individual utility level.

$$d_{iq} = \begin{cases} 1 & \text{if } U_{iq} \geq U_{jq} \quad , \quad \forall j \in A(q) \\ 0 & \text{in other case} \end{cases} \quad (5.3)$$

If we assume that the error components  $\varepsilon_{iq}$  are distributed i.i.d. Gumbel, a Multinomial Logit (MNL) model is obtained, for which it is possible to obtain an analytical expression of the choice probabilities  $P_{iq}$ , according to (5.4). Based on the attributes  $X_{ikq}$  and the choices  $d_{iq}$ , the parameters  $\theta_{ik}$  can be estimated with Maximum Likelihood.

$$P_{iq} = \frac{\exp(V_{iq})}{\sum_{j \in A(q)} \exp(V_{jq})} \quad (5.4)$$

The assumption of independence between alternatives of the MNL model can be unrealistic, as in transit networks the different lines can have overlapping links or the routes can have overlapping legs. This topological overlapping generates spatial correlation among the alternatives. To address this issue, different models (mainly extensions of the MNL model) have been proposed in the literature (Prato, 2009). The usage and analysis of these models is beyond the scope of this study.

As an example, let us consider the application of the MNL model shown in (5.5) to the route choice process under the Minimum Itineraries approach. The resulting choice probabilities of the five itineraries are: Itinerary 1: 58.1%, Itinerary 2: 2.8%, Itinerary 3: 0.1%, Itinerary 4: 1.8%, and Itinerary 5: 35.3%. These probabilities assume that the alternatives are independent, even when some itineraries share links. This solution is stable

to changes in the levels-of-service and can possibly reproduce in a more accurate way the individuals' behaviour (by adjusting the model's parameters).

$$P_i = \frac{\exp(-0.5 \cdot T_{In-vehicle\ i} - 0.5 \cdot T_{Wait\ i})}{\sum_{j \in A(q)} \exp(-0.5 \cdot T_{In-vehicle\ j} - 0.5 \cdot T_{Wait\ j})} \quad (5.5)$$

### 5.3 Transit Travel Survey

To identify and analyse the route choice strategy of transit travellers in Santiago - Chile, on June 2011 we conducted travel surveys on Plaza de Maipú, one of Santiago's most visited transit hub. The data gathered with the survey (demand information) was complemented with information regarding the levels-of-service (supply information) provided by the transit authorities. The destinations considered for the survey comprehend the central and eastern parts of the city (communes of Estación Central, Santiago, Providencia, Las Condes, Vitacura and Lo Barnechea). Figure 5-2 presents the geographic information of the survey: the dot represents the origin of the trips and the north-east zone represents the destinations.

The main objective of the survey was to gather massive information regarding mode and route choices within Transantiago, Santiago's transit system. It is important to consider that Transantiago is an integrated system, on which the mode and route choices can be made simultaneously. This way, multi-modal routes arise. Additionally, information regarding the socio-economic characteristics of the travellers was obtained.

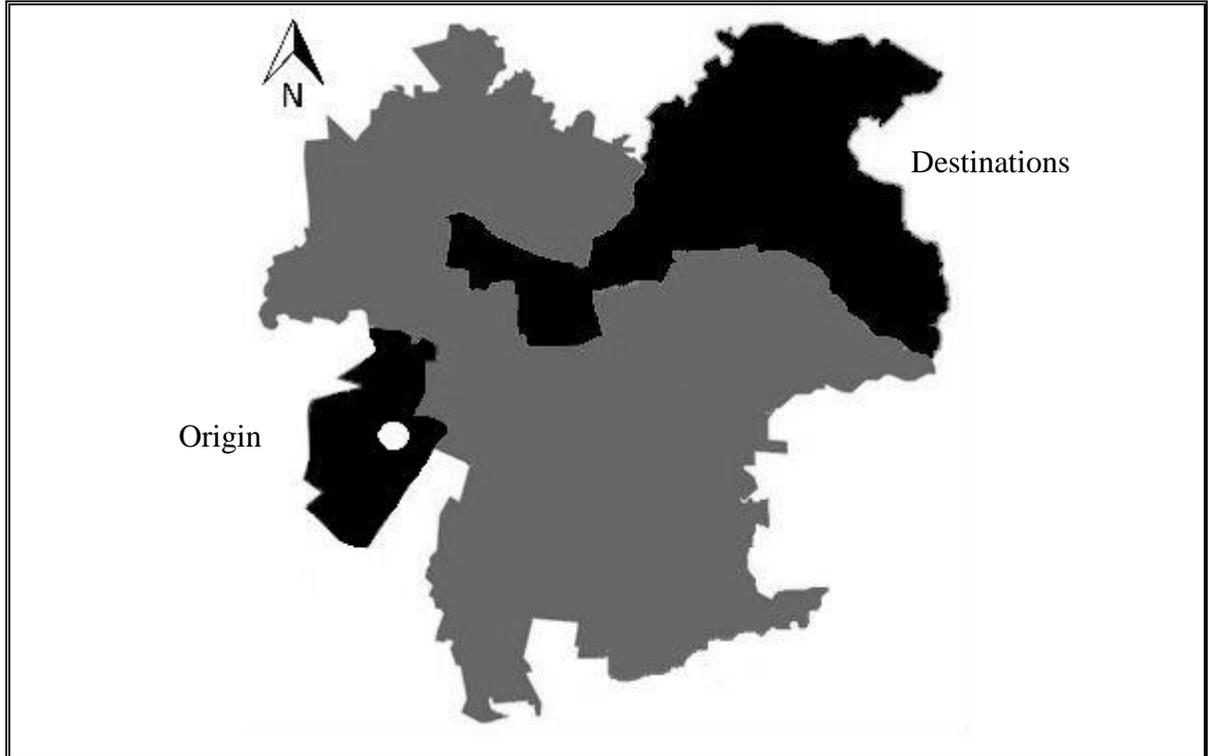


Figure 5-2: Geographic Information of the Survey

The selected origin, Plaza de Maipú, is an interesting node of Santiago's transit network as 1% of all transit trips made on the morning (approximately 30,000 trips between 7:00 hrs and 12:00 hrs) start, transfer or end at this point. Plaza de Maipú is also an interesting node in terms of available services, as it provides access to the Santiago Metro Network, two feeder (local) lines, eight trunk lines and two express lines. This way, the survey gathered information on all modal choices within Santiago's transit network, from 1,892 individuals of diverse socio-economic characteristics.

When conducting the survey, a special emphasis was placed on obtaining information regarding the route choice strategies of the individuals (i.e. if they behaved accordingly to the Minimum Itineraries, Minimum Routes or Minimum Hyper-Routes approach). As expected, all three route choice strategies were reported by the transit travellers. Although 51% of the individuals had only one travel alternative on each travel leg (mainly, those using Metro, where no lines overlap), the remaining 49% of the individuals had the

possibility of choosing their choice strategy. Among this 49% of the individuals, 67% followed Minimum Itineraries, 29% followed Minimum Routes and 4% followed Minimum Hyper-Routes. It is clear that assuming that all individuals behave according to the Minimum Hyper-Routes approach (as it is usually done in the literature) would lead to serious problems in this case. Even assuming that all transit users follow the Minimum Itineraries or Minimum Routes would result on incorrect results.

Given the low proportion of transit travellers following the Minimum Hyper-Routes approach, we will centre the analysis on understanding the differences between the 67% of individuals that do not consider common lines and the 33% of individuals that does consider common lines. It is necessary to take into account that the different travel alternatives between the origin-destination pairs tend to share the infrastructure and follow the same stopping outline, so the in-vehicle times of the different alternatives tend to be similar. This way, basically all of the available lines would be attractive and would belong to the set of common lines. Therefore, it is difficult to understand why 67% of the individuals do not consider common lines, as they are overlooking a way of reducing their waiting times (and therefore their total travel times).

