

# Multimodal Pricing and the Optimal Design of Bus Services: New Elements and Extensions

Alejandro Andrés Tirachini

Thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy in the Business School, University of Sydney, Australia

July 2012



Institute of Transport and Logistics Studies

The University of Sydney Business School

The University of Sydney

NSW 2006 Australia

# Abstract

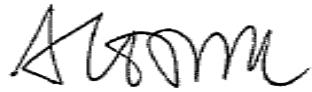
This thesis analyses the pricing and design of urban transport systems; in particular the optimal design and efficient operation of bus services and the pricing of urban transport. Five main topics are addressed: (i) the influence of considering non-motorised travel alternatives (walking and cycling) in the estimation of optimal bus fares, (ii) the choice of a fare collection system and bus boarding policy, (iii) the influence of passengers' crowding on bus operations and optimal supply levels, (iv) the optimal investment in road infrastructure for buses, which is attached to a target bus running speed and (v) the characterisation of bus congestion and its impact on bus operation and service design. Total cost minimisation and social welfare maximisation models are developed, which are complemented by the empirical estimation of bus travel times.

As bus patronage increases, it is efficient to invest money in speeding up boarding and alighting times. Once on-board cash payment has been ruled out, allowing boarding at all doors is more important as a tool to reduce both users and operator costs than technological improvements on fare collection. The consideration of crowding externalities (in respect of both seating and standing) imposes a higher optimal bus fare, and consequently, a reduction of the optimal bus subsidy. Optimal bus frequency is quite sensitive to the assumptions regarding crowding costs, impact of buses on traffic congestion and congestion level in mixed-traffic roads. The existence of a crowding externality implies that buses should have as many seats as possible, up to a minimum area that must be left free of seats.

Bus congestion in the form of queuing delays behind bus stops is estimated using simulation. The delay function depends on the bus frequency, bus size, number of berths and dwell time. Therefore, models that use flow measures (including frequency only or frequency plus traffic flow) as the only explanatory variables for bus congestion are incomplete. Disregarding bus congestion in the design of the service would yield greater frequencies than optimal when congestion is noticeable, i.e. for high demand. Finally, the optimal investment in road infrastructure for buses grows with the logarithm of demand; this result depends on the existence of a positive and linear relationship between investment in infrastructure and desired running speed.

## Statement of Originality

This is to certify that, to the best of my knowledge, the content of this thesis is my own work. This thesis contains no material previously published or written by another person unless due reference to that material is made. This thesis contains no material that has been submitted for the award of any degree or diploma in any university or other institution.



Alejandro Andrés Tirachini

03 July 2012

# Preface

This thesis is the result of the effort of several people and institutions that have taught and supported me in so many ways, throughout my studies in Chile and Australia. Even though I started my PhD in 2009, the roots of this thesis can be traced to 2004 when I was an undergraduate student at the School of Engineering of Universidad de Chile. Curious of what research in transport studies was like, I went to the library in search of master theses written by students of the transport engineering division. The one that caught my eye straightaway was Antonio Gschwender's "Microeconomic Characterisation of Urban Public Transport Operations: a Critical Analysis". This work was my first exposure to the theory of public transport economics and it captivated me. I would later develop my own master thesis on public transport, a process in which I greatly benefited from the mentoring and stimulating guidance of Cristián Cortés and Sergio Jara-Díaz. Then in 2006 I met David Hensher, who travelled to Santiago invited to a conference in our university. David kindly invited me to come to Sydney and visit him at the Institute of Transport and Logistics Studies, a trip that eventually prompted me to pursue my doctoral studies in Australia.

I owe a debt of gratitude to David as my thesis supervisor and a mentor. Aside from all the knowledge I have acquired from him, David made me feel welcome from my very first day at the University of Sydney and has been extremely supportive of the ideas and endeavours I have undertaken along the way; his encouragement has been utterly important to further develop my work and research skills. I am also indebted to my associate supervisor John Rose, whose insightful comments have been influential in the final stages of my thesis work. Three years of work at the Institute have been a great pleasure, in which I have enjoyed the friendship of several colleagues and fellow doctoral candidates, especially Claudine, Patrick, Wu, Chinh, Asif and Lorenzo. Thanks for all the moments.

I am very much indebted to Stephen Rowe, Managing Director of the Busways Group, Australia, who generously agreed to sponsor my work and granted me access to the Busways network in Blacktown, and to the company headquarters in Pymble, Sydney, where I could collect my own data to empirically estimate bus dwell and running times. My

experience in Busways was invaluable to complement the theoretical work I was doing at the university with the empirical modelling of day-to-day bus operations. A special thanks to the staff of the scheduling department who were always happy to answer my many questions, especially Clayton Davidson, Andrew Glass, Nat Dechchavalit and Greg Blackley.

I also need to thank Chile's National Commission for Scientific and Technological Research (CONICYT), for providing me with the scholarship *Beca de Doctorado por Gestión Propia* for overseas PhD studies.

Finally, none of this would have been possible without the perennial support and encouragement of my family. This thesis is dedicated to my parents Elba and Luis, my sister Jasna and my brother Luis Antonio, *con todo mi amor*.

# Contents

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	Research Topics and Questions .....	2
1.1.1	Non-motorised modes.....	4
1.1.2	Fare collection system and boarding policy .....	5
1.1.3	Crowding.....	6
1.1.4	Bus congestion.....	7
1.1.5	The provision of busways .....	7
1.2	Research Approach .....	8
1.2.1	Fare collection system and boarding policy .....	9
1.2.2	Crowding.....	9
1.2.3	Bus congestion.....	9
1.2.4	The provision of busways .....	10
1.2.5	Total cost minimisation model .....	10
1.2.6	Multimodal pricing and bus optimisation .....	11
1.3	Thesis Contributions .....	11
1.3.1	New elements and methodological refinements.....	11
1.3.2	Theoretical and scientific relevance.....	14
1.3.3	Practical relevance.....	14
1.4	Research Scope .....	16
1.5	Thesis Outline.....	17
<b>2</b>	<b>BACKGROUND: MICROECONOMIC MODELLING OF URBAN PUBLIC TRANSPORT OPERATION AND OPTIMAL PRICING .....</b>	<b>20</b>
2.1	Introduction .....	20
2.2	Setting Public Transport Fares: First Best and Second Best Models .....	21
2.2.1	First best pricing .....	22
2.2.2	Second best pricing.....	26
2.2.3	Issues that arise when subsidising public transport.....	29
2.3	Results that Matter .....	30
2.3.1	Peak versus off-peak fares.....	30
2.3.2	Effect of including other externalities beyond congestion .....	32

2.3.3	Dedicated bus lanes.....	33
2.3.4	Interactions with other sectors of the economy.....	35
2.3.5	Other public transport provision structures .....	38
2.3.6	Other relevant aspects .....	40
2.4	Optimal Public Transport Supply .....	41
2.4.1	Basic theory and main results .....	41
2.4.2	Public transport supply and road pricing .....	44
2.4.3	When the capacity constraint is binding.....	44
2.5	Summary.....	46
<b>3</b>	<b>MULTIMODAL TRANSPORT PRICING: THE INFLUENCE OF NON-MOTORISED MODES</b>	<b>47</b>
3.1	Introduction .....	47
3.2	Model Assumptions .....	49
3.3	First Best Pricing.....	52
3.4	Second Best Pricing.....	54
3.5	Extensions: Other External Costs and Collection Costs .....	58
3.6	Summary and Conclusions.....	60
<b>4</b>	<b>FARE COLLECTION SYSTEMS AND BUS BOARDING TIME: OPERATIONAL AND ECONOMIC EFFECTS .....</b>	<b>61</b>
4.1	Introduction .....	61
4.2	Definition of Alternatives.....	63
4.3	Estimation of Boarding and Alighting Times with Alternative Fare Payment Technologies and Boarding Policies .....	65
4.3.1	Background: dwell time models.....	65
4.3.2	Data collection.....	67
4.3.3	Model 1: Blacktown.....	70
4.3.4	Model 2: Inner Sydney.....	71
4.3.5	Model 3: Free CBD shuttle.....	72
4.3.6	Average passenger service times .....	74
4.4	Technology Effect and Door Effect: Non-linearity and Interdependency .....	78

4.5	Technology Effect and Door Effect: Assessment with an Empirical Bus Travel Time Model .....	79
4.5.1	Background: bus travel time models.....	79
4.5.2	Data collection and travel time model estimation.....	81
4.5.3	Percentage of time spent at each stage.....	86
4.5.4	Bus operating speed .....	88
4.5.5	Going cashless or creating busways? .....	92
4.5.6	Estimation of benefits from upgrading the fare collection system: fleet size, travel time, operator and environmental cost savings .....	95
4.6	Conclusions .....	100
<b>5</b>	<b>DETERMINANTS OF BUS CONGESTION AND ITS INCLUSION IN THE ECONOMIC ANALYSIS OF TRANSPORT POLICIES .....</b>	<b>103</b>
5.1	Introduction .....	103
5.2	Bus Congestion in the Literature .....	104
5.3	Estimation of Queuing Delays at Bus Stops.....	107
5.4	The Relationship between Bus Congestion, Fare Collection Technique and Bus Boarding Policy .....	114
5.5	Conclusions .....	117
<b>6</b>	<b>THE EFFECTS OF PASSENGER CROWDING ON PUBLIC TRANSPORT DEMAND AND SUPPLY.....</b>	<b>118</b>
6.1	Introduction .....	118
6.2	Effects of Crowding.....	120
6.2.1	Effect on in-vehicle time.....	120
6.2.2	Effect on waiting time .....	122
6.2.3	Effect on travel time reliability.....	123
6.2.4	Effect on the valuation of travel time savings and route choice .....	124
6.2.5	Impact of crowding externality on optimal supply and fare.....	126
6.3	Estimation of Crowding and Standing Costs.....	128
6.4	Conclusions .....	134



<b>7</b>	<b>BUS CONGESTION, OPTIMAL INFRASTRUCTURE INVESTMENT AND THE CHOICE OF A FARE COLLECTION SYSTEM: AN EXTENDED TOTAL COST MINIMISATION MODEL .....</b>	<b>136</b>
7.1	Introduction .....	136
7.2	Bus Round-trip Time .....	138
7.3	User Cost, Operator Cost and Problem Formulation.....	144
7.4	Results and Analysis.....	150
7.4.1	Assumptions .....	150
7.4.2	Results.....	150
7.4.3	Analysis of other scenarios.....	159
7.5	Summary and Conclusions.....	162
<b>8</b>	<b>OPTIMAL DESIGN OF BUS SERVICES ON AN EXTENDED MULTIMODAL FRAMEWORK .....</b>	<b>165</b>
8.1	Introduction .....	165
8.2	Model Set Up .....	168
8.2.1	Assumptions and definitions .....	168
8.2.2	Demand modelling and crowding .....	170
8.2.3	Travel time, congestion and bus stop delay.....	172
8.2.4	The choice of bus size and internal layout .....	174
8.2.5	Bus operator cost and problem formulation .....	175
8.3	Application .....	177
8.3.1	Physical setting and input parameters.....	177
8.3.2	Base results.....	181
8.3.3	Optimal bus frequency: The trade-off between congestion and crowding..	185
8.3.4	Optimal pricing and modal split .....	189
8.3.5	The case with increased bus-induced congestion.....	192
8.3.6	The relationship between the number of seats and optimal frequency .....	193
8.3.7	The second best scenario .....	194
8.4	Conclusions .....	196

<b>9</b>	<b>CONCLUSIONS.....</b>	<b>198</b>
9.1	Summary.....	198
9.2	Contributions .....	199
9.2.1	Result on research questions .....	199
9.2.2	Methodological contributions.....	202
9.2.3	Further results of practical relevance.....	203
9.3	Caveats of the Research.....	203
9.4	Areas for Further Research.....	205
9.5	Concluding Remarks .....	207
	<b>REFERENCES.....</b>	<b>208</b>
	<b>APPENDICES.....</b>	<b>225</b>
	Appendix A1: First Order Conditions, First Best and Second Best Pricing Models .....	225
	Appendix A2: Estimation of the Queuing Delay Function (Chapter 5).....	227
	Appendix A3: Estimation of Delay at Intersections (Chapter 7).....	228
	Appendix A4: Estimation of Parameters of the Operator Cost Functions .....	229
	Appendix A5: Bus Internal Layout: Passengers Seating and Standing and Constraints for the Determination of the Number of Seats (Chapter 8) .....	233

# List of Figures

Figure 1.1: Relationship among thesis topics.....	4
Figure 1.2: Thesis contributions .....	13
Figure 4.1: Examples of on-board fare verification systems.....	64
Figure 4.2: Load profile routes 440 and 753 .....	75
Figure 4.3: Average passenger service time.....	77
Figure 4.4: Bus network in the study area.....	82
Figure 4.5: Actual number of bus stops as a function of passengers per bus.....	89
Figure 4.6: Bus operating speed as a function of demand and fare payment and boarding policy.....	91
Figure 4.7: Bus operating speed, mixed-traffic versus segregated bus operation .....	94
Figure 5.2: Queuing delay as a function of bus frequency and queuing delay, two-berth bus stops .....	111
Figure 5.3: Bus stop queuing delay in different configurations .....	113
Figure 5.4: Queuing delay for alternative fare payment and boarding policies .....	116
Figure 6.1: Crowding can slow down both alighting and boarding of passengers .....	120
Figure 6.2: Crowding multiplier for passengers seating and standing.....	125
Figure 6.3: Different levels of occupancy in stated preference study .....	129
Figure 6.4: Crowding multiplier as a function of occupancy rate and density of standees.....	133
Figure 6.5: Value of in-vehicle time savings .....	134
Figure 7.1: Total average cost .....	152
Figure 7.2: Average queuing delay .....	152
Figure 7.3: Optimal frequency.....	154
Figure 7.4: Optimal bus running speed .....	156
Figure 7.5: Infrastructure cost per kilometre .....	156
Figure 7.6: Optimal bus stop spacing .....	157
Figure 7.7: Operating speed (running plus detentions) .....	158
Figure 7.8: Fleet size .....	159
Figure 7.9: Optimal frequency, contactless card payment <i>TnB1</i> .....	161
Figure 7.10: Queuing delay, analysis of scenarios with contactless card payment <i>TnB1</i> ....	161
Figure 7.11: Optimal distance between stops, analysis of scenarios with contactless card payment <i>TnB1</i> .....	162