The 67% of the individuals that do not consider common lines tend to be travellers that are not familiarized with the transit system or with the particular trip they are making (they either make the trip occasionally, or have not make the trip before at all) and therefore do not know other transit lines aside from the chosen one. In general, the remaining 33% that considers common lines tends to make the trip on a weekly basis. This way, the level of knowledge of the transit system and its alternatives are relevant.

Age is also a relevant factor: the 67% that does not consider common lines tends to be older than the reminding 33%. Among this 33% of the individuals tend to be those younger than 30 years old, possibly because they are more familiarized with Santiago's current transit system (which was completely redesigned in 2007), while those older than 30 years old might are more likely to try to replicate the path they followed before the redesign of the system, instead of making a proper use of the current transit lines.

The purpose of the trip is also a relevant factor. Among the 33% of the individuals that consider common lines tend to be those travelling to work or study. These individuals have temporal restrictions (i.e. they have to be at the destination at a strict time) and therefore probably analyse and optimize their travel choices. The reduction of the waiting times associated to considering common lines is valuable for these individuals. On the remaining 67% the presence of leisure travellers that do not have temporal restrictions is higher. For them, the waiting time reductions are less valuable.

Finally, income plays a relevant role between the individuals that consider common lines and those that do not. The 33% that considers common lines has a higher average income than the 67% that does not consider common lines. Although it is possible to identify individuals with different income levels on both groups, those with monthly incomes below US\$600 tend to be in the 67% that does not consider common lines, those with monthly incomes between US\$600 and US\$1,200 distribute themselves between both groups, and those with monthly incomes above US\$1,200 tend to be in the 33% that considers common lines. This can be seen as an indirect effect of the level of education of the individuals: as their income increases, higher education levels are achieved, and therefore their analytical capacity of analysing the different travel alternative (identifying a set of common lines to reduce the total travel time) also increases.

#### **5.4 Modelling Route Choice Strategies**

Based on the analysis conducted on Section 5.3 and the relationship between individuals' socio-economic characteristics and the route choice strategies they follow, to appropriately model the decisions of transit travellers in Santiago we propose treating route choice strategies as an endogenous socio-economic characteristic of the individuals (which can be modelled based on their characteristics). Once the strategy is modelled, we can analyse and study the factor that the travellers take into account when choosing travel routes within the public transport system.

### 5.4.1 Modelling Travellers' Strategies

To model the route choice strategy (reduced in this case to considering or not common lines) we use a binary MNL model, using as explanatory variables the socio-economic characteristics of the individuals. The estimated results for modelling the probability of considering common lines are presented in Table V-3. It can be seen that all the considered socio-economic characteristics are statistically significant when understanding the transit travellers' route choice strategies.

The signs of the parameters are consistent with the analysis conducted on Section 5.3. The probability of considering common lines is greater for those transit travellers that: (i) make the particular trip more frequently, (ii) have a higher income level, and (iii) are younger. The purpose of the trip was omitted from the model, as it is highly correlated with the frequency of the trip. With the estimated parameters it is possible to obtain the probabilities of considering common lines for each socio-economic category. These probabilities are shown in Table V-4. It can be seen that there is a wide range of probability values, varying from 11% to 65% depending on the socio-economic characteristics of the individuals.

Table V-3: Results of the Behaviour Strategies Model

<b>Variable</b>	<b>Parameter</b>	<b>t-Test</b>
At Least Weekly Travel	1.322	4.98
At Least Monthly Travel	0.766	3.71
Occasional/First Travel	0.000	Base Category
High Income (over US\$1,200)	0.940	3.22
Medium Income (US\$1,200 to US\$600)	0.327	3.45
Low Income (less than US\$600)	0.000	Base Category
Young Age (less than 30 years old)	0.399	2.90
Adult Age (more than 30 years old)	0.000	Base Category
Constant	- 2.051	- 5.76

Table V-4: Probabilities of Considering Common Lines

<b>Socio-Economic Category</b>		<b>Young Age</b>	<b>Adult Age</b>
<b>At Least Weekly Travel</b>	<b>High Income</b>	65 %	55 %
<b>At Least Weekly Travel</b>	<b>Medium Income</b>	50 %	40 %
<b>At Least Weekly Travel</b>	<b>Low Income</b>	42 %	33 %
<b>At Least Monthly Travel</b>	<b>High Income</b>	51 %	41 %
<b>At Least Monthly Travel</b>	<b>Medium Income</b>	36 %	28 %
<b>At Least Monthly Travel</b>	<b>Low Income</b>	29 %	22 %
<b>Occasional/First Travel</b>	<b>High Income</b>	33 %	25 %
<b>Occasional/First Travel</b>	<b>Medium Income</b>	21 %	15 %
<b>Occasional/First Travel</b>	<b>Low Income</b>	16 %	11 %

#### 5.4.2 Modelling Travellers' Route Choice

With the route choice data gathered in the survey, transit mode and route choice models were estimated. Based on GPS data provided by the Santiago transit authorities, in-vehicle times and waiting times were obtained for each alternative (both considering and not considering common lines). Additionally, the fare, the walking time (when transferring) and the number of transfers were considered as explanatory variables. Given the fare scheme in Santiago's transit network, the fare of each mode/route alternative depends exclusively of the usage of Metro in any of the trip legs (using Metro has an additional cost of US\$0.16 in the peak period and US\$0.04 in the off-peak period). The waiting time considers the initial wait at Plaza de Maipú and all successive waits at the transfer stations.

The modelling approach is sequential: each individual in our database has a probability of considering or not common lines, and then a conditional probability of choosing a given alternative depending on the particular strategy followed. If the individual does not consider common lines, a Minimum Itineraries approach will be followed and the alternatives will be itineraries; if the individual considers common lines, a Minimum Routes approach will be followed and the alternatives will be routes. Given the low proportion of transit travellers following a Minimum Hyper-Routes approach, this route choice strategy was omitted from the analysis to avoid biasing the results.

We consider a linear systematic utility level  $V_{iq}$  for each alternative  $i$ , composed of the attributes mentioned above. The parameters of the model distinguish two different classes of individuals: those who consider common lines, and those who do not. Although some other factors may influence the decisions of the transit travellers (Raveau *et al.*, 2011), the specification we considered for the systematic utility is a robust approximation of the general behaviour of Santiago's transit users.

The result of the estimation of a MNL model for mode/route choice, considering the five explanatory variables mentioned, are presented in Table V-5. It can be seen that all variables have the expected sign (negative, as they represent a disutility) and are statistically significant at 95% confidence. Differences arise between the parameters for individuals that consider common lines and the parameters for individuals that do not consider common lines. These differences support the fact that they are different behaviour patterns among the transit travellers, and that it is necessary to distinguish them.

Table V-5: Results of the Mode/Route Choice Model

Variable	Individuals that Consider Common Lines		Individuals that Do Not Consider Common Lines	
	Parameter	<i>t</i> -Test	Parameter	<i>t</i> -Test
Fare (US\$)	- 21.04	- 2.33	- 25.05	- 2.45
In-Vehicle Time (min)	- 0.623	- 2.17	- 0.479	- 2.39
Waiting Time (min)	- 1.598	- 4.35	- 1.217	- 3.77
Walking Time (min)	- 1.849	- 2.11	- 1.349	- 2.43
Transfers	- 2.585	- 2.45	- 1.569	- 1.98

Based on the obtained parameters, it is possible to calculate monetary valuations and marginal rates of substitution for the time and transfers variables (Table V-6). It can be seen that individuals that consider common lines have higher monetary valuations than the individuals that do not consider common lines (this is possibly because they have a higher income, and therefore a higher value of time). In particular, for individuals that consider common lines transfers are twice as valuable. Regarding the temporal marginal rates of substitution, they do not vary much between both types of individuals. This way, although

the monetary valuations are significantly different, the ratio between them remains relatively constant. On the other hand, there are differences between the marginal rates of substitution of the transfers regarding the in-vehicle time, between both types of individuals.