Figure 8.1: Transport corridor diagram .....	169
Figure 8.2: Test corridor, Military Road .....	178
Figure 8.3: Origin-Destination matrix.....	179
Figure 8.4: Optimal frequency.....	186
Figure 8.5: Average speed M1.....	186
Figure 8.6: Optimal seat supply.....	188
Figure 8.7: Optimal seating+standing supply.....	188
Figure 8.8: Optimal bus frequency on shared and dedicated right-of-way.....	189
Figure 8.9: Optimal toll and bus fare.....	190
Figure 8.10: Modal split M2.....	191
Figure 8.11: Modal split per trip length, M2 base case (total demand=19,234) .....	191
Figure 8.12: Optimal bus frequency M1, double equivalency factor for buses.....	192
Figure 8.13: Optimal bus frequency M2, double equivalency factor for buses.....	193
Figure 8.14: Optimal bus frequency for suboptimal numbers of seat.....	194
Figure 8.15: Optimal toll minus bus fare, first best and second best scenarios .....	195
Figure 8.16: Optimal bus frequency, first best and second best scenarios .....	196
Figure A4.1: Infrastructure cost per kilometre versus operating (commercial) speed.....	230

# List of Tables

Table 4.1: Summary of fare payment and bus boarding alternatives.....	65
Table 4.2: Characteristics of the services surveyed .....	68
Table 4.3: Estimation of parameters for dwell time models.....	71
Table 4.4: Boarding and alighting times per passenger .....	74
Table 4.5: Average passenger service time (s/pax).....	77
Table 4.6: Estimation of travel time model.....	85
Table 4.7: Percentage of time spent at each stage .....	87
Table 4.8: Fleet size (FS – buses) and in-service time ratio (STR) .....	97
Table 4.9: Fuel cost savings with respect to on-board cash payment .....	98
Table 4.10: Potential labour cost savings with respect to on-board cash payment.....	99
Table 4.11: Environmental cost savings with respect to on-board cash payment .....	99
Table 4.12: Average user cost savings with respect to on-board cash payment.....	100
Table 5.1: Queuing delay parameters .....	110
Table 6.1: Estimation of parameters, MNL models.....	131
Table 7.1: Proportion of passengers boarding and alighting at each door, as a function of the number of doors per bus, regime $TnBn$ .....	142
Table 7.2: Proportion of passengers alighting at each door, as a function of the number of doors per bus, regime $TnB1$ .....	143
Table 7.3: proportion of passengers boarding and alighting at the busiest door.....	143
Table 7.4: Optimal bus size [m] .....	153
Table 8.1: Parameter values .....	181
Table 8.2: Base case results.....	183
Table A4.1: Cost items related to bus size .....	231
Table A4.2: Cost items related to fare collection technology.....	232
Table A5.1: Area occupied by passengers sitting, standing and other objects.....	234

# Notational Glossary

Variable/ parameter	Description	Units
$\alpha_0, \alpha_1$	Parameters of BPR travel time function	
$\alpha_{max}$	Fraction of bus passengers that traverse the most loaded section of the line	
$\beta_j$	Parameters of bus queuing time	
$\Delta$	Fare payment technology and boarding policy	
$\varepsilon$	Ratio of the average waiting time to the average bus headway	
$\eta$	Safety factor for the calculation of the fleet size	
$\kappa$	Safety factor to have spare capacity on buses	
$\lambda^+, \lambda^-$	Number of passengers boarding and alighting a bus, respectively	<i>pax</i>
$\mu$	Ratio of the value of home waiting time savings to the value of station waiting time savings	
$\tau_a$	Car toll	<i>\$/veh</i>
$\tau_b$	Bus fare	<i>\$/pax</i>
$\varphi$	Passenger car equivalency factor of a bus	
$a, a_i$	Average alighting time per passenger	<i>s/pax</i>
$b, b_i$	Average boarding time per passenger	<i>s/pax</i>
$B$	User benefit	<i>\$</i>
$C_t$	Total cost	<i>\$/h</i>
$C_o$	Total operator cost	<i>\$/h</i>
$C_a$	Access time cost	<i>\$/h</i>
$C_w$	Waiting time cost	<i>\$/h</i>
$C_v$	In-vehicle time cost	<i>\$/h</i>
$c_1$	Busway infrastructure and land costs	<i>\$/km-h</i>
$c_2$	Station infrastructure and operation cost	<i>\$/station-h</i>
$c_3$	Personnel costs (crew) and vehicle capital costs	<i>\$/veh-h</i>
$c_4$	Bus running costs	<i>\$/veh-km</i>
$c_5$	Implementation cost related with the fare payment technology	<i>\$/h</i>
$c_i$	Average cost, mode <i>i</i>	<i>\$/trip</i>
$\bar{c}_m$	Money cost, mode <i>m</i> (fare in case of public transport)	<i>\$</i>
$c_{oc}$	Bus door opening and closing time	<i>s</i>
$C_T$	Traffic light cycle time	<i>s</i>
$D_i$	Inverse demand, mode <i>i</i>	<i>\$</i>
$d_1$	Non-random delay at intersections	<i>s/veh</i>
$d_2$	Overflow delay at intersections	<i>s/veh</i>
$EC$	External costs other than congestion	<i>\$</i>

$f_a$	Car flow	veh/h
$f_b$	Bus frequency	bus/h
$F$	Fleet size	bus
$f_{min}, f_{max}$	Minimum and maximum frequencies, respectively	veh/h
$g$	Effective green at traffic light	s
$\bar{h}_i$	Average number of stops per intersection	stops/veh
$h_m$	Headway between two consecutive vehicles, mode $m$	min
$I$	Number of traffic signal intersections along the route	
$I_u$	Marginal utility of income	
$K_b$	Bus capacity	pax/h
$K_{max}$	Maximum bus capacity	pax/h
$K_r$	Road capacity	veh/h
$K_I$	Capacity of intersections	veh/h
$l_i$	Average trip length in direction $i$	km
$L$	Route length (one way)	km
$n_{den}$	Density of standees per square metre (bus or train)	pax/m <sup>2</sup>
$N_{bi}$	Bus demand in direction $i$	pax/h
$N_b$	Total bus demand ( $N_b = N_{b1} + N_{b2}$ )	pax/h
$N_{km}$	Number of bus passengers per kilometre	pax/bus-km
$N_f$	Total number of passengers per bus	pax/bus
$o_r$	Average car occupancy rate	person/car
$P_a$	Value of access time savings	\$/h
$P_w$	Value of waiting time savings	\$/h
$P_v$	Value of in-vehicle time savings	\$/h
$p_a$	Proportion of passengers alighting at the busiest bus door	
$p_b$	Proportion of passengers boarding at the busiest bus door	
$p_d$	Percentage of passengers that boards buses at high demand stops	
$p_h$	Percentage of high demand stops	
$p_c$	Percentage of passengers that alights buses at high demand stops	
$p_{seat}$	Proportion of seats been used inside bus or train	
$q_i$	Amount of travel, mode $i$	trips
$r_a$	Bus acceleration rate	m/s <sup>2</sup>
$r_d$	Bus deceleration rate	m/s <sup>2</sup>
$R$	Number of roundabouts	
$s_b$	bus length	m
$s_f$	bus saturation flow	veh/h

$SW$	Social welfare	$\$$
$S$	Number of bus stops	
$S_{km}$	Average number of bus stops per kilometre	
$T_c$	Bus round-trip time (cycle time)	$h$
$T_r$	Bus running time	$h$
$T_i$	Bus delay at intersections	$h$
$T_s$	Bus delay at bus stops (total)	$h$
$T_k$	Bus slack time	$h$
$T_{ti}$	Bus travel time in direction $i$	$h$
$TnB1$	Bus boarding policy given number of doors: "Total $n$ doors, boarding at front door only"	
$TnBn$	Bus boarding policy given number of doors: "Total $n$ doors, boarding at all doors"	
$t_0, t_1$	Parameters of waiting time cost	
$t_i$	Delay due to intersections	$s$
$t_r$	Delay due to roundabouts	$s$
$t_N$	Average boarding and alighting time per passenger	$s/pax$
$t_b$	Bus running time	$min$
$t_{iw}$	Internal waiting delay at bus stops	$s$
$t_{ac}$	Acceleration and deceleration delay	$s$
$t_d$	Dwell time	$s$
$t_s$	Bus delay per bus stop	$s$
$t_L$	One-kilometre non-stop bus travel time	$s/km$
$t_{am}$	Access time, mode $m$	$min$
$t_{em}$	Egress time, mode $m$	$min$
$t_{vm}$	In-vehicle time, mode $m$	$min$
$U_m$	Utility of mode $m$	
$u$	Ratio of effective green time	
$v_w$	Walking speed	$km/h$
$v_0$	Bus running speed	$km/h$
$v_{min}, v_{max}$	Minimum and maximum running speed, respectively	$km/h$
$V$	Bus operating (commercial) speed	$km/h$
$x_b$	Degree of saturation at intersections	
$Z_2, Z_3$	Dummy variables for the number of berths per bus stop	



# Chapter 1

## Introduction

The 20<sup>th</sup> century has delivered new challenges for the way we live and interact in both urban and rural environments. Transport is at the core of the basic need of moving passengers and freight from one place to the other; however, attached to the social and economic benefits of transport are a number of externalities, including congestion, pollution and accidents, with impacts that extend beyond the transport sector to affect several related economic fields, the wider community and the environment. In Australia, total travel in urban areas has increased ten-fold over the past sixty years, and the social cost of congestion in urban roads has been estimated to be \$9.4 billion in 2005, a figure that is projected to rise to \$20 billion by 2020 (BITRE, 2007)<sup>1</sup>.

Increasing levels of congestion and environmental externalities are persuasive indicators of the need for re-designing our transport systems, in order to decrease the amount of car traffic and encourage travellers to use more sustainable forms of mobility, such as public transport –including rail and bus alternatives- walking and cycling. In this context, it has been long recognised that implementing and delivering public transport services

---

<sup>1</sup> The Bureau of Transport and Regional Economics (BITRE) of Australia calculates the cost of “avoidable” congestion, which accounts for trips where the benefit for road users of travelling in congested conditions are outweighed by the cost imposed on other road users and the community. The BITRE estimations comprise private and business time costs, extra vehicle operating cost and air pollution.

that are attractive, efficient and affordable is a key element for the present and future sustainability of our cities and the quality of life of their residents.

This thesis is concerned with the pricing and design of urban transport systems; in particular we study the optimal design and efficient operation of bus services, the pricing of both public and private transport, and the role of non-motorised transport alternatives in the development of transport pricing policies. In doing so we explore both policy problems (e.g., what should be the bus fare in different scenarios?) and operational issues (e.g., what is the operating speed gain for buses from upgrading the fare collection system?), whilst in addition analysing the effects of selected transport externalities on both demand (e.g., what is the impact of passengers' crowding in the public transport patronage?) and supply (e.g., what is the effects of bus stop congestion on optimal frequency?). The methodological approach and specific research questions addressed are explained next.

## **1.1 Research Topics and Questions**

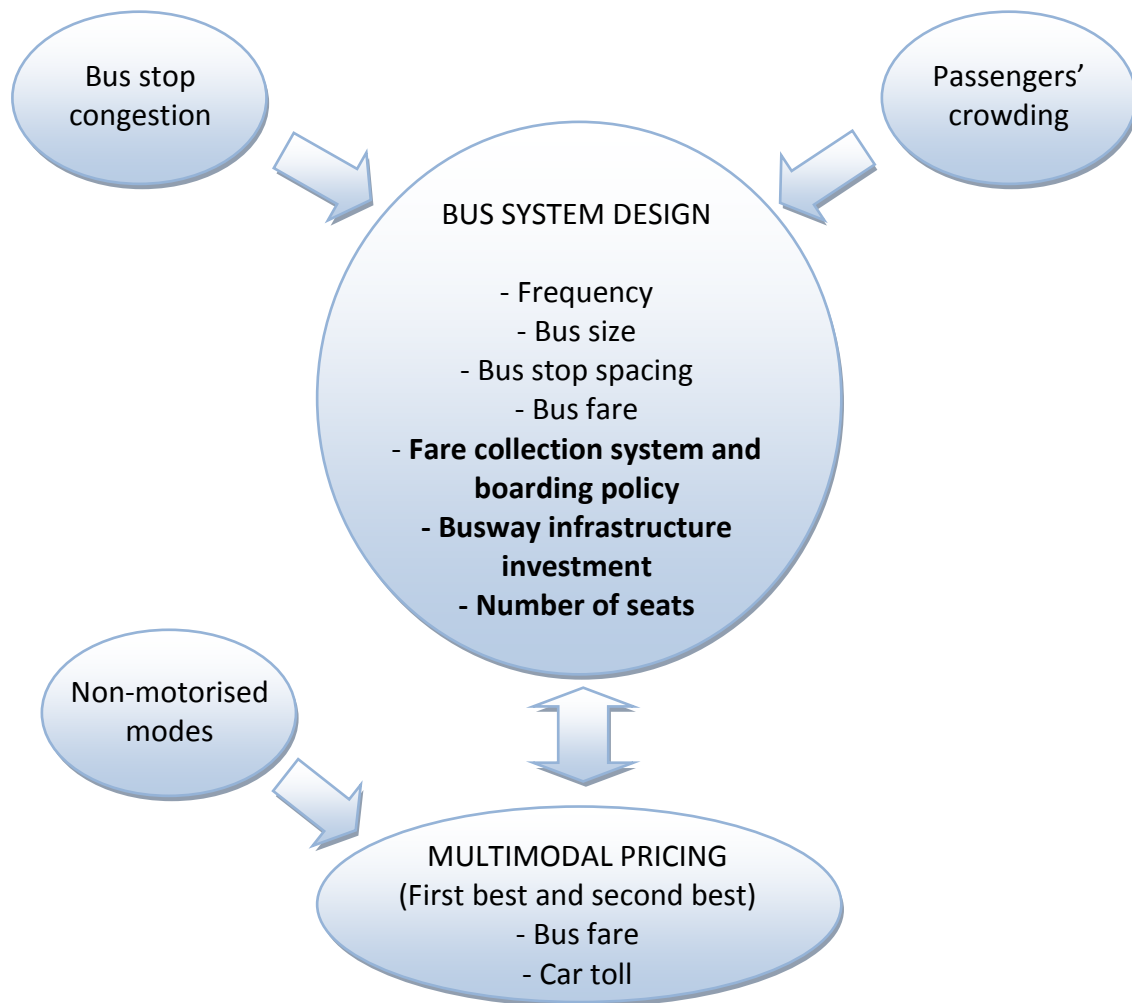
Providing an efficient and effective public transport service involves a large number of decisions that require close scrutiny, such as the design of a network, the choice of mode(s) (e.g., bus, tram, light rail, metro), the fare regime, the nature and level of investment in infrastructure, the number of services per day or hour, the fare collection method and the location of stations or bus stops, among several others. The resulting choices have a significant impact on the cost of the system (and potential subsidy required) and the level of service provided. Accessibility, travel time, crowding, comfort and other factors of importance to users, depend on the design of the service, and hence understanding the economic nature of urban public transport operations continues to attract the attention of researchers, practitioners and policy makers. From a transport planner's perspective, the challenge behind the design of public transport services lays in the myriad number of trade-offs that need to be considered at once, in order to determine an *optimal* service design from an economic perspective, for example:

- i. Increasing bus frequency reduces waiting time for users, but increases the cost of operation.
- ii. Increasing the number of bus stops reduces users' access time, but increases bus riding time.
- iii. Investing in a quicker fare collection technology and dedicated road infrastructure for buses reduces bus travel time (and consequently may reduce operating cost), but increases capital cost.
- iv. Increasing the number of seats on buses improves the quality of service by allowing seat access to more people, but reduces the capacity of vehicles.