Table V-6: Subjective Monetary Valuations and Marginal Rates of Substitution

	<b>Individuals that Consider Common Lines</b>	<b>Individuals that Do Not Consider Common Lines</b>
<b>Variable</b>	<b>Subjective Monetary Valuations</b>	
In-Vehicle Time	US\$ 1.78 per hour	US\$ 1.15 per hour
Waiting Time	US\$ 4.56 per hour	US\$ 2.92 per hour
Walking Time	US\$ 5.28 per hour	US\$ 3.24 per hour
Transfers	US\$ 0.12 per transfer	US\$ 0.06 per transfer
<b>Variable</b>	<b>Marginal Rates of Substitution regarding In-Vehicle Time</b>	
1 min of Waiting Time	2.56 min of in-vehicle time	2.54 min of in-vehicle time
1 min of Walking Time	2.97 min of in-vehicle time	2.82 min of in-vehicle time
1 Transfer	4.15 min of in-vehicle time	3.28 min of in-vehicle time

## 5.5 Conclusions

In this study we analysed the route choice strategies that Santiago transit travellers follow, in order to contrast the traditional modelling assumptions with the actual choices made. Although in the literature is commonly assumed that transit travellers behave according to a Minimum Hyper-Routes approach, our empirical evidence shows that only 4% of the respondents follow that strategy. Additionally, 67% of the individuals do not even consider common lines when travelling. This way, at least for the case of Plaza de Maipú, adopting a single route choice strategy for modelling the behaviour and decisions of the transit travellers seems to be an erroneous approach, especially when that single route choice strategy is Minimum Hyper-Routes.

The analysis conducted shows a strong relationship between the individuals' socio-economic characteristics and the route choice strategies they follow. Based on this result, we propose treating the route choice strategies as an additional socio-economic

characteristic of the transit travellers, depending on some of their other characteristics (in particular, the frequency of the trip, their age, and their income). This way the mode/route choice models obtained are capable of reproducing and characterizing the different route choice strategies present among the individuals.

Finally, by distinguishing between different types of transit travellers (based on whether they consider common lines or not) we are capable of identifying significant differences in their mode/route choice decision-making process. In particular, differences in subjective monetary valuations and marginal rates of substitution may arise. This way, ignoring the behavioural differences in terms of route choice strategies can lead to bias results.

### **Acknowledgments**

This study was financed by Project Fondecyt 1110720: *Planning and analysis of express services for an integrated public transport system* and Project Fondef D10I1049: *Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano*.

### **Publication History**

The basis of this article was submitted in 2013 to *Transportation Research Record*, under the title “Analyzing route choice strategies on transit networks”. The results covered in this article were presented at the *XVI Chilean Conference on Transportation Engineering* (Santiago, 2013), and at the *93rd Annual Meeting of the Transportation Research Board* (Washington, 2014).

### **References**

Chriqui, C., and Robillard, P. (1975). Common bus lines. *Transportation Science*, 9(2), 115-121.

de Cea, J., and Fernández, J.E. (1993). Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science*, 27(2), 133-147.

Lam, W.H.K., Gao, Z.Y., Chan, K.S. and Yang, H. (1999). A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B*, 33(5), 351-368.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers of Econometrics*. Academic Press, New York, 105-142.

Nguyen, S., and Pallottino, S. (1988). Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, 37(2), 176-186.

Prato, C.G. (2009). Route choice modelling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), 65-100.

Raveau, S., Muñoz, J.C. and de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A*, 45(2), 138-147.

Shimamoto, H., Kurauchi, F., Schmöcker, J.D. and Bell, M.G. (2010). Transit assignment model considering the correlation of vehicles' arrival. *89h Annual Meeting of the Transportation Research Board*. Washington D.C.

Spiess, H., and Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B*, Vol. 23(2), 83-102.

Ziliaskopoulos, A., and Wardell, W. (2000). An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research*, 125(3), 486-502.

**6 A PLANNING TOOL FOR PUBLIC TRANSPORT ANALYSIS****Sebastián Raveau**Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile**Juan Carlos Muñoz**Department of Transport Engineering and Logistics  
Pontificia Universidad Católica de Chile**Carlo Giacomo Prato**Department of Transport  
Technical University of Denmark**Abstract**

In this article we present an integrated methodological framework for modelling travel decisions on public transport networks. The modelled decisions correspond to the selection of stops, modes, services and routes in multi-modal public transport systems. The objective of this study is to enhance and expand the traditional framework of modelling these travel decisions by analyzing in detail the behaviour, perceptions and preferences of the travellers. The final result of the application of the proposed methodology consists in the assignment of a trip matrix to a public transport network and hence in the flows for the different services. The methodology is applied to the public transport network of Santiago, Chile. The results present a high goodness-of-fit with relation to the observed flows in the network, capturing on a macroscopic level the travel patterns in the network.

## 6.1 Introduction

Understanding how public transport users make their travel decisions and being able to predict their behaviour is essential in transportation planning. The route choice variables normally included in traditional route choice models limit to some basic service levels attributes of the alternative routes, such as travel time and fare (Ortúzar and Willumsen, 2011). However, other variables, related to both the level of service and the travellers' perceptions, influence the user's route choice process but are generally ignore in traditional modelling. The choice of stop is traditionally incorporated implicitly in the route choice, or simplistically modelled as a minimum-time decision.

Particularly in public transport networks with uncertainty on the waiting times (i.e. frequency based systems without timetables), travellers can reduce their expected total travel time by following different route choice strategies. In literature, it is usual to assume that all travellers are capable of considering high-complexity strategies (which might require developed analytical capacities). Similarly, in literature is usually assumed that all travellers have perfect information regarding the levels-of-service of all available alternatives. As expected, these assumptions might not be true for a considerable proportion of the travellers.

The objective of this study is to enhance and expand the traditional framework of modelling these travel decisions by analyzing in detail the behaviour, perceptions and preferences of the travellers. Based on the results of the study, two planning tools are developed: (i) a tactical planning tool for authorities, planners and operators, and (ii) a trip planning tool for travellers.

The reminder of the study is organized as follows. In Section 6.2 we present the methodological framework for analysing and modelling travel demand in public transport systems; in Section 6.3 we present the application of the proposed methodology to the

public transport system of Santiago, Chile; and finally in Section 6.4 we present the main conclusions of the study.

## **6.2 Methodological Framework**

This study presents an integrated methodological framework for modelling travel decisions on public transport networks. The modelled decisions correspond to the selection of stops, modes, services and routes in public transport systems. The final result of the application of the proposed methodology consists in the assignment of a trip matrix to a public transport network and hence in the flows for the different services.

The general structure of the methodology is as follows, given a trip matrix of origin-destination pairs. For each origin zone, a set of attractive stops to start the trip (regardless of the public transport modes they serve) is generated. For each of these stops, a set of attractive routes to the destination zone is generated. For each stop-route alternative (i.e. a complete travel alternative within the public transport network), the choice probabilities are calculated with a stochastic model based on their attributes (Dial, 1971). Based on the choice probabilities, the trips are assigned to the network.

To model the decisions made by the travellers, we follow a sequential approach. First, we model their choice strategies through their socio-economic characteristics. Then we model the successive choice of stop and route, from the origin zone to the destination zone. Finally, based on the choice criteria, we can assign the trip matrix to the public transport network.

### **6.2.1 Modelling Travel Strategies**

A novelty of the proposed methodological framework is the modelling of different route choice strategies, traditionally used in the literature. These strategies deal with the travellers' consideration of a set of common lines when choosing how to travel: the

travellers may not choose a single service, but may board the first service from a set of lines (Chriqui and Robillard, 1975; Spiess and Florian, 1989). The traditional assumption is to assume that all travellers behave in the same way (i.e. all/none of them consider common lines), while in our proposed framework we model users behaviour according to each of both strategies (Raveau and Muñoz, 2014).

We segment the population into two classes: whether they consider common lines or not. The distinction is made based on their individual socio-economic characteristics (such as income level, age, familiarity with the system, and gender). This way, different travellers' behaviour can be accommodated, improving the assignment results.