The objective of this thesis is improving our understanding of five significant elements of urban bus service provision:

- The influence of considering non-motorised travel alternatives (walking and cycling) in the estimation of optimal public transport fares.
- The choice of a fare collection system and bus boarding policy.
- The influence of passengers' crowding on bus operations and optimal supply levels.
- The optimal investment in road infrastructure for buses, which is attached to a target running speed for buses.
- The characterisation of bus congestion at bus stops and its impact on bus travel time, frequency and capacity.

We provide a new look to classical problems that have driven a large amount of research in the public transport economics literature (such as problems i and ii above), and develop a framework to analyse a number of problems that have been neglected in the literature (such as iii and iv above). A graphical representation of the issues addressed in this thesis is presented in Figure 1.1 (with a bold text to highlight decision variables that are new to the literature).



**Figure 1.1: Relationship among thesis topics**

The key elements of interest and specific research questions of this research are discussed next.

### **1.1.1 Non-motorised modes**

Pricing models that take into consideration only two modes, car and public transport (bus or rail), have found that subsidies for public transport are desirable, with fares below marginal cost due to the underpricing of car use (e.g., Sherman, 1971; Bertrand, 1977; Glaister and Lewis, 1978; Else, 1985; Ahn, 2009; Parry and Small, 2009, among several others). However, as put forward by Kerin (1992), this approach neglects the existence of other modes, notably walking and cycling, that play a significant and growing role in urban mobility, especially for short trips. Recognising the role of non-motorised transport

is relevant because low public transport fares not only deter some drivers from using their cars, but also divert walkers and cyclists into public transport, which is not necessarily a desirable outcome. As such, a pricing and public transport optimisation model that also includes non-motorised transport seems desirable in order to estimate the impact of walking and cycling on (possibly decreasing) optimal subsidies for public transport.

Based on the above, the following research question is formulated:

**Question 1:** *What is the effect of including non-motorised transport alternatives in the optimal pricing of motorised modes (public transport and car)?*

### **1.1.2 Fare collection system and boarding policy**

The existing economic literature on bus transport considers that bus travel or in-vehicle time is either fixed or increases with the dwell time, i.e., time spent transferring passengers at bus stops (e.g., Mohring, 1972, Jansson, 1980), which in its simplest formulation is presented as the passengers' boarding and alighting time plus the time necessary to open and close doors. When the dwell time is considered as a variable, and consequently the total travel time is influenced by the level of demand, analysts have to date assumed that the average boarding and alighting time per passenger is exogenously given, ignoring that currently there are several alternative boarding and alighting policies and technological options for fare collection that have an impact on travel times, operator costs, as well as the complexity and image of the public transport service.

In this thesis we analyse the following research questions:

**Question 2:** *What is the impact of alternative fare collection systems and bus boarding policies (boarding allowed at one or all doors) on bus travel times and associated costs (e.g., fleet size, operating cost, environmental cost)?*

**Question 3:** *What is the optimal fare collection system and bus boarding rule, given demand and operator cost parameters?*

### **1.1.3 Crowding**

The experience of riding on a bus or train may differ both in terms of comfort and perception of the trip given the number of people on board vehicles and at stations. For example, the more users who take a bus, the more stops it has to make to get people on and off (if buses are allowed to skip designated stops when there is neither boarding nor alighting passengers), and the longer dwell times would be if there are more passengers to board and alight buses per stop, incrementing the travel time for everyone. If demand is high enough to produce on-vehicle crowding, the trip is likely to be less comfortable for everyone, a number of users have to stand in aisles, people may experience delays moving inside and getting off and on, or even have to wait for another vehicle if the one they attempt to get on is full.

The analysis of the economic effects of crowding and standing costs inside public transport vehicles has focused, on the one hand, on estimating how the perception of travel time changes with levels of crowding, i.e., the influence of crowding and standing on the value of travel time savings (Maunsell and Macdonald, 2007; Whelan and Crockett, 2009; Hensher *et al.*, 2011; Wardman and Whelan, 2011), and on the other hand, in determining the effect of this discomfort on the optimal bus fare (Kraus, 1991) and the optimal values of bus frequency and size (Jara-Díaz and Gschwender, 2003).

This thesis analyses the impact of crowding on a broader set of policies that includes the optimal design of vehicles in terms of number of seats per bus, and the influence of crowding in the optimal distance between bus stops under congested operations. The following research questions are to be addressed:

**Question 4:** *What is the impact of disregarding the effects of crowding on people's preferences on the design of the optimal road pricing and public transport service and fare levels?*

**Question 5:** *Considering both crowding and standing disutilities, how many seats should buses provide?*

#### **1.1.4 Bus congestion**

Bus congestion is an issue when services are provided with a frequency high enough to produce interactions between buses (including bunching effects). The few authors that have assessed bus congestion in the economic analysis of transport policies have used bus flow-delay functions ‘borrowed’ from car traffic models, such as the linear function implemented by Ahn (2009) in his analysis of bus services and road pricing, and the Bureau of Public Roads (BPR) function used by Fernández *et al.* (2005) in their analysis of bus cost structure. These functions do not fully take into account differences between the sources of congestion in car traffic and in bus systems. A recent theoretical improvement was made by Basso and Silva (2010), who assume that part of the dwell time at bus stops is transferred to cars as extra delays. In this thesis we model the formation of bus queues at bus stops when all the stop berths are being used as the main source of congestion in bus systems, and analyse how optimal values of frequency, capacity and infrastructure investment change when bus congestion is properly accounted for. In this respect, the following research question is addressed:

**Question 6:** *What is the effect on the design of bus systems of misrepresenting bus congestion (or not considering it at all) for scenarios with high bus demand (and which are consequently, subject to bus congestion).*

#### **1.1.5 The provision of busways**

Bus running speed (the cruising speed that buses attempt to maintain in between two consecutive stops) is commonly treated as an exogenous parameter in the microeconomic literature of bus transport, with the assumption that bus speed is given by the physical conditions and regulations (speed limits) of the bus routes under study and by car traffic in the case of mixed-traffic circulation. Nevertheless, bus running speed can be a decision variable, if an investment in infrastructure, like upgrading or building new busways, is designed to have a positive impact on the running speed of buses. The provision of busways goes beyond the simple analysis of whether or not bus lanes should be provided (Mohring, 1983; Berglas *et al.*, 1984; Basso and Silva, 2010), because in reality several degrees of bus segregation (at different capital costs) may yield different speed gains. In this work, we analyse the relationship between investment in dedicated

infrastructure for buses and running speed, which leads to the following research question:

***Question 7:** If bus speed can be increased by investment in infrastructure in dedicated bus corridors, what is the optimal level of investment in busways (which in turn determines the running speed of buses)?*

## **1.2 Research Approach**

The ultimate goal of this research is the optimisation of urban bus routes and pricing regimes for both cars and public transport (Figure 1.1). As explained in the previous section, the focus is on elements that have received partial treatment in the literature, including the effect of crowding and bus congestion on optimal bus supply, the choice of a fare collection system and infrastructure investment in bus corridors, and the influence of non-motorised transport modes on the optimal level of public transport fares. The research approach is summarised in a sequential way as follows:

- First, it is necessary to develop microeconomic models for the optimisation of urban public transport that account for bus congestion, crowding, fare collection systems and bus infrastructure investment.
- In the microeconomic literature on public transport operations, the most common modelling approach is the minimisation of total cost, which comprises operator cost and the users (time) cost (Mohring, 1972; Jansson, 1980; Chang and Schonfeld, 1991; Jara-Díaz and Gschwender, 2003). An extended total cost minimisation model is developed in this thesis, in order to determine an optimal value of bus frequency, size and distance between stops, plus the decision on infrastructure investment (an associated running speed), fare collection technology and boarding policy (one-door versus all-door boarding).
- A multi-modal framework that integrates car, public transport and non-motorised modes (bicycle and walking) is formulated, to account for the relationship between road pricing, non-motorised modes, pricing and service levels for public



transport. The objective function is one of maximising social welfare, comprising consumer surplus and the profit of public transport operator and the road pricing regime.

### **1.2.1 Fare collection system and boarding policy**

Alternative fare payment systems and bus boarding rules differ in requirements of infrastructure support, the ability to integrate fares across routes and modes, security, operating cost including transaction costs, evasion control, and capacity to handle different fare structures. All these features should be weighed up against each other when deciding on a fare payment system for a specific bus service or network. This study focuses on differences in travel time and operator costs. To this end, average boarding times of several alternative fare payment systems and boarding policies are estimated, including on-board payment with cash, magnetic strip, contactless card, and off-board payment. For the estimation of boarding times with cash, magnetic strip and off-board payment, dwell times survey are performed in Sydney, whereas the average boarding time with contactless card for fare validation inside buses is taken from a study in Santiago, Chile (Fernández *et al.*, 2009). For the estimation of the capital cost associated with each fare payment technology, we take into account the cost of software, vending machines, card validation devices and tickets readers, with cost indicated in the Bus Rapid Transit Planning Guide published by Wright and Hook (2007).

### **1.2.2 Crowding**

We consider that crowding increases the valuation of travel time savings. Using data from a mode choice experiment conducted in Sydney, that includes attributes on the number of seats available and the number of passenger standing inside vehicles for the public transport alternatives (bus, train and metro, the experiment is described in Hensher *et al.*, 2011), we estimate the impact of crowding and standing disutilities in the value of travel time savings using a multinomial logit model (MNL).

### **1.2.3 Bus congestion**

Bus congestion is a phenomenon that has received little attention in the literature, possibly because most of the research on urban transport modelling and pricing is

associated with developed countries where it is relatively unusual to find situations with bus frequencies high enough to cause congestion. We estimate the formation of bus queues behind a station when all the stop berths are being used to transfer passengers by preceding vehicles. After analysing bus stop operations with the simulation model IRENE<sup>2</sup>, Fernández *et al.* (2000) found that the queuing delay grows exponentially with the frequency of buses that enter the bus stop.

As argued by Fernández and Planzer (2002), a simulation approach is well suited to analyse key performance measures of bus stops (like the queuing delay), because the processes involved in the arrival of buses, passengers and the interaction between them are very complex and usually random, which suggests that analytical steady-state approaches such as the Highway Capacity Manual (HCM) formula to calculate bus stop capacity (TRB, 2000) have a limited application. Consequently, we use IRENE to estimate queuing delays as a function of frequency, bus length, number of berths and average dwell time.

#### **1.2.4 The provision of busways**

A linear relationship between infrastructure cost per kilometre and running speed is proposed and embedded into a total cost minimisation model for bus operation, based on a positive correlation between infrastructure investment and commercial speed (total speed including stops), empirically identified by comparing data from a number of Bus Rapid Transit (BRT) systems (Wright and Hook, 2007). We show that a target speed increases the investment in infrastructure but also reduces the travel time between stops, and hence a compromise running speed is selected as the optimal solution.

#### **1.2.5 Total cost minimisation model**

The economics of urban bus transport is first analysed by modelling a single bus corridor, segregated from other modes (e.g., cars and trucks). The total cost minimisation model

---

<sup>2</sup> IRENE is a bus stop simulator that calculates the capacity, queuing delay, dwell time, berth usage and other indicators of the performance of a bus stop, as a function of a number of inputs such as the boarding and alighting demand, number of berths, stochasticity of both user and bus arrivals. Parameters exogenously given are frequency and the mean number of passengers boarding and alighting per bus. For a detailed description of the program see Gibson *et al.* (1989) and Fernández and Planzer (2002).

includes operator and user costs, the former comprising the cost of buses, dedicated infrastructure, bus stations, crew and operating cost, whereas the latter comprises access, waiting and in-vehicle time costs. To the traditional framework of optimising frequency, bus size and/or distance between stops we add the aforementioned decisions on boarding rules and fare collection technique and investment in road infrastructure for buses, in a framework that considers bus congestion in the form of queuing delay at bus stops. Total cost is minimised subject to constraints on capacity, minimum and maximum frequency and running speed.

### ***1.2.6 Multimodal pricing and bus optimisation***

Finally, the optimisation of bus services is analysed in a multimodal setting that includes cars and walking as travel alternatives. We present a deterministic social welfare maximisation model in order to analytically explore the impact of non-motorised transport in first best and second best pricing of motorised private and public transport. Then, a numerical analysis is performed by assuming that modal choice follows a multinomial logit model, which is applied to a transport corridor in Sydney, divided in zones in order to analyse spatial differences in mode choice (e.g., number of walking trips as a function of trip length), i.e., modal choice is different per origin and destination. The location of bus stops is fixed in this model, which allows us to know the number of passengers inside buses in each segment of the route (between two consecutive stops), information that is used to determine the number of seat that buses should have, considering crowding and standing disutilities.

## **1.3 Thesis Contributions**

### ***1.3.1 New elements and methodological refinements***

The main contributions of this thesis are classified into two groups:

**New elements:** These are factors and variables that have not been previously analysed in the literature. For example, the influence of different fare collection systems in the optimal design of a bus route (frequency, bus size and distance between stops).

**Methodological refinements:** These encompass the analysis of effects that have been previously addressed in the literature; however the contribution of this work lies in a more comprehensive treatment of the phenomenon. For example, the influence of bus congestion on optimal frequency, capacity and distance between stops.

A graphical summary of the contributions of this thesis is shown in Figure 1.2. The scientific and practical relevance of this work is discussed next.

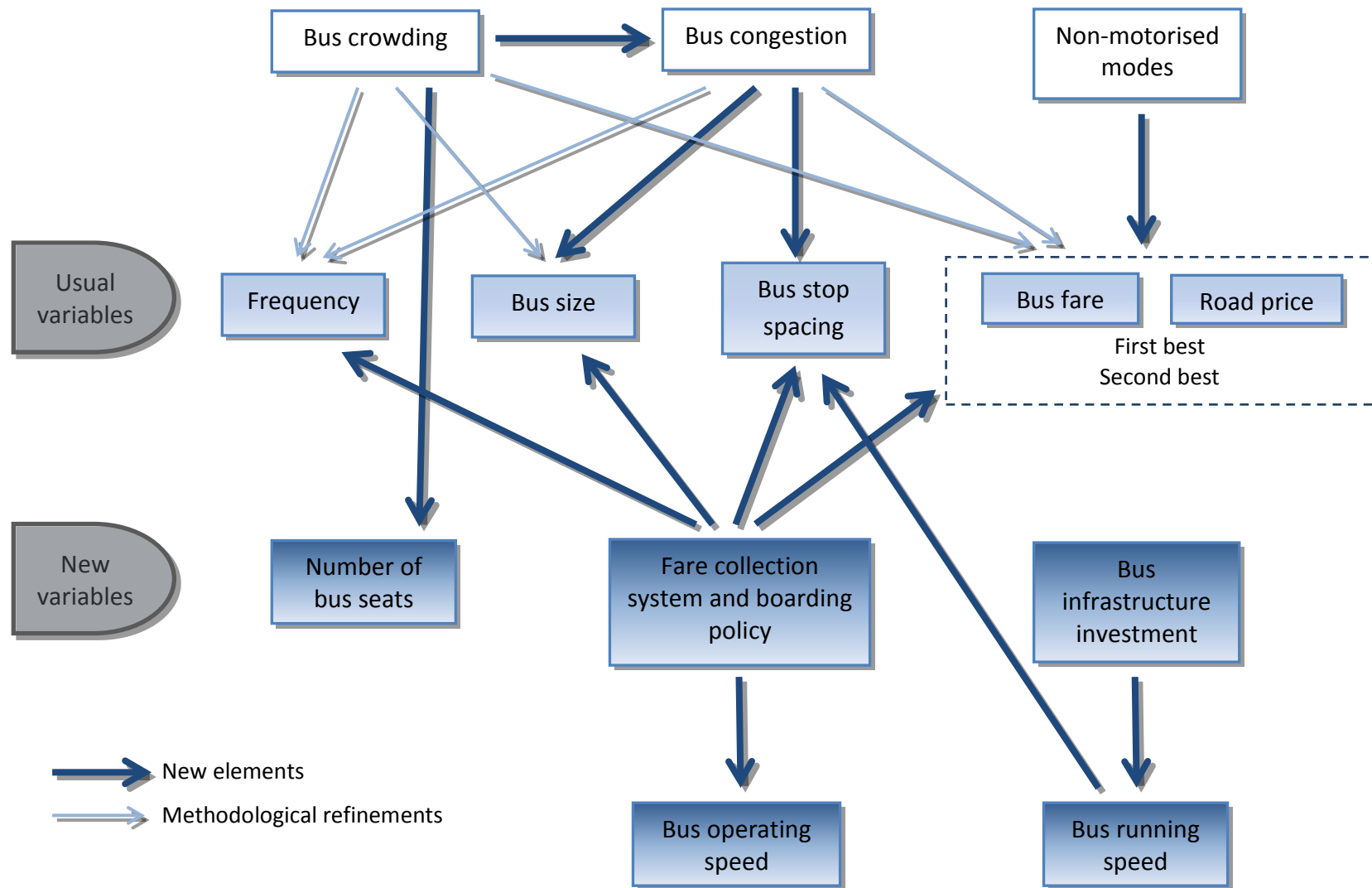


Figure 1.2: Thesis contributions

### **1.3.2 Theoretical and scientific relevance**

The main scientific contributions of this thesis are the following.

- Derivation of second-best bus fare with explicit account of underpriced cars and an uncongestible non-motorised mode as travel alternatives. We obtain analytically the conditions that lead to the underestimation or overestimation of the optimal bus fare when non-motorised transport is ignored (Chapter 3).
- The decision of a fare collection system and bus boarding policy is embedded into microeconomic models for the optimisation of bus systems. Faster bus boarding techniques (e.g., upgrading from of-board cash payment to off-board contactless card payment validation) present the trade-off of reducing riding time and increasing capital cost (Chapters 4 and 7).
- Bus congestion in the form of queuing delays behind bus stops is estimated using simulation. The delay function depends on the bus frequency, bus size, number of berths and dwell time (which is given by the number of passengers boarding and alighting, the number of doors per bus and the fare collection technology). Therefore, models that use flow measures (including frequency only or frequency plus traffic flow) as the only explanatory variables for bus congestion are incomplete (Chapters 5 and 7).
- The crowding externality and standing disutility for passengers inside buses is used to determine the optimal number of seats that buses should have. Explicit constraints are considered for the determination of seating and standing areas on buses (Chapters 6 and 8).
- Calculation of optimal bus road infrastructure investment (and bus running speed) in dedicated bus corridors (Chapter 7).

### **1.3.3 Practical relevance**

This thesis presents a number of contributions for practice, as summarised next.

- The potential substitution between motorised and non-motorised modes (walking and cycling) should be considered when estimating the second-best public

transport fare. If the modal substitution between public transport and non-motorised modes is strong relative to the substitution between car and public transport, and between car and non-motorised modes, it is more likely that the optimal public transport fare is underestimated if non-motorised alternatives are ignored (Chapter 3).

- Boarding times with different fare collection systems and boarding policies (boarding allowed at one or all doors) estimated can be used to empirically assess benefits for users and operators of upgrading the fare collection system in terms of running time, fleet size, operating and environmental cost. With an empirical model on bus running times with and without the influence of traffic congestion, a comparison of providing bus lanes versus upgrading the fare collection system is also possible (Chapter 4).
- Empirical estimation of the actual number of bus stops as a function of scheduled bus stops and demand, based on data from an on-call bus service in Sydney. This function can be used to estimate the number of stops per bus ride given the number of passengers boarding and alighting, and consequently total riding time (Chapter 4).
- The bus congestion function estimated with the simulator IRENE is useful to analyse the influence of several factors on the occurrence of queuing delays at bus stops, such as the number of passengers boarding and alighting, the fare collection technology, the number of doors to board and alight, the bus frequency and size and the number of berths (Chapter 5).
- The investment on quicker fare collection systems is justified as demand grows. Using a total cost minimisation framework, demand thresholds for the introduction of a more sophisticated fare payment technology can be identified (Chapter 7).
- Decreasing total average cost is observed when boarding is allowed at all doors, whereas increasing average costs occur for high demand if boarding is restricted to the front door only. The highest total cost is associated with on-board cash payment, followed by payment with magnetic strip and contactless card. This is because buses spend more time in bus stops boarding passengers when payment

is on-board at the front door only, which triggers bus queues that in turn increase travel time and operator costs. This highlights the importance of having an efficient fare payment system and bus boarding policy, as a way to avoid bus congestion as much as possible (Chapter 7).

- Disregarding bus congestion in the design of the service would yield greater frequencies and smaller buses when congestion is noticeable, i.e. for high demand (Chapter 7).
- Optimal bus frequency results from a trade-off between the level of *congestion inside buses*, i.e., passengers' crowding, and the level of *congestion outside buses*, i.e., the effect of frequency on slowing down both buses and cars in mixed-traffic. In particular, optimal bus frequency is quite sensitive to the assumptions regarding crowding costs, the impact of buses on traffic congestion and the overall congestion level. We show that if crowding matters, bus frequency increases (for a given bus size) with demand even under heavy congestion. However, that might not be the case if the crowding externality is not accounted for, in which case an increase of total demand might be met by a decrease of both frequency and number of seats per bus, at the expense of crowding passengers inside buses and making more passengers stand while travelling (Chapter 8).
- Regarding the relationship between crowding, standing and the number of seats, it is shown that in an scenario with no capacity constraints, buses should have as many seats as possible (given constraints on minimum areas for aisles and around doors that must not be allocated to seating), and that if the number of seats is reduced, frequency should be increased (Chapter 8).

#### **1.4 Research Scope**

Given the large number of problems that this thesis addresses, we have limited the scope of the theoretical approach in several respects.

In terms of geographical or spatial scope, the analysis is reduced to a single corridor, either for the optimisation of bus services only (i.e., a single route in Chapter 7) or for the



multimodal social welfare maximisation model that includes cars and non-motorised modes (Chapter 8). Transport networks are not considered.

The empirical estimation of bus travel times in Chapter 4 considers several time periods; however, the total cost minimisation and social welfare maximisation models are developed in a single period framework, which is assumed to be the morning peak in the numerical applications.

General equilibrium issues and tax distortions are discussed as part of the literature review on car and public transport pricing (and subsidies) in Chapter 2; however the economic model and pricing rules developed consider the transport market only, abstracting from distortions in the rest of the economy.

Regarding bus operations, the models assume that buses maintain a regular headway, i.e., the issue of bus bunching is ignored. The optimisation of bus timetables is not addressed either, which might be important when translating the potential time benefits of, for example, upgrading the fare collection system from a slow to a quicker technique. The physical and fare integration between buses and other public transport modes like rail are also ignored.

## **1.5 Thesis Outline**

Chapter 2 provides a literature review on the microeconomic models of public transport operation and optimal pricing rules. The concepts of first best and second best pricing are revisited, together with the theoretical foundations for subsidising public transport. The setting of bus frequency and capacity, the choice of providing bus lanes and the influence of other sectors of the economy on transport pricing, among other topics, are discussed in light of the relevant literature in this field.

In Chapter 3 a social welfare maximisation model is set up with the objective of revealing the influence of walking and cycling alternatives in the optimal pricing of motorised public and private transport. The impact of an active capacity constraint in the public

transport mode, and of considering externalities other than congestion, is also revisited in the proposed three-mode framework.

Chapters 4 to 6 are concerned with the operation of bus routes. Chapter 4 provides a review of issues regarding the choice of a fare collection system in urban bus services. We estimate boarding and alighting times with alternative fare payment methods and boarding and alighting rules (regarding number of doors to board and alight). Dwell time models are presented and estimated using data collected in Sydney for payment with on-board cash, magnetic strip (pre-paid ticket) and a free bus service which is used as a proxy for off-board fare payment and validation. A bus running time model (including all stages of a trip) is also estimated using empirical data from Sydney in order to calculate potential benefits from upgrading the fare collection system (for example, from on-board cash payment to on-board magnetic strip validation or off-board fare validation), in particular, savings on bus running and crew cost, environmental cost, fleet size and travel time for passengers.

Chapter 5 reviews models for the inclusion of bus congestion in the economic analysis of urban transport. Using the bus stop simulator IRENE (Gibson *et al.*, 1989; Fernández and Planzer, 2002), queuing delays at bus stops are estimated as a function of the bus frequency and size, number of berths and the dwell time at the stop. The influence of alternative fare collection systems and boarding policies on bus queuing delays is analysed.

Chapter 6 provides a review of the influence of passengers' crowding in public transport demand and supply, including the effect on riding time, waiting time, the valuation of travel time savings, travel time variability and optimal fare. New crowding cost functions are estimated for Sydney.

Next, we set up microeconomic models for the optimisation of bus supply and multimodal pricing, including the elements discussed in Chapters 4 to 6. Chapter 7 sets out an extended total cost minimisation model (including both users and operator costs) for the design of a bus route. The emphasis is on the optimal choice of a fare collection

system, the optimal investment in bus infrastructure (which has associated a target running speed for buses running in a dedicated corridor), and the effects on the optimal design of the queuing delays that arise at bus stops in high-frequency high-demand scenarios.

In Chapter 8 we develop an extended social welfare maximisation model for the optimisation of a bus route including bus fare and road price, assuming that mode choice is governed by a multinomial logit model. This model integrates the design issues discussed in Chapters 4 to 7 in a multimodal framework that includes walking, car and bus as travel alternatives.

Finally, Chapter 9 summarises the main findings, methodological contributions and policy implications of this thesis, and provides directions of further research.

## Chapter 2

# Background: Microeconomic Modelling of Public Transport Operation and Optimal Pricing<sup>3</sup>

### 2.1 Introduction

Travellers usually have several modal alternatives available in specific trip purpose contexts and geographical jurisdictions, including private and public transport options, motorised and non-motorised. Aside from availability, monetary cost and trip time outlays are usually the most important attributes when choosing a mode, and hence when the price or fare can be changed, it becomes a tool to influence modal demand, and consequently the level of transport-related externalities, as well as the social welfare and distributional impacts of the transport system. The economics of transport pricing has attracted the interest of economists for over a century, since Dupuit (1844), Pigou (1920) and Knight (1924) analysed tolls as a way to recover costs or increase the economic efficiency of roads<sup>4</sup>. Researchers have devoted considerable effort to analyse the merits of road pricing as a tool to manage congestion and other externalities derived from transport activity. Economic models of transport pricing focus on either a single mode or

---

<sup>3</sup> The literature review in this chapter is partially published in Tirachini and Hensher (2011; 2012).

<sup>4</sup> The history of the idea of road pricing and the evolution of the research on this topic is extensively reviewed by Lindsey (2006).

multimodal approach; however the analysis of road pricing for private transport has received a disproportionate amount of attention relative to public transport and multimodal analysis, including free personal travel alternatives. This chapter reviews a large set of issues associated with the pricing of urban transport modes and the economics behind the operation of public transport systems.

## **2.2 Setting Public Transport Fares: First Best and Second Best Models**

The analysis of transport pricing schemes usually distinguishes between first best and second best policies. A situation in which all prices match marginal costs is known as *first best*. As reviewed by Quinet (2005), in the first best world there are no external effects, no public goods, firms are price-takers, there is no tax or taxes are optimal, there is no uncertainty or asymmetry in information, there are no transaction costs and no redistribution problems. However, transport systems in the real world do not match all these conditions and several departures can be found, such as the influence of external factors, non-competitive markets, non-optimal taxes and so on, associated with a condition known as *second best*<sup>5</sup>. Technological or acceptability constraints are common factors that impose second best situations within the transport sector, given the impossibility of taxing at marginal cost all modes or all locations in a network.