### **6.2.2 Modelling Stop Choice**

In the first choice step, we model the choices of public transport stops (i.e., the access from the origin zone to the public transport system). The set of attractive stops for a given origin zone does not depend on the destination. This means that the potential alternative stops are the same for any trip originated in a zone (the accessibility of the network is independent to the travel patterns). Their level of attractiveness (and their consequent probability of being chosen), on the other hand, will depend on the particular destination.

The definition of the set of attractive stops for a given origin zone is comprised of three criteria: (i) all the stops within the zone are attractive, (ii) all the additional stops within a certain distance from the zone's centroid are attractive (such distance can vary depending on the particular public transport mode, e.g. people might be willing to walk more to a metro station than to a bus stop), and (iii) all the additional stops within a certain distance from the zone's borders are attractive. This procedure is summarized in Figure 6-1. Depending on the size of the zones and the sizes of the "attractive distance" (either from the centroid or the borders), many stops might be attractive for trips that begin in different zones. Both attractive distances must be calibrated.

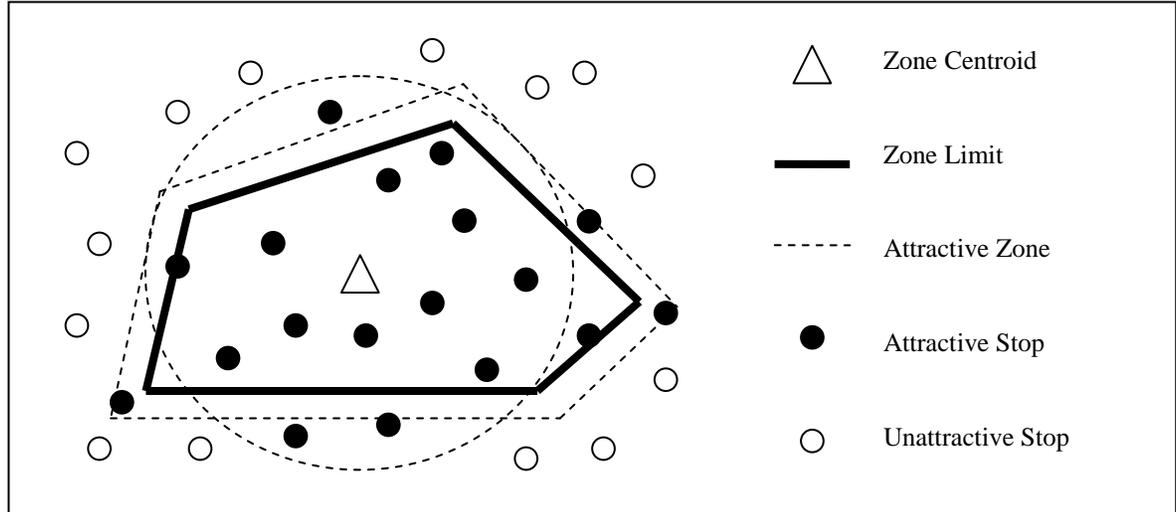


Figure 6-1: Definition of Attractive Stops

Among the attributes related to the particular stops, that will determine their probability of being chosen, the proposed methodology considers; (i) walking time to the stop, (ii) waiting time at the stop (dependent on frequency, queuing and crowding), (iii) types of services (dependent on the presence of buses, metro, trams, etc.), (iv) number of services, (v) transfers, (vi) reliability of the services, (vii) safety, and (viii) socio-economic characteristics from the travellers (such as gender or age).

The remaining element that composes the attractiveness of the stop is an aggregate measure of the level of service from the stop to the destination, which depends on the characteristics of the alternative route from the stop to the destination. This aggregate measure is defined as a Log-Sum type (Train, 2009), and is calculated from the route choice modelling.

### 6.2.3 Modelling Route Choice

In the next step, we model the choices of public transport routes from the selected stop to the destination zone. On integrated systems, these routes could be composed of different modes and services. For this purpose, a set of attractive routes is constructed based on their

generalized level-of-service, which is composed by attributes such as fare, different time components, reliability, transferring experience, vehicle crowding, and network topology (Raveau *et al.*, 2011).

For every combination of origin stop and destination zone, a reasonable number of alternative routes are generated based on “similar performance”. The reasonable number of routes to be generated depends on the particular trip combination. Two route sections have “similar performance” if they satisfy these three criteria: (i) they physically overlap over a certain percentage of their length, (ii) they have comparable generalized costs (which does not excessively exceed the minimum cost alternative), and (iii) they have comparable specific attributes (i.e., frequency, travel time, fare, crowding).

To generate the alternative routes, the link penalty algorithm (De la Barra *et al.* 1993) is applied. The criterion to generate the routes is a generalized cost function that is calibrated through model estimation from the available survey data. Once the generalized cost function is estimated with the parameters necessary to account for all the elements characterizing the routes and hence the choice of stop, it is passed to the choice set generation technique. The set of attractive routes is also constructed taking into account the overlapping of the alternatives, in order to avoid high correlation and similarity between the obtained routes. It should be noted that the generalized cost function is consistent in the choice set generation and the assignment.

#### **6.2.4 Travel Assignment**

Given the generalized cost function and the generated alternative routes, it is possible to calculate the generalized cost of the routes and the attractiveness of stops. These measures allow calculating the probability of choosing a stop within the set of attractive stops, and the probability of choosing a route conditional on the choice of the stop. The probability of choosing a specific route is consequently the product of the two probabilities.

The model for the selection of the stop may be formulated as a Multinomial Logit model (McFadden, 1974) considering the attractiveness of each stop as a utility. The model for the selection of the route conditional on the stop choice may be formulated as a Path Size Logit model (Cascetta *et al.*, 1996) considering the route cost of the routes as a disutility and the similarity across the alternatives.

As some of the relevant attributes such as vehicle crowding and other variables related to comfort depend on the travel decisions of the individuals, the methodological framework results in a fixed-point problem. The application method is iterative until the equilibrium is found. It can be proved that the iterative process converges to a unique equilibrium.

### **6.3 Application to a Real Network**

The methodology is applied to the public transport network of Santiago, Chile (6 million inhabitants). In Santiago, over 4 million trips are made daily on the public transport modes. The public transport system (Transantiago) consists of 191 feeder (local) bus lines, 118 trunk bus lines, and 5 metro lines. To calibrate the modelling approach, a travel survey was conducted in one of Santiago's main public transport hub. The data gathered with the survey (demand information) was complemented with information regarding the levels-of-service (supply information) provided by the public transport authorities.

The network is modelled with 616 one-way bus lines and 10 one-way metro lines. These lines generate a network with 852,548 line segments (which can be grouped to generate 663,696 route segments when common lines are considered). There are 11,113 bus stops and 108 metro stations (which are modelled through 216 directional stops). The demand is composed by 663,599 trips in the morning peak period (6:30 AM to 8:30 AM) (Munizaga and Palma, 2012) and divided into 779 zones. The zoning was obtained from the Origin-Destination Survey of Santiago.

### 6.3.1 Modelling Travellers' Behaviour

Based on the information obtained in the survey, three socio-economic variables were significant to explain the consideration of common lines: (i) the age (two categories, with a threshold in 30 years old), (ii) the monthly income (three categories, with thresholds in US\$600 and US\$1,200), and (iii) the frequency on which the trip is made (Raveau and Muñoz, 2014). Table VI-1 summarizes the resulting probabilities of considering common lines when travelling, which depend on the individual's characteristics.

Table VI-1: Probabilities of Considering Common Lines

Socio-Economic Category		Young Age	Adult Age
At Least Weekly Travel	High Income	65 %	55 %
At Least Weekly Travel	Medium Income	50 %	40 %
At Least Weekly Travel	Low Income	42 %	33 %
At Least Monthly Travel	High Income	51 %	41 %
At Least Monthly Travel	Medium Income	36 %	28 %
At Least Monthly Travel	Low Income	29 %	22 %
Occasional/First Travel	High Income	33 %	25 %
Occasional/First Travel	Medium Income	21 %	15 %
Occasional/First Travel	Low Income	16 %	11 %

As a first approach, we consider a linear systematic generalized cost for each alternative route, composed of: (i) the fare, (ii) the in-vehicle time, (iii) the waiting time, (iv) the walking time, and (v) the number of transfers. The parameters of the model distinguish the two different classes of individuals: those who consider common lines, and those who do not. Although some other factors may influence the decisions of the public transport travellers (Raveau *et al.*, 2011), the specification we considered for the systematic utility is a robust approximation of the general behaviour of Santiago's public transport users.