A second best situation is the most likely to exist when designing a transport pricing reform initiative. Transport economists know that the derivation of optimal charge levels is more difficult for second best policies than in a theoretical first best environment, as the former should take into account all the inefficiencies and distortions in the market, whereas in the latter situation only the marginal costs need to be determined (Rouwendal and Verhoef, 2006). Even though the first best is practically unachievable, it is useful as a benchmark to compare any second best policy that is proposed. Next we review the

---

<sup>5</sup> See Rouwendal and Verhoef (2006) or Small and Verhoef (2007) for a more detailed discussion on second best issues.

concepts of first best and second best pricing for public transport, in light of the relevant literature in this topic.

### **2.2.1 First best pricing**

The principles of marginal cost pricing of private transport have a long history. In the context of automobiles, it had been recognised from very early days that establishing a cost function for the study of demand and welfare must include travel time as a key influence. In the study of public transport pricing, such a realisation appears to have occurred only in the 1970's, with the works of Mohring (1972), Turvey and Mohring (1975) and Jansson (1979). The recognition that user costs should be included in the determination of optimal fares for public transport is illustrated in the following quotes:

*“The right approach is to escape the implicit notion that the only costs which are relevant to optimisation are those of the bus operator. The time-costs of the passengers must be included too, and fares must be equated with marginal social costs.”*  
(Turvey and Mohring, 1975, p. 280)

*“(…) in the wide field of scheduled transport it has only recently been realised that the principle of marginal cost pricing is practically impossible to apply correctly unless all users sacrifices and efforts are, at least conceptually, treated as costs on a par with producers costs.”*  
(Jansson, 1979, pp. 270-271)

The addition of user time costs as an input in the social cost function of public transport proved to have remarkable consequences for the application of a marginal cost pricing rule. When an increase in demand is met by an increase in the frequency of service, i.e., an increase in scale, the travel cost of all users decrease due to savings in waiting time (assuming that waiting time is inversely related to frequency), a phenomenon that is not observed when the operator or producer cost is the only item considered in the cost function of public transport<sup>6</sup>. Consequently, marginal cost lies below average cost, which

---

<sup>6</sup> Note that the operator cost may also exhibit scale economies as shown by Allport (1981) for trains and buses. In the particular case of railways, the existence of scale economies due to the high fixed costs of infrastructure has been extensively used as an argument to justify rail subsidies (Preston, 2008).

is the first best argument for subsidising public transport operation, as introduced by Mohring (1972) and Turvey and Mohring (1975). Intuitively, a lower fare will encourage more travellers to use public transport, which would be accompanied by an increase in the optimal frequency, which lowers waiting times, therefore benefiting all passengers (Jansson, 1993).

The first best fare is the one that maximises social welfare, defined as the summation of users' and operators' benefit. The unrestricted solution of this problem is a well-known result, namely that the optimal public transport fare equals total marginal cost (i.e., the summation of users and operator marginal cost) minus the average users cost (e.g., Else, 1985; Tisato, 1998). Then, the general principle of marginal cost pricing as a means to reach economic efficiency applies to public transport services, but with the subtraction of what users already "pay" when using the service, i.e., their own time (Jara-Díaz, 2007).

The scale economies approach of Mohring predicts that the optimal subsidy per passenger decreases with demand, a result that depends on the assumption of a strictly inverse relation between waiting time and frequency, which exists when passengers arrive randomly at stations or bus stops, i.e., when the service frequency is high (average headways shorter than 10 or 15 minutes). However, on low frequency services, passengers usually follow a published timetable and plan their trips accordingly, thus arriving at bus stops or train stations just before the scheduled departure of the service (assuming adherence to schedule). The consideration of this binary passenger behaviour has consequences on the shape of the first best fare and subsidy, as shown by Jansson (1993) and Tisato (1998). In both low and high frequency regimes, there is a schedule delay cost (waiting time cost for the high frequency case), because departures are not at the desired time. However, in the low frequency scenario, most of the schedule delay is spent at home or work, where passengers can allocate their time in a more useful way than when waiting at bus stops, and hence the valuation of schedule delay savings is higher when passengers do not follow a timetable (high demand-high frequency) than

when a timetable is followed (low demand-low frequency). Given that the savings in waiting time due to the optimal adjustment of frequency is the main argument for marginal pricing below average operator cost (first best subsidy), Jansson (1993) and Tisato (1998) show that, in each regime (low frequency and high frequency), the optimal average subsidy is decreasing in demand, but with demand levels in the transition between the two regimes, the optimal subsidy per passenger may actually be increasing with patronage.

The work of Tabuchi (1993) marks the beginning of a renewed interest in the properties of the bi-modal equilibrium between private and public transport under different pricing regimes. Instead of assuming static congestion for the automobile, Tabuchi assumes a dynamic bottleneck that arises when the flow of cars exceeds the capacity of the road (Vickrey, 1969; Arnott *et al.*, 1993). A competing rail service is provided, which exhibits economies of scale due to fixed capital costs. In a highly stylised model that ignores travel time as a cost for rail users and capacity constraints, different first best and second best pricing policies are analysed in terms of social costs, showing the advantages of fine (i.e., dynamic) over coarse (i.e., uniform) optimal tolls, and that as demand grows it is more attractive to have a rail based alternative competing with cars, due to economies of scale in the former and congestion externality in the latter mode.

Tabuchi's two-mode model has been subsequently extended by a number of researchers. Danielis and Marcucci (2002) analyse how the optimal road price should be modified given that rail is priced at average cost instead of marginal cost, i.e., there is a budget constraint imposed that prevents public transport subsidy (also referred to as Ramsey pricing, see Section 2.3.5). At the same time, Huang (2002) applied a multinomial logit model to empirically represent mode choice instead of the deterministic approach of the original Tabuchi's model. The results support the same general principles for first best and second best pricing, regardless of the deterministic or stochastic nature of the equilibrium demands.



Interestingly, it is common that theoretical models that compare automobile and bus assume that the economies of scale for public transport are in the user cost component (Mohring, 1972), whereas models that compare automobile with rail tend to assume that the economies of scale are given by the fixed capital cost of rail (e.g., Tabuchi, 1993; Danielis and Marcucci, 2002; Huang, 2002). Both assumptions are simplifications, since bus service provision also has non-negligible fixed costs; furthermore the schedule delay cost should not be ignored for rail, even when the service provider publishes a timetable.

A different approach was presented by Kraus and Yoshida (2002) who adopted the highway bottleneck model of Vickrey (1969) for the modelling of rail commuting, assuming that users arrive at stations at the same time as trains do. The authors show that the average users cost increases with demand, i.e., the opposite result to the decreasing average users cost of all of the Mohring's type of models, a result explained in part because the length of the peak period is not fixed, such that as demand grows, the peak period enlarges (i.e., some passengers take earlier trains), which increases the schedule delay cost at the destination, given that the desired arrival time is fixed. Kraus and Yoshida (2002) provide an important insight into how the scheduling considerations of users affect average costs of travelling; however their approach is less appropriate for modelling high frequency services, in which it has been empirically observed that passengers arrive at stations or bus stops randomly at a more or less constant rate (e.g., RAND, 2006). Therefore, waiting time at stops exists even if the capacity constraint is not binding, and consequently, the economies of scale induced by increasing frequency should be accounted for.

At this point we need to mention that not only are additional benefits for users attached to a more frequent public transport service, costs could also be incurred if providing additional bus kilometres has a negative effect on speeds for both buses and cars, especially when frequency is high enough for buses to actually slow each other down, as well as other vehicles that share the road with them (more details in Chapter 4 on bus

congestion). In this case, an increase in frequency can augment total average cost, and Mohring (1972)'s scale economies argument for bus subsidies could no longer apply. In this case, buses may operate with an operational surplus if marginal cost pricing is in place, as found in a later study by Mohring (1983). Nonetheless, there are a number of strategies that can be used to make bus transport more efficient in order to minimise or avoid bus congestion. For example, in Chapter 7 it is shown that increasing total costs are observed for a demand over 3,000 pax/h-direction, if passengers are allowed to pay fares on board buses and boarding is at the front door only, in which case frequency is over 60 bus/h (cash payment); however, decreasing total costs are still obtained even for higher frequencies when the fare payment is performed with pre-paid collection technologies and boarding is allowed at multiple bus doors simultaneously.

### **2.2.2 Second best pricing**

As widely recognised in the literature, several departures from ideal first best conditions exist in reality. In the case of public transport pricing, the most evident and analysed case is that buses or trains compete with underpriced cars, as cultural, technical, political or social constraints impede the setting of marginal cost road pricing. When this inefficiency is present, the optimal pricing analysis in public transport is referred to as second best pricing (although as previously mentioned, the second best concept may encompass several other distortions both within and outside the transport sector). The classical argument is that if cars are underpriced, there is an excess of car travel, therefore it would be welfare improving to reduce the public transport fare in order to attract some drivers to use trains or buses, in turn reducing the level of congestion and other traffic externalities on the road network. This is a second economic-based rationale to subsidise public transport, after the economies of scale (first best) argument<sup>7</sup> (Preston, 2008; Parry and Small, 2009). Therefore, as argued by Small (2008), a conclusion from first best and second best fare analyses is that congestion charging could be seen as a way to reduce

---

<sup>7</sup> Other arguments in favour of subsidising public transport include pursuing distributional or social objectives and option values, which are not treated in this thesis (see Kerin, 1992; Preston, 2008).

the financial needs of public transport, since an optimal road charge should decrease the subsidy required for public transport, even if the revenue from road pricing is not hypothecated to public transport.

Formal proofs that an alternative mode should be priced below marginal cost when cars are priced at average instead of marginal cost can be traced to Lévy-Lambert (1968), Marchand (1968) and Sherman (1971). The idea, linked to competitive neutrality, was extended by Glaister (1974), who finds a second best bus fare below marginal cost, not only in the peak but also in the (congestion free) off-peak period, the latter due to two effects - a low off-peak bus fare can attract peak car users, and peak bus users are attracted to travelling by bus during the off-peak, which relieves pressure in the peak, and therefore decreases the peak bus fare, which in turn attracts more car travellers into public transport.

A slightly different approach is introduced by Jackson (1975), who instead of calculating the optimal second best fare, estimates the optimal second best subsidy directly, assuming the underpricing of highway travel, and that average cost per bus user is constant. The optimal fare subsidy depends on the level of congestion on the road, the own cost elasticity of transit demand, and the cross cost elasticity of demand between private and public transport. Interestingly, Jackson (1975) proposes a method to determine the optimal subsidy for bus speed improvements, instead of covering fare reductions, with the result depending on how large the increase in operator cost is to achieve improvements in speed. Illustrative examples show that the welfare improvements largely rely on, and increase with, the degree of congestion associated with highway travel, and the cross elasticity of demand with respect to the generalised cost of public transport. Jackson's case for second best welfare maximisation through an increase in the quality of bus service, as opposed to a fare reduction, parallels the contributions of Mohring (1972) and Turvey and Mohring (1975), who identify a first best

justification for bus subsidies due to the reduction in bus waiting times if frequency is optimally adjusted as demand grows, which is another way of speeding up bus travel.

In terms of congestion externalities, most transport pricing studies focus on automobiles (and trucks in the freight context) as the major source of traffic congestion imposed on all modes that share the right of way; however this is not necessarily the case for high frequency bus services that may slow down both cars and buses. The effect of this congestion effect of buses on cars was analysed by Else (1985), who shows the impact of the external congestion cost of public transport over the second best fare and subsidy. Bus congestion by itself increases the optimal fare and reduces subsidy, however, using British data, Else (1985) shows that even when recognising that public transport contributes to congestion, the optimal fare does not cover operating cost, and an optimal subsidy is required.

Finally, we mention the work of Parry and Small (2009), who show that substantial gains in social welfare are accrued from diverting car drivers into public transport (second best argument) in peak periods, whereas the case to subsidise fares due to the reduction of users costs (scale economies – first best argument) is stronger in the off-peak. When there are two public transport modes (bus and rail), they consider that reducing the fare on one has consequences over the other; for example a drop in the rail fare would attract bus passengers, resulting in increased waiting and access times scale economies, negative effect), but decreased bus operator cost, in-vehicle crowding and externalities (positive effect).

In summary, the above suggests that setting public transport fares below average operator cost is supported by most of the formal analysis of pricing, resulting in the call for an 'optimal' subsidy regardless of whether it is based on first best or second best grounds. Despite the rigorous analytical approaches and empirical evidence, the extant literature has a number of limitations associated in particular with the omission of non-

motorised modes such as walking and cycling, and the distortionary effect of bus subsidies, as identified by Kerin (1992), who more precisely states that “the results of the second-best pricing studies are derived under conditions that are probably unduly favourable to second-best bus pricing. If the key omitted factors could be incorporated into the trade-off process, the optimal second best subsidy level would probably be much lower than that suggested by existing formal models” (Kerin, 1992, p.39). Some of these factors have been accounted for in more recent research, such that possible inefficiencies associated with subsidy (Section 2.2.3), the existence of tax distortions and their interaction with the transport system (Section 2.3.4), and the impact of bus congestion on travel times and operation costs (Chapter 5). The influence of non-motorised transport on optimal pricing decisions is addressed with a multimodal pricing model for the maximisation of social welfare in Chapter 3.

### **2.2.3 Issues that arise when subsidising public transport**

Observed practice has shown a number of problems associated with public transport subsidies that stylised first best and second best models have ignored. The realisation of the efficiency gains that optimal subsidies in theory yield in practice, depends on several factors, such as the form of the subsidy (operating subsidy per passenger or passenger-kilometre versus one-off grant), the structure of the service provider (private or public company) and the relationship between the provider and the subsidising body (Else, 1985). Moreover, the authority may not have sufficient information on costs and demand to estimate the optimal level of subsidy (Frankena, 1983).

A potentially major problem is the inefficiency induced in the operation of public transport services by some types of subsidy. Several authors have shown that when a subsidy is outlaid to cover the gap between operating cost and revenue, in particular for public bus operators protected from competition, it may distort the incentives of managers and workers, reducing productivity and increasing labour wages, as empirically shown in the 80's by Bly *et al.* (1980), Cervero (1984), and Pickrell (1985) among others.

Recent research has shown that there are ways to contain the cost spiral in the presence of subsidies, through performance-based benchmarking and the use of service quality indicators in service contracts (Hensher and Prioni, 2002; Hensher and Stanley, 2003; Mazzulla and Eboli, 2006; Gatta and Marcucci, 2007), action by the regulator to enforce penalties for poor performance, and the application of competitive tendering (Hensher and Houghton, 2004; Hensher and Wallis, 2005).

A related issue discussed by Preston (2008) comes from the distinction between capital and operating subsidies. One-off subsidies targeted specifically to capital investment may condition the decisions of policy makers and operators towards over-investing in capital, for example, acquiring more sophisticated or newer vehicles instead of spending on the maintenance of the current fleet. Pre-defined rail-specific capital subsidies may also lead to unjustified rail investments in areas with low demand for public transport, with the second round effect of inducing an unnecessarily large subsidy for operation. Therefore, the correct *ex-ante* determination of capital and operating subsidies is crucial to ensure efficiency in the allocation of resources to public transport service provision.

In summary, the way in which a (supposed to be optimal) subsidy is paid matters, and the business environment should be defined in a way to minimise or eliminate potential money waste induced by ill-designed subsidies. The design of contracts to tackle this problem is a topic of ongoing research and continuous learning and adjustment in public transport agencies around the world.

## **2.3 Results that Matter**

### **2.3.1 *Peak versus off-peak fares***

Several authors have moved beyond considering a single period of operation, distinguishing peak and off-peak periods. This issue was first addressed qualitatively by Turvey and Mohring (1975), who make a case for a higher first best fare in the peak period and in the peak direction, given that bus occupancy is higher, and therefore, on the

one hand more passengers are affected when an extra passenger boards a bus, and on the other hand the probability of a passenger not being able to board a full bus is greater.

A more complex way to consider this problem is that when travellers face an increase in the price of one mode in one period, they may not only switch mode, but also switch departure time, given values of cross-mode and cross-period demand elasticities (Glaister, 1974; Glaister and Lewis, 1978; De Borger *et al.*, 1996; Hensher, 2002). In a two-mode two-period second best pricing model, Glaister (1974) finds that both peak and off-peak bus fares are below marginal cost, but the relationship between the two cannot be determined a priori, and more specifically, peak bus fare is not necessarily greater than off-peak bus fare. This is despite the fact that a reduction in the peak fare should attract more peak car users than a reduction in the off-peak fare, due to the assumption that the cross elasticity between modes is greater during the peak than between the peak and the off-peak period. The rationale behind low off-peak bus fares is two-fold: first, it attracts peak car users, and second, some peak bus passengers are also transferred to the off-peak, which reduces peak bus demand, which in turn decreases peak fares and therefore makes bus in the peak period more attractive for motorists.

Glaister's model was extended by Glaister and Lewis (1978) who added rail as a travel option to bus and car in a two-period framework, producing a total of six mode-period travel alternatives. Travel times by car and bus depend on both car and bus demand to account for congestion, whereas rail travel times are fixed. A system of equations for the optimal bus and rail fares are identified. In terms of estimation of elasticities, the authors use off-peak elasticities that are between two and three times the peak elasticities, based on London data, and assume that peak to off-peak elasticities are relatively low, around five percent of the corresponding within-period elasticity.

A second argument to charging a higher fare in the off-peak than in the peak appears in Jansson (1993), who argue that if, in the peak, the frequency is high enough for

passengers to arrive randomly at bus stops, and in the off-peak, the frequency is low enough for passengers to plan their trip following a timetable, the lower value of the schedule delay savings value in the latter reduces the optimal frequency and increases the occupancy rates with respect to the peak, therefore increasing the marginal cost per passenger and, consequently, the off-peak fare.

At this stage, it is worth noting that even though it is theoretically possible that optimal fares can be shown to be higher in the off-peak than in peak periods, as analysed by Glaister (1974) and Jansson (1993), such an outcome is rare in numerical applications of marginal cost pricing principles reported in the literature. Peak fares are greater than their off-peak counterparts in both second best (Glaister and Lewis, 1978) and first best (De Borger *et al.*, 1996; De Borger and Wouters, 1998) scenarios. An exception is Proost and Van Dender (2008) who estimated optimal fares for bus and metro in London and Brussels, and found that off-peak fares are lower in London but higher in Brussels, relative to the optimal peak fares. In practice, the peak fare is commonly higher than the off-peak fare due to the higher marginal cost and capacity constraints that are characteristic of peak periods in bus and rail systems, at least in large cities.

### ***2.3.2 Effect of including other externalities beyond congestion***

A relatively few authors have studied the impact of other external costs of transport on pricing and subsidy decisions. Buses and trains do cause accidents, pollution and noise, and therefore consideration of these externalities on the setting of optimal fares is justified on economic grounds.

When environmental externalities are included in first best pricing models, optimal prices increase for motorised modes, which would in turn reduce the first best subsidy calculated for public transport (Kerin, 1992). However, the second best analysis is different. Taking the case of fuel emissions, one bus is likely to pollute more than one car, but it carries more people with a single vehicle, thus reversing the result of comparing



vehicles only, i.e., the marginal external cost of car users is usually larger than that of public transport riders<sup>8</sup>, therefore it is expected that the fare premium of considering externalities other than congestion on optimal prices is greater for private than for public transport. On second best grounds, this would tend to reduce the bus fare even more and subsequently justify higher subsidies (Else, 1985).

Among the multimodal pricing models that include environmental, accident or noise externalities we can cite De Borger *et al* (1996), De Borger and Wouters (1998), Proost and Van Dender (2008), Parry and Small (2009) and Jansson (2010). Because of data limitations, models that are applied to cities usually use a constant value for the marginal accidents and environmental externality costs per kilometre, independent of traffic flow; nevertheless, it is possible to use more sophisticated models that relate pollution or accidents costs to traffic speed and flow (e.g., Shepherd, 2008). The contribution of environmental and accident externalities to optimal fares relative to the congestion externality, strongly depends on the specific application, in particular on the degree of congestion observed. It is common that in peak periods in highly urbanised areas, the marginal cost of congestion is much higher than that of other externalities, whereas in the off-peak the external costs of congestion, accidents and pollution have the same order of magnitude, as reported by De Borger *et al* (1996) for Belgium, and Parry and Small (2009) for London and U.S. cities. Therefore, we can conclude that ignoring externalities other than congestion should not have a substantial impact on fares in the peak period, but it does matter for off-peak travel.

### **2.3.3 Dedicated bus lanes**

The study of private and public transport pricing options is different if modes share the right of way or run on segregated roads. In this section, a review of studies that look at the convenience of providing segregated lanes for buses (at the cost of reducing road space for cars) is undertaken.

---

<sup>8</sup> As empirically found for pollution and accidents, but not for noise (De Borger *et al.*, 1996).

Mohring (1983) analyses the convenience of having reserved lanes for buses with first best pricing and with a number of second best scenarios, including a suboptimal toll for cars, and zero-bus-fare and zero-bus-deficit constraints. Using Minneapolis data, it is found that the travel cost savings of providing dedicated road infrastructure for buses are small when marginal cost pricing is in place (a result also obtained by Small, 1983), but considerable benefits are accruable when toll and fare constraints are present, to the point that the travel cost in a situation with exclusive bus lanes, toll and bus fare constraints, is only slightly higher than when first best pricing is implemented on mixed traffic (bus-automobile) roads. At the same time, Small (1983) finds that when total demand exceeds a threshold, the benefits of segregating a bus lane overcome the increased congestion cost for auto users.

Berglas *et al.* (1984) study first best and second best pricing for bus and automobile with and without segregated facilities for buses, optimising fares and road width in each case. They show that if travel cost decreases with road width, and the cost of separating the right of way for buses and cars is nil, the mixed traffic operation is never superior, and is more likely inferior than providing exclusive lanes for buses and cars, given that a bus passenger has a lower contribution to congestion than a car user. The congestion interaction between buses and cars suggests that when an auto traveller switches to bus, both buses and cars obtain speed gains, whereas when there are segregated busways, only car users obtain a gain in travel time (assuming that an extra bus passenger has no effect on bus travel time). This argument is used by the authors to show that under segregated operation and second best pricing, the need of subsidies for buses is reduced.

The superiority of providing exclusive bus lanes is also supported by Basso and Silva (2010), who using data from Santiago, Chile find that the provision of one bus lane on a corridor increases social welfare with respect to any scenario in which bus and car share the right of way (even when optimal pricing is applied in mixed traffic but not for exclusive bus lanes). The optimal operation with dedicated bus lanes is translated into a lower

requirement on the number of buses, a lower bus fare and higher frequency, providing large benefits for bus users. A distributional effect of the dedicated lane policy shows that the highest income group is the only segment worse off in terms of consumer surplus if this policy is applied, whereas middle and low income groups are better off.

Summing up, implementing dedicated bus road infrastructure to fight congestion problems is exposed by the extant literature as being slightly worse (Mohring, 1983) or better (Basso and Silva, 2010) than providing marginal cost pricing on mixed traffic conditions, with the extra advantage that bus lanes as a transport policy tool are likely to be more politically and socially acceptable than imposing marginal cost pricing (Mohring, 1983), a fact that is evident when comparing the number of cities in which marginal cost pricing has been implemented, versus cities with dedicated bus lanes. A limitation of all economic models on bus lanes is that they abstract from the extra cost of reserved bus lanes produced by diversions and extra delays on intersections, as some movements need to be prohibited. This consideration is likely to reduce the welfare gain estimates of segregated bus lanes but is unlikely to change the main conclusions obtained by the authors.

#### ***2.3.4 Interactions with other sectors of the economy***

The previous discussed research is based on partial equilibrium models that are concerned with the transport sector only, and therefore, abstract from the interaction between transport and other sectors of the economy. This is a significant issue because the result of a partial equilibrium model establishes, for example, the need to subsidise public transport as a way to maximise social welfare in the transport sector, but says nothing about how such a deficit should be covered, i.e., where that subsidy should come from and what are its repercussions on the wider fiscal system. In order to answer these questions, one needs to go further than a transport specific approach, into a general equilibrium model, which can be used to estimate the impact of transport pricing reforms

on the government budget, the labour market, land use, firms and so on<sup>9</sup>. Large scale general equilibrium models are usually too complex to be able to develop analytic solutions, due to the large number of interrelations between sectors in the economy that must be accounted for, and therefore need to be solved numerically as, for example, the LUSTRE model calibrated for Washington DC (Safirova *et al.*, 2007) and the TRESIS-SGEM framework under development for Sydney (Hensher *et al.*, 2012).

Things are simpler when the rest of the economy is reduced to a limited number of relevant markets, like for instance the labour market, and distortions in the economy are then collapsed into the labour tax only, as implemented in Parry and Bento (2001), who argue that public transport subsidies increase labour supply because the cost of commuting is reduced. Nonetheless, they conclude that spending road pricing revenue on cutting labour taxes directly is more welfare improving than earmarking the revenue to reduce fares. Their model assumes rail-car competition with no congestion interaction between modes, which is a relevant issue, because in a situation where there is congestion interdependence such as when buses and cars share the right of way, a transfer of passengers from cars to buses has a positive effect on bus speed due to the reduced number of cars in the road. This is a benefit from public transport subsidies that is not counted in bimodal rail-car models.

The question of how to fund public transport subsidies was first analytically addressed by Dodgson and Topham (1987), who investigated the efficiency of raising the subsidy for public transport through a rise in the tax on other goods, in particular, the property tax on land and fixed structures. The convenience of such a subsidy strongly depends on the marginal cost of public funds (MCF), which measures the welfare loss for society in raising additional revenue to finance public spending through the application of distortionary taxes (Browning, 1976; Kleven and Kreiner, 2006). The MCF increases with the initial tax

---

<sup>9</sup> For an extended discussion on the advantages of general equilibrium models, see Calthrop *et al.* (2010)

rate on the taxed good used to finance public transport; and the larger the marginal cost of public funds, the less likely that reducing fares through subsidies is welfare improving.

The MCF does depend on what tax instrument is used to increase government revenue (e.g., uniform lump sum tax, income tax). Consequently, given the objective of subsidising public transport, the welfare analysis of transport pricing policies depends on the source of the money required to cover financial deficits or investments (Proost *et al.*, 2007; Calthrop *et al.*, 2010) or how the revenue is allocated if there is a surplus.

The question therefore arises as to what is the impact of wider fiscal considerations in the calculation of optimal fares and (possibly) subsidies? It is expected that estimated public transport subsidies would decrease given that when there is no account on how the subsidies are funded, their cost is misrepresented in the social welfare analysis (Kerin, 1992). This issue can be analysed in a simple (but not complete) way that avoids dealing with general equilibrium models, by simply including the marginal cost of public funds in partial equilibrium models, as implemented in De Borger *et al.* (1996) and Proost and Dender (2008) among others, who apply a weight to net transport revenues. De Borger and Swysen (1999) analyse the impact of the MCF on the optimal bus fare, showing that the consideration of the MCF makes the optimal bus fare deviate from its marginal social cost, even when cars are priced at marginal cost. The numerical application to Brussels and London of Proost and Van Dender (2008) shows that, as expected, road prices and public transport fares increase in the presence of costly public funds, as the benefits of generating revenue to be used elsewhere in the economy (or the benefits from reducing the subsidy for public transport) are taken into account. A similar conclusion is reached by Parry and Small (2009) who suggest that fiscal considerations would decrease optimal public transport subsidies, but not to the point of jeopardising their need on second best grounds.

In summary, approaching the problem by including the MCF to the net revenues of a transport intervention in a partial equilibrium model is useful as a first approximation to answering the question of how tax distortions affect, and are affected by, reforms in the transport sector, but a full understanding of this issue still requires a general equilibrium model, since for example, the impacts of a reduced congestion on other markets (as shown by Parry and Bento, 2001, for the income tax), derived from a public transport subsidy, are not going to be captured with an approach that only considers the MCF as representing the rest of the economy.

### **2.3.5 Other public transport provision structures**

Thus far, optimal pricing rules assuming social welfare maximising behaviour by the authority or price regulator have been examined. Nevertheless, as discussed by Proost *et al.* (2007), in the real world that need not to be the case, as governments are subject to constraints and pressures from lobby groups, or simply have a different objective function (e.g., maximising votes instead of maximising welfare). The study of these issues is beyond simple economic optimisation models that are common in welfare maximising models, and falls under the regime of political economy analysis (Proost *et al.*, 2007).

Other departures from the social welfare maximisation assumption are simpler to analyse, such as the case of private operator profit maximisation and welfare maximisation subject to a budget constraint. In this section we briefly describe the main findings in the literature.

In the numerical estimation of second best fares for public transport, it is common to find very low fares as a competitive neutrality argument to compensate for the underpricing of cars, or even negative fares for buses (e.g., Ahn, 2009)<sup>10</sup>. This issue has resulted in

---

<sup>10</sup> When the in-vehicle time does not depend on the number of passengers and only the waiting time effect of frequency is accounted for, the first best bus fare is zero (Chang and Schonfeld, 1991; Ahn, 2009; Basso and Jara-Díaz, 2010), as under this assumption carrying an extra passenger is costless. Consequently the second best fare is negative, in order to attract travellers from the underpriced car.

researchers introducing a budget constraint to the welfare maximisation problem (De Borger *et al.*, 1996; De Borger and Wouters, 1998; Danielis and Marcucci, 2002; Ahn, 2009; Basso and Silva, 2010), by either setting a maximum allowable subsidy or setting public transport fare equal to average operator cost. As expected, any financial constraint reduces the social welfare relative to the unconstrained case.