The results of the estimation of a Multinomial Logit model, obtained from Raveau and Muñoz (2014), are presented in Table VI-2. It can be seen that differences arise between the parameters for individuals that consider common lines and the parameters for

individuals that do not consider common lines. These differences support the fact that they are different behaviour patterns among the public transport travellers, and that it is necessary to distinguish them. With these parameters, the aggregate measure of the level of service from the stop to the destination can be obtained.

Table VI-2: Mode/Route Choice Model Parameters

	<b>Individuals that Consider Common Lines</b>	<b>Individuals that Do Not Consider Common Lines</b>
<b>Variable</b>	<b>Parameter</b>	<b>Parameter</b>
Fare (US\$)	- 21.04	- 25.05
In-Vehicle Time (min)	- 0.623	- 0.479
Waiting Time (min)	- 1.598	- 1.217
Walking Time (min)	- 1.849	- 1.349
Transfers	- 2.585	- 1.569

To define the set of attractive stops for each zone, the maximum walking distance was calibrated based on the information of the survey. The attractive distances from the centroid are 600 metres for bus stops and 1,000 metres for metro stations; when these distances are considered for the selected zoning, the additional criterion of distance from the borders (described in Section 6.2.2) is not needed. Additionally, transfer links were generated between stops within 400 metres.

### 6.3.2 Tactical Planning Tool

The first outcome from the application of the proposed methodology is a tactical planning tool, aimed for public transport authorities, planners and operators. This tactical planning tool provides better models in terms of behavioural interpretation and fit, when forecasts are made. Properly understanding the decision-making process of the travellers is fundamental to obtain reliable predictions and evaluating tactical changes to the network (such as modifications to the existing lines, the construction of new lines and changes in the operation).

The resulting flows of the application from the methodology are presented in Figure 6-2. The results present a high goodness-of-fit with relation to the observed flows in the network (particularly for the most important services, such as the metro lines and main bus corridors), capturing on a macroscopic level the travel patterns in the network.



Figure 6-2: Assignment Flows for Santiago

To illustrate the goodness-of-fit of the assignment, Figure 6-3 presents the load profiles of metro Line 1 (the most loaded line in the public transport network). The correlation between observed and modelled loads is 98% in the eastbound and 99% in the west bound. Satisfactory results are also obtained in the other metro lines. Given that the metro network is the structural backbone of the system, a proper forecasting of its usage is basic for any tactical planning tool.

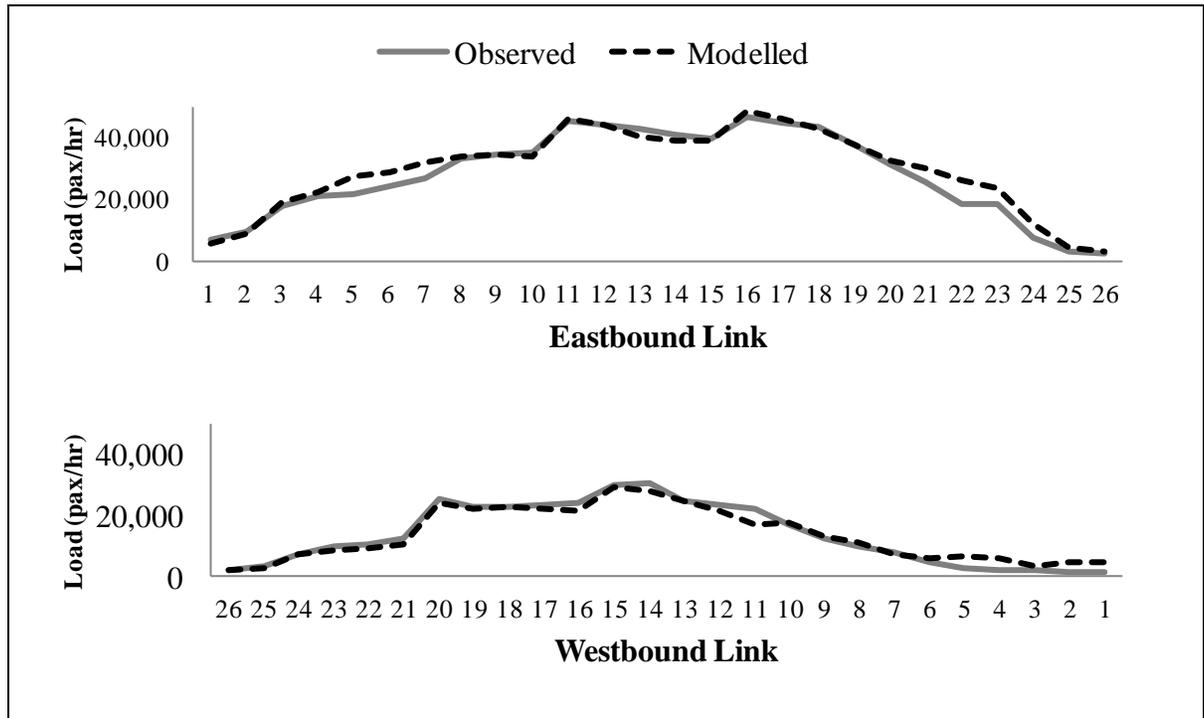


Figure 6-3: Load Profiles on Metro Line 1

To analyse the goodness-of-fit of the results on the bus lines, Figure 6-4 presents the load profiles of ten lines (five trunk lines and five express buses with limited stops) serving approximately 7.5 Kilometres of segregated bus corridor Santa Rosa, in the south of Santiago. The observed and modelled load profiles are compared. The aggregate flows of the ten lines are shown on panel (a), while the specific flows of each line are shown on panels (b) through (f).

Figure 6-4 shows that the model successfully recovers the aggregated flows in the corridor (with a correlation of 98% between observed and modelled data). This way, on a macroscopic level, the tactical planning tool is capable of reproducing the decisions of the travellers. The flow predictions for each of the individual lines have a slightly lower fit to the observed data (which is expectable), with correlations between the observed and modelled data varying between 76% (line 206) and 99% (line 209). The tactical planning tool also reproduces the general travel patterns. Interestingly, there is a tendency to underestimate the flows on the trunk lines, and over estimate the flows on the express

lines. This is probably due to the absence of capacity constraints in the modelling, which in practice tend to activate for the express lines.

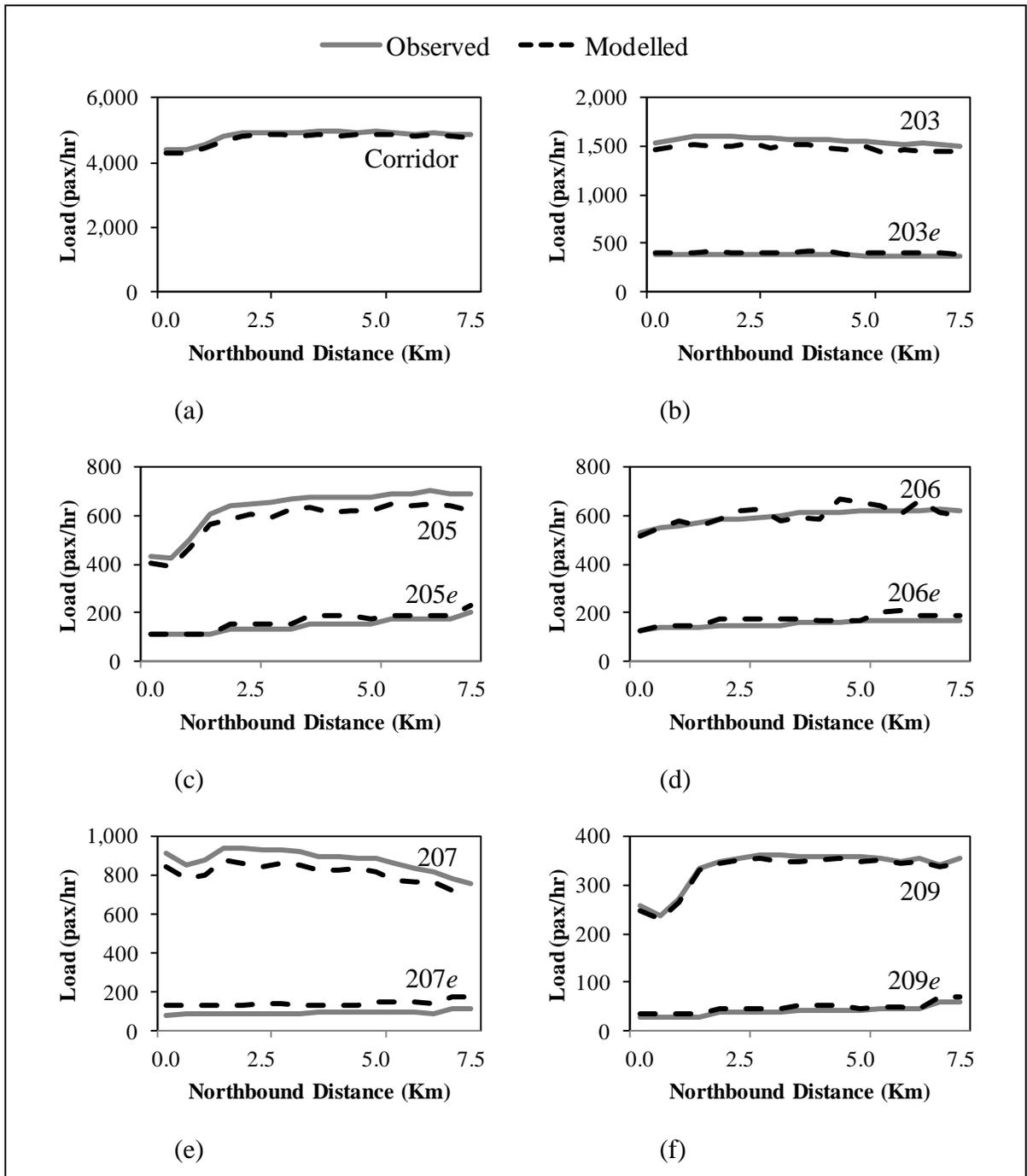


Figure 6-4: Load Profiles on Santa Rosa Corridor

To assess the tactical planning tool's forecasting capability in terms of stop choice, we analyse the decisions observed and predicted in Plaza de Maipú, one of Santiago's main public transport hubs. In Plaza de Maipú there are two trunk lines stops, two express lines stops, and a metro station. Table VI-3 shows the observed and modelled choice percentages for each of the five stop alternatives. It can be seen that the general choice patterns are well predicted, with a mean absolute error of 4.2%. The metro is overestimated, while the express lines are underestimated.

Table VI-3: Stop Choice Probabilities on Plaza de Maipú

<b>Stop</b>	<b>Observed Probability</b>	<b>Modelled Probability</b>
Metro Station	36.0 %	40.6 %
Trunk Lines Stop #1	12.4 %	10.7 %
Trunk Lines Stop #2	13.3 %	19.2 %
Express Line Stop #1	20.4 %	14.1 %
Express Line Stop #2	17.8 %	15.4 %
Plaza de Maipú	100.0 %	100.0 %

Finally, in terms of execution times, the tactical planning tool reached a convergence of 2% in five hours, after computing ten assignment iterations. The convergence is measured in terms of weighted flow difference in successive iterations. For a big network as the Santiago public transport system, the convergence tends to be flat and slowly decreasing. In this case, 20% convergence is reached in two iterations and 5% convergence is reached in five iterations. Additional iterations (over the ten iterations considered) have little impact on the convergence.

### **6.3.3 Trip Planning Tool**

The second outcome from the application of the proposed methodology is a trip planning tool, aimed for the system users. This trip planning tool looks to offer better alternatives for the travellers, adjusting the proposed alternative to the individuals' behaviour and preferences. Traditional public transport trip planners tend to offer the user single-factor optimizations, such as suggesting the route of minimum time or the route with least

transfers. Nevertheless, travellers consider those factors and many others (all at the same time) when deciding how to travel (Raveau *et al.*, 2011).

Based on the proposed methodological framework, and the tactical planning tool implemented in Section 6.3.2, we have implemented a trip planning tool that considers a more detailed representation of the behaviour of the individuals. This way, the trip planning tool can also offer the users the route that represents the least generalized cost (composed by many attributes), instead of only minimizing single attributes.

Figure 6-5 shows an application example of the trip planning tool. An additional advantage of this trip planning tool, in comparison with traditional trip planning tools, is the explicit consideration of common lines. This can be seen in lower left corner, where three different lines are offered for the first trip leg. With this information, the user can board the first line that passes through the stop, instead of waiting for a particular line (and therefore reduce the waiting time).

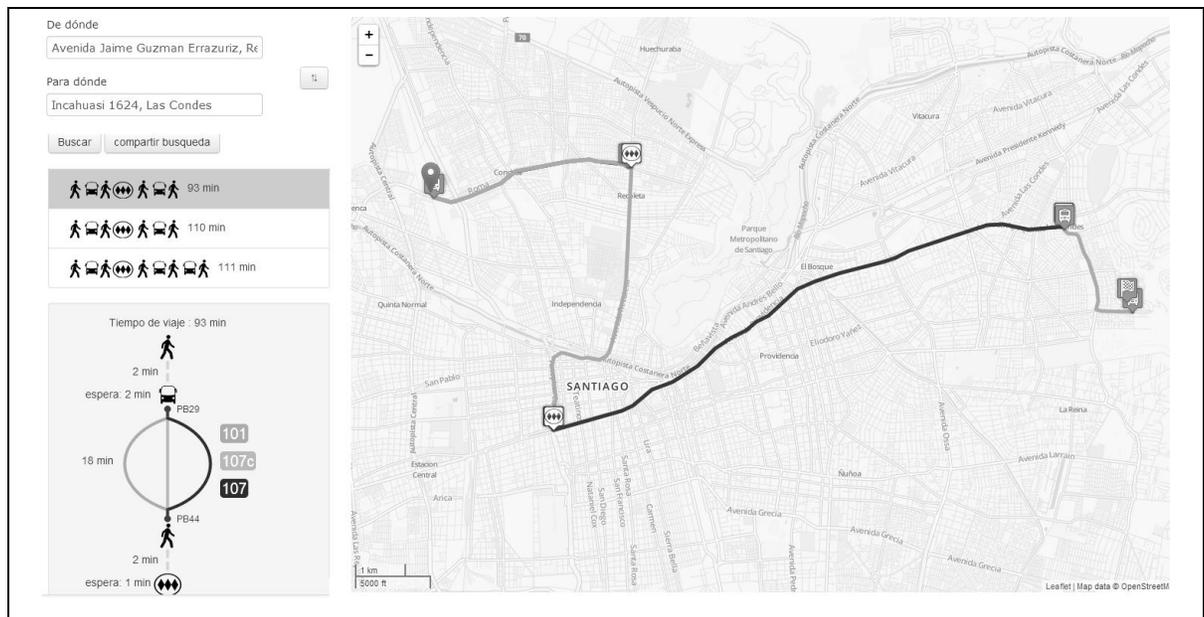


Figure 6-5: Trip Planning Tool Interface

This trip planning tool also contemplates active interaction from the travellers. Users can sign for a follow-up stage after they make the trip, and provide feedback regarding their experienced level-of-service. This information can be very helpful in order to adjust and correct the parameters of the model (both for the tactical planning tool and the trip planning tool).