Beyond constrained or unconstrained social welfare maximisation as the objective function to define optimal fares, another problem often examined is operator profit maximisation (e.g., Chang and Schonfeld, 1991; Chien and Tsai, 2007; Pels and Verhoef, 2007; Wichensin *et al.*, 2007; Ahn, 2009). The general result when comparing profit and welfare maximisation is that the profit maximising fare exceeds the welfare maximising one by a monopolistic mark-up, which in the analysis of Ahn (2009) is a function of the degree of substitution between public transport and car; in particular, the lower the cross elasticity of demand, the higher the monopolistic fare.

Proost and Van Dender (2004) compare first best pricing against the observed current situation and two alternative pricing policies: average cost pricing for all modes and Ramsey pricing<sup>11</sup> with budget equilibrium in the transport sector with respect to the reference situation. Results are obtained for several European cities using the partial equilibrium model TRENEN, and show that the more stringent a policy is in terms of budget constraints, the lower the social welfare it provides, to the point that average cost pricing (which requires cost recovery in public transport) yields welfare losses compared to the reference situation in which public transport is subsidised (even though it is not an optimal second best subsidy).

---

<sup>11</sup> The Ramsey pricing rule maximises welfare given a cost recovery constraint, and therefore optimal fares are higher than in the unconstrained case; the rule establishes that the price deviation should be inversely proportional to the price elasticity of demand (Ramsey, 1927; Hensher and Brewer, 2000; Jara-Díaz and Gschwender, 2005).

In terms of service outputs, Pels and Verhoef (2007) show that the rules for optimal frequency and capacity are the same under both profit and welfare maximisation, assuming congestion independence (car-rail model). However, this result does not hold when there is congestion interaction between cars and buses due to the effect of bus frequency on traffic congestion (Ahn, 2009). It is crucial to realise that even if the frequency rule is the same under profit and welfare maximisation, this does not necessarily mean that the frequency should be set at the same level under both regimes, as the frequency value depends on the final demand, which is likely to be price sensitive. Therefore, as the optimal fare is higher when maximising profit, it is expected that frequency is going to be lower than when maximising social welfare. This issue has gained renewed interest as van Reeve (2008) uses that the same functional form for the optimal frequency may hold under profit and welfare maximisation to suggest that a monopolistic public transport operator behaves as a social welfare maximiser, and that the service can be operated with a positive profit, therefore public transport subsidies would not be justified on first best grounds. This conclusion strongly depends on van Reeve's demand specification, as later shown by Basso and Jara-Díaz (2010) and Savage and Small (2010), who by means of a more general analysis on the price sensitivity of demand, find that the value of the optimal frequency is different under profit maximisation, an expected result when fares are different and demand is price-sensitive, as discussed before.

A different problem was addressed by Wichensin *et al.* (2007) in their analysis of car, bus and rail pricing. They compare the cases in which the public transport market behaves as a monopoly or as a non-cooperative duopoly, maximising profit in either case. They show that social welfare and consumer surplus are larger in the duopoly case due to the lower fares for both bus and rail, which also implies a higher modal share but lower profits for public transport relative to a profit maximising monopoly.



### **2.3.6 Other relevant aspects**

In this section, two elements that have not been addressed thus far are briefly discussed, namely travel distance and distributional implications of pricing. These elements are expected to play a role in the setting of optimal prices in urban transport.

The impact of travel distance on fare has been analysed in depth by Kerin (1992) and Jara-Díaz and Gschwender (2005). There is no definitive answer on whether or not fares should increase with travel distance; the outcome depends on the conditions of the trip. Turvey and Mohring (1975) show that first best bus fare can decrease with travel distance in radial corridors, as in the peak direction (towards the CBD), passengers that travel shorter distances get on buses when more passengers are already on board, which represents a greater social cost. This inverse relation between fare and travel distance might not hold if we consider that long distance passengers yield greater discomfort externalities than short distance passengers (Kraus, 1991), as the latter are less likely to have a seat available when boarding a bus, due to the presence of the former. As stated by Turvey and Mohring (1975), what matters is the marginal contribution of a passenger to the social cost of the public transport service, which may be related to travel distance in a proportional or inverse way, depending on the conditions of the trip.

On distributional and equity issues, it has been widely recognised that pricing decisions not only have welfare implications in absolute terms, but also distributional effects which are generally ignored because they greatly complicate the analytical treatment of the pricing problem. Distributional concerns emphasise the need of a general equilibrium approach, as it seems necessary to identify who is affected by, for example, an increase in the labour or land properties tax to finance and increase the subsidy for public transport (Dodgson and Topham, 1987; Proost *et al.*, 2007)<sup>12</sup>.

---

<sup>12</sup> Dodgson and Topham (1987) find that the existence of social benefits of fares subsidies financed by additional taxes depends on the income elasticities of demand for private and public transport, and for the taxed good.

## 2.4 Optimal Public Transport Supply

### 2.4.1 Basic theory and main results

In the microeconomic literature on public transport operations, the most common objective is the maximisation of social welfare, defined by the summation of the consumers and producer surplus, which is equivalent to minimising the total cost  $C_t$ , i.e., users' cost  $C_u$  plus operators' cost  $C_o$ , when demand is parametric (Jara-Díaz, 1990).

Users cost is usually divided into access time cost  $C_a$ , waiting time cost  $C_w$  and in-vehicle time cost  $C_v$ , therefore the total cost can be defined as:

$$C_t = C_o + C_u = C_o + C_a + C_w + C_v \quad (2.1)$$

The cradle of this body of literature is a paper by Mohring (1972), who attempts to find the optimal value of the service frequency for a single bus route. In Mohring's simplest model it is assumed that frequency affects waiting time, but not in-vehicle (riding) time, that the number of bus stops is held constant and that the unitary cost of operating a bus per hour is fixed,  $c$ . In this case, total cost (2.1) is minimised by providing an optimal frequency that grows with the square root of demand  $N$ :

$$f^* = \sqrt{\frac{P_w}{2t_c c}} N \quad (2.2)$$

where  $P_w$  is the value of waiting time savings and  $t_c$  is the (constant) cycle time<sup>13</sup>. This first version of the *square root formula* clearly shows the trade-off between the interests of users and operators; the value of waiting time savings and demand push frequency up (users want a high frequency to reduce waiting time), unit operator cost  $c$  acts in the opposite direction reducing frequency, as fewer buses mean a lower cost for the operator. Frequency (2.2) leads to the existence of scale economies in the optimal design of transit services, and subsequent need for subsidies under marginal cost pricing (as discussed in Section 2.3).

---

<sup>13</sup> In (2.2) it is also assumed that average waiting time is half of the average headway between two consecutive buses, which explains the factor 2 in the denominator.

An important and sometimes forgotten outcome is that the square root rule does not necessarily mean that optimal frequency depends on the square root of demand; that is a result of the first Mohring model (Mohring, 1972). Subsequent extensions with more accurate representations of the users cost function - for example, including boarding and alighting time that makes the total travel time to increase with the number of passengers - have shown that, even though the square root form is maintained, demand under the root appears to a degree higher than one, for example, the quadratic formulation in Jansson (1980) who included the boarding and alighting effect on travel time. Therefore, even though the functional form for the optimal frequency is a square root when a single route is considered, it can vary with demand to a power higher than 0.5.

The total cost minimisation approach has been subsequently extended to jointly optimise frequency and other variables, such as bus size (Jansson, 1980; Oldfield and Bly, 1988; Jara-Díaz and Gschwender, 2003a), route density (Kocur and Hendrickson, 1982; Chang and Schonfeld, 1991), the spatial structure of bus services, comparing corridors (with transfers) and direct lines to link origins and destinations (Jara-Díaz and Gschwender, 2003b), and the optimal design of alternative fleet assignment strategies, including expressing (Leiva *et al.*, 2010), short turning (Delle Site and Filippi, 1998; Tirachini *et al.*, 2011) and short turning integrated with deadheading (Cortés *et al.*, 2011). Other methodological contributions include showing that bus operating cost is proportional to its size (Jansson, 1980), that the bus occupancy rate or load factor may increase average waiting time as buses are more prone to be full (Oldfield and Bly, 1988), and that the value of in-vehicle time savings is an increasing function of the load factor, in order to account for the disutility of crowding and standing (Kraus, 1991; Jara-Díaz and Gschwender, 2003a). Some of these authors have been able to arrive to closed forms for the optimal values of the two more common optimisation variables, bus frequency and bus size (e.g., Jansson, 1980; Chang and Schonfeld, 1991; Cortés *et al.*, 2011), whereas in other cases the mathematical complexity of the models require numerical methods to locate a solution (e.g., Delle Site and Filippi, 1998; Leiva *et al.*, 2010).

### **2.4.2 Public transport supply and road pricing**

A relevant issue for the economic analysis of pricing options is the determination of the optimal change in public transport frequency and capacity when road pricing is introduced. The answer is not straightforward; for instance, Jansson (2010) finds that bus frequency, when car travel is underpriced, should be lower than when marginal cost road pricing is in place, due to the negative impact of frequency on the environment, and excessive congestion derived from the greater than optimal car traffic. However, the bimodal rail-car analysis of Kraus (2003) concludes that both rail frequency and capacity should increase if cars are underpriced, assuming no congestion interaction and disregarding the environmental cost associated with rail, assumptions that are relaxed in the model derived in Chapter 3. The existing literature does not offer unambiguous evidence for the direction of change in frequency and capacity of public transport after applying road pricing; indeed the outcome seems to depend on the modelling assumptions. Bus (and rail) frequency should be increased with congestion pricing in situations where the expected modal switching (given the relevant cross price elasticities) might lead to a shortage of service capacity, at least in peak periods. The anticipation of modal switching in London and Stockholm delivered increased buses in advance of the application of cordon pricing, which was used to show that the revenue raised from the congestion charge was being hypothecated back to the transport sector for the benefit of modal switchers.

### **2.4.3 When the capacity constraint is binding**

Transport capacity on a public transport route is given by the product of the frequency of service and the capacity of vehicles. This transport capacity sets the maximum flow that the service is able to accommodate in a given period of time. If the maximum passenger flow equals the transport capacity, it is said that the capacity constraint is binding, and, as expected, this has an influence on the determination of outputs like optimal fare and frequency, because in this case an increase in demand must be met by an increase in supply (e.g., more vehicles, bigger vehicles).

To our knowledge, the first transport pricing study in which capacity considerations are accounted for is Glaister (1974), who solves the second best problem with a constraint that limits the amount of travel in the peak to the capacity of the bus system, finding that a shadow price of capacity, i.e., the extra social benefit achieved if capacity is increased by one unit, should be incorporated into the bus fare. Glaister does not provide an expression for the shadow price of capacity because capacity is not an optimisation variable in his model. As an aside, he concludes that if the capacity constraint is active in the peak, but not in the off-peak period, it is more likely that the optimal peak fare will be greater than the off-peak counterpart.

It has been subsequently shown that when transport capacity is optimised together with the fare, the shadow price of capacity is identified as a function of users and operators cost parameters (Pedersen, 2003; Small and Verhoef, 2007). For example, when the optimal frequency cannot accommodate total demand, and needs to be consequently increased, there is a positive effect on users cost and a negative effect on operator cost, that should somehow show up in the optimal fare, as shown by Pedersen (2003) using a public transport demand and supply model in which transport capacity is an optimisation variable.

Another argument to increase bus fares when the capacity constraint is binding is provided by Turvey and Mohring (1975), who argue that higher fares should be levied when buses run full (or close to full), as this increases the probability of passengers not being able to board the first bus that arrives at their stop, and having to wait for one or more buses to continue their trip.

In conclusion, transport capacity appears to play a role in increasing both first best and second best fares when the system is operating at capacity. Nevertheless, the fact that the capacity constraint is binding does not necessarily mean that the provided frequency and bus size are not optimal. This issue will be addressed in Chapter 3, where it is shown that

what really matters is not whether the capacity constraint is binding or not, but whether the supplied transport capacity is optimal. In other words, the first best fare equals marginal total cost minus average users cost when the transport capacity is optimal, even if the capacity constraint is binding (disregarding Turvey and Mohring' argument of an increased probability of buses not stopping to collect passengers).

## **2.5 Summary**

The basic theory on urban transport pricing and public transport optimisation has been reviewed, highlighting the main methodological contributions found in the literature. The review indicates that there are sufficient theoretical grounds to set public transport fares below average operator costs, and therefore an optimal subsidy seems justified on first best grounds, second best grounds, or both. Nevertheless, the answer as to which is the appropriate level of fare and subsidy does depend first on the modelling approach (first or second best, what externalities are included, possibility of day of time substitution), and second, on the actual context or city. As a result, estimated optimal bus fares vary from negative figures to values that actually cover operating cost (without considering capital investment). The optimisation of bus frequency and size is also extensively discussed, with reference to elements that influence their optimal level such as active capacity constraints and the setting of road pricing.

This thesis adds to this body of literature by analysing the influence of non-motorised transport on public transport pricing, introducing new decision variables like the choice of a fare collection technique and level of infrastructure investment for bus corridors, and by providing a more comprehensive methodological framework for the introduction of bus congestion and passengers crowding in the optimisation of bus services and setting of fare and road pricing. These elements are presented in the next chapters.

## Chapter 3

# Multimodal Transport Pricing: the Influence of Non-motorised Modes<sup>14</sup>

### 3.1 Introduction

Second best pricing models that take into consideration only two modes - cars and public transport (bus or rail) - have found that subsidies for public transport are desirable, with fares below marginal cost due to the underpricing of cars. However, as put forward by Kerin (1992), this approach neglects the existence of other modes, notably walking and cycling, that play an important and growing role in urban transport systems, especially for short trips. Disregarding non-motorised transport is a growing concern because low bus fares not only deter some drivers from using their cars, but also divert walkers and cyclists onto trains or buses, which is not necessarily a desirable outcome. As such, a pricing model that also includes non-motorised transport seems desirable in order to estimate the impact of these modes on (possibly decreasing) optimal subsidies for public transport. Even though there are no analytical models that address the issue of the influence of non-motorised transport on urban transport pricing policy, we do find that walking and cycling are considered as travelling alternatives in applied models (Safirova *et al.*, 2006; Proost

---

<sup>14</sup> The multimodal pricing model presented in this chapter is published in Tirachini and Hensher (2012).

and Van Dender, 2008), but no attempt is made to identify how the design of the pricing instrument would change by considering or ignoring walking and cycling.