#### **6.4 Conclusions**

Understanding public transport users' preferences and decision-making process is an essential step in transportation planning, in order to correctly predict their travel decisions and the resulting flows on the public transport network. The objective of the proposed modelling framework is to enhance the traditional behavioural considerations. The results present a satisfactory goodness-of-fit in terms of macroscopic predicted flows, when compared with the actual decisions of the travellers.

The proposed methodology for modelling travel decisions on public transport networks is applied to develop a tactical planning tool (oriented for authorities, planners and operators) and a trip planning tool (oriented for users). This way, the potential advantages of the proposed methodology can be extended to make actual improvements and contributions on a given public transport system, instead of simply being a theoretical contribution.

An important aspect on the application of the proposed methodology, through either of the planning tools, is the interaction with the user. As mentioned in Section 6.3.3, the trip planning tool contemplates a follow-up stage after the user makes the trip, in order to obtain valuable feedback that can be used to enhance the performance of both tools.

It is important to take into account that the application of the proposed methodology to the tactical planning tool and the trip planning tool is designed to be in constant improvements. The modelling approach is susceptible to be adjusted, either in terms of specification (e.g. changing the models' formulation) or composition (e.g. changing the intervening variables

or their parameters). The models are yet to be calibrated for a non-peak period, and there is a necessity to compare the proposed methodology's performance with other planning tools available.

### **Acknowledgments**

This research was supported by FONDEF D10I1049 “Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano”. The authors gratefully acknowledge the research support provided by the Centre for Sustainable Urban Development (CEDEUS) and the Across Latitudes and Cultures - Bus Rapid Transit (ALC-BRT) Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF). The author would like to thank Alexandra Soto and Alfonso Salgado for their contribution to the implementation of both planning tools, and Carlos Melo and Felipe González for their contribution to the methodological framework.

### **Publication History**

The basis of this article was submitted in 2014 to *Transportation Research Part A*, under the title “Modelling mode and route choices on public transport systems”. The results covered in this article are scheduled for the *XVIII Panamerican Conference on Transportation Engineering, Traffic and Logistics* (Santander, 2014).

### **References**

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: Lesort, J.B. (ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Pergamon, Lyon, 697-711.

Chriqui, C., and Robillard, P. (1975). Common bus lines. *Transportation Science*, 9(2), 115-121.

De la Barra, T., Perez, B., and Anez, J. (1993). Multidimensional path search and assignment. *Proceedings of the 21st PTRC Summer Annual Meeting*, Manchester, 307-319.

Dial, R.B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5, 83-113.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers of Econometrics*. Academic Press, New York, 105-142.

Munizaga, M.A. and Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C*, 24(1), 9-18.

Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Edition, John Wiley and Sons, Chichester.

Raveau, S., and Muñoz, J.C. (2014). Analyzing route choice strategies on transit networks. *93rd Annual Meeting of the Transportation Research Board*. Washington D.C.

Raveau, S., Muñoz, J.C. and de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A*, 45(2), 138-147.

Spiess, H., and Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B*, Vol. 23(2), 83-102.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. 2<sup>nd</sup> Edition, Cambridge University Press, Cambridge.

## 7 CONCLUSIONS

Understanding public transport travellers' preferences and decision-making processes is essential in transport planning, in order to correctly predict travel decisions and the resulting flows on public transport networks. The purpose of this thesis has been to understand public transport travellers' behaviour, identifying the relevant factors that are taken into account by them, quantifying the impact that different characteristics of the system have on their decisions, and enhancing the mathematical models used for transport planning.

Mode and route choices (either on private or public transport networks) are traditionally modelled considering only some tangible explanatory variables, such as travel time and fare. These variables, although relevant, fail to accommodate all aspects of traveller's behaviour, as there are other – mainly intangible - factors (such as comfort and reliability) that also affect the decisions made by travellers. These non-traditional factors tend to be ignored in classical models, as they are hard to measure. In this study, two different approaches are followed to enhance the explanatory and forecasting capability of choice models: (i) hybrid discrete choice models with latent variables, and (ii) traditional discrete choice models with a wider specification of preferences.

Following an application of hybrid discrete choice models to understand the choice of travel modes in Santiago, latent variables such as accessibility/comfort, safety and reliability were constructed based on individual characteristics and perception indicators from the travellers. This approach presented superior results in terms of goodness-of-fit when compared to traditional discrete choice models that do not incorporate latent variables.

Analysing route choices on the Santiago Metro network, the thesis proposed a wide specification of travellers' preferences, by considering many factors that are usually ignored when modelling. Five different types of variables were found to be significant

when explaining the decisions made by travellers: (i) three different time components; (ii) transfer-related variables; (iii) occupancy and comfort; (iv) topological perceptions, and (v) socio-demographic information. All these variables contribute to understand travellers' behaviour, and should be considered when modelling. This reassures the idea that public transport users take into account a wide variety of attributes when travelling.

Results show that, when the proposed psychological and behavioural factors are not considered in the models (and therefore their specification is incomplete), there are biases in the estimated parameters. These biases can lead to mistaken conclusions regarding the effect of different variables on the decision-making process, and even lead to errors in the application of the results on public transport planning. In fact, problems were found when marginal rates of substitutions (such as values of time) were estimated. This problem is due to relevant information being omitted from the models. The inclusion of non-traditional factors in the modelling (either through latent variables or through a wider specification of the perceptions and preferences) helps correcting this problem.

This thesis also provides a behavioural comparison between two metro systems: the Santiago Metro and the London Underground. The comparison defined first a common specification for the route choice models of both cities, which is not commonly done in the literature when comparisons are made. The specification considers a wide variety of factors that capture individuals' preferences and perceptions. The comparison allows the quantification and contrast of how the factors influence decisions on both systems. Although the factors might be perceived differently, all of them were significant on both cities. This way, the proposed specification could be used to analyse route choice decisions on other systems.

When analysing route choices in the Santiago Metro and the London Underground systems, a significant finding was the quantification of the impact that schematic maps have on travellers' decisions. Public transport schematic maps are usually distorted in order to accommodate all the information that needs to be presented to the travellers. These distortions can affect the spatial perceptions of the individuals, and therefore influence

their decisions. The impact of schematic maps on travellers' decisions had not been analysed prior to this study. The schematic map distortions can even be exploited by the planners to induce desired route choices and thereby make the best use of the public transport system.

This research also provides an empirical analysis about the route choice strategies that travellers might follow: either board the first line (within a previously selected set of lines) that arrives to the stop, or wait for a specific line. Although the traditional modelling approach assumes that all travellers behave the same way, the empirical data for Santiago shows that different travellers follow different strategies. The analysis conducted shows a strong relationship between the travellers' socio-economic characteristics and the route choice strategies they tend to follow. Based on this result, the route choice strategy was treated as an endogenous socio-economic characteristic of the travellers, which depends of other socio-economic and trip features (in particular, the frequency of the trip, their age, and their income). The resulting mode/route choice models are capable of reproducing the travel decisions and resulting flow patterns in a better way.

Explicitly distinguishing between different types of public transport travellers (based on whether they consider a set of attractive lines or not), allows the identification of significant differences in their preferences and decision-making processes. In particular, there are differences in the subjective monetary valuations and marginal rates of substitution that are obtained for different strategy-followers. This way, ignoring the behavioural differences in terms of route choice strategies can also lead to bias results.

The proposed modelling approach to understand travel decisions on public transport networks was applied to develop two planning tool: (i) a tactical planning tool (oriented for authorities, planners and operators), and (ii) a trip planning tool (oriented for users). Traditional planning tools tend to use over-simplified behavioural assumptions and might result on inaccurate predictions and recommendations. The potential advantages of enhancing the behavioural modelling of travellers' decisions can be extended to improve the planning tools. An important aspect of the application of the proposed models for

public transport planning is the interaction with the user. Valuable feedback can be obtained to enhance the performance of the models and planning tools. It is of essence to never forget that public transport modellers and planners deal with real individuals, whose behaviour is hard to understand and model.

## 8 REFERENCES

Ashok, K., Dillon, W. and Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research*, 39(1), 31-46.

Ben-Akiva, M.E., Bergman, M.J., Daly, A.J. and Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. In J. Volmuller and R. Hamerslag (eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*. VNU Science Press, Utrecht, 299-330.

Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A. (2002). Integration of choice and latent variable models. In H. S. Mahmassani (Ed.), *In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges*. Pergamon, Amsterdam. 431-470.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. and Fetiariison, M. (2009). Estimation of discrete choice models: extending BIOGEME. *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland

Bolduc, D. and Alvarez-Daziano R. (2010). On estimation of hybrid choice models. In S. Hess and A. Daly (eds.), *Choice Modelling: The State-of-the-Art and the State-of-Practice*. Emerald Group Publishing, Bingley.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. John Wiley and Sons, Chichester.

Bovy, P.H.L. and Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. *Transportation*, 32(4), 341-368.

Bovy, P.H.L. and Stern, E. (1990). *Route Choice. Wayfinding in Transport Networks*. Kluwer Academic Publishers, Norwell, MA.

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: Lesort, J.B. (ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Pergamon, Lyon, 697-711.

Cherchi, E. and Cirillo, C. (2008). A modal mixed logit choice model on panel data: accounting for systematic and random heterogeneity in preferences and tastes. *86th Annual Meeting of the Transportation Research Board*. Washington D.C.

Chriqui, C., and Robillard, P. (1975). Common bus lines. *Transportation Science*, 9(2), 115-121.

Conroy-Dalton, R. (2003). The secret is to follow your nose: route path selection and angularity. *Environment and Behavior*, 35(1), 107-131.

de Cea, J., and Fernández, J.E. (1993). Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science*, 27(2), 133-147.

De la Barra, T., Perez, B., and Anez, J. (1993). Multidimensional path search and assignment. *Proceedings of the 21st PTRC Summer Annual Meeting*, Manchester, 307-319.

Dial, R.B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5, 83-113.

- Gaudry, M.J.I., Jara-Díaz, S.R. and Ortúzar, J. de D. (1989). Value of time sensitivity to model specification. *Transportation Research Part B*, 23(2), 151-158.
- Guo, Z. (2011). Mind the map! The impact of transit maps on path choice in public transit. *Transportation Research Part A*, 45(7), 625-639.
- Guo, Z. and Wilson, N.H.M. (2011). Assessing the cost of transfer inconvenience in public transport systems: a case study of the London Underground. *Transportation Research Part A*, 45(2), 91-104.
- Hensher, D.A. (2007). Sustainable public transport systems: moving towards a value for money and network-based approach and away from blind commitment. *Transport Policy*, 14(1), 98-102.
- Hochmair, H. (2009). The influence of map design on route choice from public transportation maps in urban areas. *The Cartographic Journal*, 46(3), 242-256.
- Hunt, J.D. (1990). A logit model of public transport route choice. *ITE Journal*, 60(12), 26-30.
- Jankowski, P., Andrienko, N. and Andrienko, G. (2001). Map-centred exploratory approach to multiple criteria spatial decision making. *International Journal of Geographical Information Science*, 15(2), 101-127.
- Koppelman, F.S. and Hauser, J.R. (1978). Destination choice behaviour for non-grocery-shopping trips. *Transportation Research Record*, 673(1), 157-165.
- Koppelman, F.S. and Pas, E.I. (1980). Travel-choice behaviour: models of perceptions, feelings, preference, and choice. *Transportation Research Record*, 765(1), 26-33.

Lam, W.H.K., Gao, Z.Y., Chan, K.S. and Yang, H. (1999). A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research Part B*, 33(5), 351-368.

Liu, Y., Bunker, J. and Ferreira, L. (2010). Transit users' route-choice modelling in transit assignment: a review. *Transport Reviews*, 30(6), 753-769.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed.), *Frontiers of Econometrics*. Academic Press, New York, 105-142.

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4), 275-297.

Mobidrive (2000). *Mobidrive questionnaires*. *Arbeitsberichte Verkehrs-und Raumplanung* 52. Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau, ETH Zürich (in German).

Montello, D.R. (1991). Spatial orientation and the angularity of urban routes: a field study. *Environment and Behavior*, 23(1), 47-69.

Munizaga, M.A. and Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C*, 24(1), 9-18.

Muñoz, J.C., Ortúzar, J. de D. and Gschwender, A. (2009). Transantiago: the fall and rise of a radical public transport intervention. In W. Saleh and G. Sammer (eds.). *Travel Demand Management and Road User Pricing: Success, Failure and Feasibility*. Ashgate, Farnham.

Nguyen, S., and Pallottino, S. (1988). Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, 37(2), 176-186.

Nicolaidis, G.C. (1975). Quantification of the comfort variable. *Transportation Research*, 9(1), 55-66.

Ortúzar, J. de D. and Willumsen, L.G. (2011). *Modelling Transport*. 4<sup>th</sup> Edition, John Wiley and Sons, Chichester.

Ovenden, M. (2007). *Transit Maps of the World*. Penguin Group, New York.

Papinski, D., Scott, D.M. and Goherty, 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*, 12(4), 347-358.

Prato, C.G. (2009). Route choice modelling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), 65-100.

Prato, C.G., Bekhor, S. and Pronello, C. (2012). Latent variables and route choice behaviour. *Transportation*, 39(2), 299-319.

Prashker, J.N. (1979). Mode choice models with perceived reliability measures. *Transportation Engineering Journal*, 105(3), 251-262.

Prashker J.N. and Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, 24(4), 437-463.

Ramming, M.S. (2001). *Network Knowledge and Route Choice*. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D. and Ortúzar, J. de D. (2010). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record*, 2156(1), 131-139.

Raveau, S., and Muñoz, J.C. (2014). Analyzing route choice strategies on transit networks. *93rd Annual Meeting of the Transportation Research Board*. Washington D.C.

Raveau, S., Muñoz, J.C. and de Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A*, 45(2), 138-147.

Raveau, S., Yáñez, M.F. and Ortúzar, J. de D. (2012). Practical and empirical identifiability of hybrid discrete choice models. *Transportation Research Part B*, 46(10), 1374-1383.

Rundmo, T. and Hale, A.R. (2003). Managers' attitudes towards safety and accident prevention. *Safety Science*, 41(7), 557-574.

Shimamoto, H., Kurauchi, F., Schmöcker, J.D. and Bell, M.G. (2010). Transit assignment model considering the correlation of vehicles' arrival. *89th Annual Meeting of the Transportation Research Board*. Washington D.C.

Spieß, H., and Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B*, Vol. 23(2), 83-102.

Stapleton, D.C. (1978). Analyzing political participation data with a MIMIC model. *Sociological Methodology*, 9(1), 52-74.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. 2<sup>nd</sup> Edition, Cambridge University Press, Cambridge.

Turner, A. (2001). Angular analysis. *Proceedings of the 3rd International Symposium on Space Syntax*, Georgia Institute of Technology, Atlanta, 30.1–30.11.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23(2), 131–138.

Vredin Johansson, M., Heldt, T. and Johansson, P. (2006). The effects of attitudes and personality traits on mode choice. *Transportation Research Part A*, 40(6), 507-52.

Yáñez, M.F., Mansilla, P. and Ortúzar, J. de D. (2010a). The Santiago Panel: measuring the effects of implementing Transantiago. *Transportation*, 37(1), 125-149.

Yáñez, M.F. and Ortúzar, J. de D. (2010). Modelling choice in a changing environment: assessing the shock effects of a new transport system. In S. Hess and A.J. Daly (eds.), *Choice Modelling: the State-of-the-Art and the State-of-Practice*. Emerald, Bingley.

Yáñez, M.F., Raveau, S. and Ortúzar, J. de D. (2010b). Inclusion of latent variables in Mixed Logit models: modelling and forecasting. *Transportation Research Part A*, 44(9), 744-753.

Ziliaskopoulos, A., and Wardell, W. (2000). An intermodal optimum path algorithm for multimodal networks with dynamic arc travel times and switching delays. *European Journal of Operational Research*, 125(3), 486-502.