In this chapter, a multimodal pricing model is developed, including three modes - automobile, public transport (either bus or rail) and non-motorised transport (either walking or cycling), with the objective of maximising social welfare. This model extends the previous literature by identifying the role that non-motorised transport can play in the optimal setting of fares for public transport. Emphasis is given to the effect of bus demand on car congestion when both modes share the right of way, and the way in which the optimal fare, frequency and vehicle size should be determined when the capacity constraint is binding for a public transport service, i.e., when demand meets the capacity offered by the operator (see Section 2.4.3). We also include in the framework the cost of externalities other than congestion, such as accidents, pollution and noise, and the toll collection cost, all of which increase the marginal cost of motorised transport compared with walking and cycling alternatives. The emphasis of this chapter is not on the determination of the empirical value for optimal fares and subsidies (where applicable) but with the economic principles behind them<sup>15</sup>.

With reference to the outcomes of this multimodal pricing framework, it is shown that the effect of considering non-motorised transport alternatives on optimal public transport fares depends on the demand substitution between modes; the stronger is the demand substitution between public transport and non-motorised modes, relative to the substitution between car and public transport, and car and non-motorised modes, the more likely it is that a higher public transport fare would result from the allowance for the role of walking or cycling on fare setting. On the other hand, a capacity constraint on public transport plays a role in optimal pricing only when the transport capacity cannot be set at its optimal level. Finally, the internalisation of externalities other than congestion is

---

<sup>15</sup> For numerical comparisons on fares and subsidies among several studies, see Proost and Van Dender (2008) and Parry and Small (2009).



likely to increase optimal fares and road charges, therefore increasing the generalised cost of motorised transport modes relative to a non-motorised alternative.

The remainder of the chapter is organised as follows. Definitions, assumptions and formulation of the social welfare maximisation model are presented in Section 3.2. The first best and second best problems are solved and analysed in Sections 3.3 and 3.4, respectively. Section 3.5 extends the model by including external costs other than congestion and toll collection costs into the social welfare objective function. Finally, a summary and the main conclusions are given in Section 3.6.

### **3.2 Model Assumptions**

Consider a single origin-destination pair and three modes: automobile (*a*), public transport (*b*) that could be a bus or rail based mode, and a non-motorised mode (*e*) that could be walking or cycling. At this point it is necessary to distinguish between non-motorised modes as being complementary or an alternative to motorised modes; walking is commonly an access and/or egress mode in a trip chain that includes driving or riding a bus or train, in which case the modes are complementary. In this model, it is assumed that walking or cycling are a (linehaul) mode, i.e., are an alternative to choosing a motorised mode (walking and cycling as an access mode is included into the motorised alternatives).

The competitiveness of walking and cycling is mainly associated with trip distance and factors like steepness of (some part of) the route, weather, availability of safe walking and cycling facilities, etc. In all situations, walking as a substitute to motorised modes typically declines as distance increases, for example, in Sydney 65.8 percent of trips shorter than one km are walking-only trips, a fraction that is 23.6 percent for trips between one and two km, and 5.4 percent for trips between two and five km (TDC, 2010). Donoso *et al.* (2006) reports a similar pattern for Santiago, Chile. As such, there is a (location specific) distance range in which walking is an alternative for motorised modes.

Without loss of generality, it is assumed in the model that the public transport mode is a bus that shares the right of way with cars, resulting in congestion dependence between the two modes. The case of trains or buses running on segregated busways is a particular case of the above, derived after assuming congestion independence between modes, as usually assumed by researchers that address the rail-car pricing problem (Tabuchi, 1993; Arnott and Yan, 2000; Pels and Verhoef, 2007). The decision variables are optimal prices for both automobile and public transport, and frequency and size (capacity) of the public transport mode. We consider only one period of operation<sup>16</sup>, which allows us to find a closed form formulae for the optimal prices of automobile and public transport to shed light on the impact of non-motorised transport and capacity constraints. Road capacity is fixed and tax distortions are ignored (see Section 2.3.4).

We follow much of the notation of Small and Verhoef (2007). Ignoring income effects, the joint demand for the three modes can be obtained from the benefit function  $B(q_a, q_b, q_e)$ , which expresses the consumers' willingness to pay for a particular combination  $\{q_a, q_b, q_e\}$  of travel by automobile, public transport and non-motorised mode. The inverse demand function  $D_i$  for mode  $i$  is given by:

$$D_i(q_a, q_b, q_e) = \frac{\partial B(q_a, q_b, q_e)}{\partial q_i} \quad i \in \{a, b, e\} \quad (3.1)$$

Let  $C_i$  and  $c_i$  be the total and average cost functions of mode  $i$  respectively (including both time and operation costs), that is:

$$C_i = q_i c_i \quad (3.2)$$

Let  $c_a(q_a, q_b, f_b, K_b)$  and  $c_b(q_a, q_b, f_b, K_b)$  be the average cost of car and bus travel, respectively. Further, it is assumed that these cost functions depend on demand  $q_a$ , bus

---

<sup>16</sup> Examples of multiperiod analyses are Glaister (1974), Glaister and Lewis (1978), De Borger *et al.* (1996) and Proost and Van Dender (2008).

frequency  $f_b$  and capacity  $K_b$  (related to bus size), and the activity of buses at bus stops, which is given by  $f_b$ , demand  $q_b$  and capacity  $K_b$  if dwell time increases with crowding. The relationship between car demand  $q_a$  and car flow  $f_a$  is  $f_a = v_a q_a$ , where  $v_a$  is the inverse of the average occupancy rate per car<sup>17</sup>. Bus cost  $c_b$  includes users cost  $c_u$  (access, waiting and in-vehicle time costs) and operator cost  $c_o$  (which depends on bus frequency and size); hence

$$c_b = c_u + c_o \quad (3.3)$$

We assume that the travel time associated with walking or cycling is fixed and independent of demand or flow of any mode, i.e., the non-motorised mode is uncongestible.

In equilibrium, the marginal benefit is equal to the generalised price,  $c_a + \tau_a$  and  $c_u + \tau_b$  for cars and public transport, respectively (equation 3.4), where  $\tau_a$  is the road use charge for the auto occupant and  $\tau_b$  is the fare for public transport.

$$\frac{\partial B}{\partial q_a} = c_a + \tau_a \quad \frac{\partial B}{\partial q_b} = c_u + \tau_b \quad (3.4)$$

The social welfare function  $SW$  (3.5) is maximised subject to a capacity constraint for public transport vehicles, given by expression (3.6), which states that the transport capacity  $f_b K_b$  must be sufficient to carry demand  $q_b$ .

$$SW = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e \quad (3.5)$$

$$q_b \leq f_b K_b \quad (3.6)$$

---

<sup>17</sup> We assume that the occupancy rate does not change with pricing reforms, i.e., we ignore the possibility of car-pooling if road price increases.

### 3.3 First Best Pricing

To solve the constrained maximisation problem (3.5)-(3.6), we set the Lagrange function  $L$  given by (3.7).

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda [f_b K_b - q_b] \quad (3.7)$$

where  $\lambda$  is the Lagrange multiplier associated with constraint (3.6), i.e. the marginal social benefit of increasing bus transport capacity by one unit.

After applying first order conditions (see Appendix A1) we find:

$$\tau_a = q_a \frac{\partial c_a}{\partial q_a} + q_b \frac{\partial c_b}{\partial q_a} \quad (3.8)$$

$$\tau_e = 0 \quad (3.9)$$

Equation (3.8) is the well-known Pigouvian tax for cars, including here the marginal cost on bus due to car demand (second term), whereas (3.9) shows that the price for walking or cycling is zero (the uncongestible mode).

The solution for the optimal bus fare, frequency and capacity depends on whether or not the capacity constraint (3.6) is binding.

*Case 1: Capacity constraint is not binding*

In this case  $\lambda = 0$  and the optimal fare is obtained as (3.10):

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b} \quad (3.10)$$

The optimal frequency and capacity are obtained by solving the following system of equations:

$$q_a \frac{\partial c_a}{\partial f_b} + q_b \frac{\partial c_b}{\partial f_b} = 0 \quad (3.11a)$$

$$q_a \frac{\partial c_a}{\partial K_b} + q_b \frac{\partial c_b}{\partial K_b} = 0 \quad (3.11b)$$

*Case 2: Capacity constraint is binding*

In this case constraint (3.6) is active, i.e.,  $q_b = f_b K_b$  and the Lagrange multiplier is  $\lambda \neq 0$ .

From Appendix A1 (equation A1.3):

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b} + \lambda \quad (3.12)$$

From equation (A1.5), the marginal welfare benefit of capacity can be expressed as (3.13):

$$\lambda = \frac{1}{K_b} \left[ q_a \frac{\partial c_a}{\partial f_b} + q_b \frac{\partial c_b}{\partial f_b} \right] \quad (3.13)$$

and using that  $\frac{1}{K_b} = \frac{f_b}{q_b}$  we obtain:

$$\tau_b = c_o + q_a \frac{\partial c_a}{\partial q_b} + q_b \frac{\partial c_b}{\partial q_b} + f_b \left[ \frac{q_a}{q_b} \frac{\partial c_a}{\partial f_b} + \frac{\partial c_b}{\partial f_b} \right] \quad (3.14)$$

Equation (3.14) shows the effect of the capacity constraint on the optimal bus fare. A similar result was obtained by Pedersen (2003) in a model with no car-bus interactions. When the capacity constraint is binding, one possibility is to increase the frequency to satisfy constraint (3.6) to a higher than optimal value. In that case the term in brackets in (3.14) is positive and represents the impact on car and bus marginal cost of the increased frequency necessary to deal with a demand that the optimal frequency (solution of equation 3.11a) cannot meet. Nevertheless, note that frequency and capacity can be optimal and the capacity constraint can indeed be binding, if for example there is no extra benefit of providing excess capacity (no crowding or comfort costs) and therefore, once the frequency has been optimised, the vehicle size is obtained as the minimum value that

satisfies (3.6). In this case, expression (3.14) is valid, but the capacity related term in brackets is zero because frequency is optimal (solution of equation 3.11a), and then (3.14) is reduced to the optimal fare (3.11) with no capacity constraints. Therefore, an important outcome of equation (3.14) is that what really matters when setting optimal fares is not if the capacity constraint is binding, but whether or not the operator provides the optimal transport capacity.

### 3.4 Second Best Pricing

We can solve the same problem assuming that there is no road price for cars, i.e.,  $\tau_a = 0$ .

The Lagrange function is:

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda [f_b K_b - q_b] + \gamma_a \left( c_a - \frac{\partial B}{\partial q_a} \right) + \gamma_b \left( c_u + \tau_b - \frac{\partial B}{\partial q_b} \right) + \gamma_e \left( c_e - \frac{\partial B}{\partial q_e} \right) \quad (3.15)$$

The first order conditions are given in the Appendix A1. We can simplify the differential notation as follows:

$$\frac{\partial^2 B}{\partial q_i \partial q_j} = \frac{\partial^2 B}{\partial q_j \partial q_i} \equiv B_{ij} \quad (3.16a)$$

$$\frac{\partial c_i}{\partial q_j} \equiv c_{ij} \quad (3.16b)$$

$B_{ij}$  is the derivative of the inverse demand  $d_i$  given in equation (3.1) with respect to  $q_j$ .

That is,  $B_{ij}$  measures a marginal change in willingness to pay for mode  $i$  due to a marginal change in the amount of travel on mode  $j$ . If there is no substitution between two modes, then  $B_{ij} = 0$ . If all modes are substitutes (e.g., an increase in bus fare would increase the amount of car and non-motorised travel), then  $B_{ij} \leq 0 \quad \forall i, j$ . On the other hand,  $B_{ij} > 0$  implies that  $i$  and  $j$  are complements. Moreover, as Kraus (2003) discusses, following standard microeconomic theory for utility maximising consumers, it should hold that

(assuming that trip demand is independent of income)  $B_{ii} \leq 0$  and  $B_{ii}B_{jj} > B_{ij}B_{ji}$  for any modes  $i$  and  $j$ .

*Case 1: Capacity constraint is not binding*

After some algebraic manipulation we obtain the second best bus fare  $\tau_b^{SB}$  as expressed in (3.17).

$$\tau_b^{SB} = \tau_b - (q_a c_{aa} + q_b c_{ba}) \frac{c_{ab} - B_{ab} + \frac{B_{ae}B_{be}}{B_{ee}^2}}{c_{aa} - B_{aa} + \frac{B_{ae}}{B_{ee}}} \quad (3.17)$$

where  $\tau_b$  stands for the expression for the first best fare in (3.10). Unlike the first best pricing rule, under the second best rule, the non-motorised mode plays a role through the substitution parameters  $B_{ae}$ ,  $B_{be}$  and  $B_{ee}$ . Note that if car is an uncongestible mode and does not interact with buses, then the second best correction is zero (second term at the right hand side of equation 17), and consequently the second best fare is equal to the first best fare,  $\tau_b^{SB} = \tau_b$ , analogous to a two-link road pricing analysis when one link is uncongestible (e.g., Knight, 1924; Verhoef *et al.* 1996).

Two new results can be derived from equation (3.17). First, if we assume that there is no substitution between modes  $a$  and  $e$ , and  $b$  and  $e$ , then  $B_{ae} = B_{be} = 0$ , and (3.17) is reduced to

$$\tau_{b0}^{SB} = \tau_b - (q_a c_{aa} + q_b c_{ba}) \frac{c_{ab} - B_{ab}}{c_{aa} - B_{aa}} \quad (3.18)$$

which is the second best bus fare considering only two modes, as obtained by Small and Verhoef (2007) for the case in which there is no congestion interaction between modes, i.e.,  $c_{ab} = c_{ba} = 0$ , and by Ahn (2009) who considered that bus demand does not affect car travel time, that is  $c_{ab} = 0$ . If  $c_{ab} = 0$ , the second best bus fare equals the first best price

( $\tau_{b0}^{SB} = \tau_b$ ) when there is no cross demand elasticity between car and bus, i.e., when  $B_{ab} = 0$ , and therefore a low bus fare has no effect on mode shifting, as noted by Small and Verhoef (2007) and Ahn (2009). Nevertheless, when delays related to bus passenger activities affect cars ( $c_{ab} \neq 0$ ), the second best fare (18) is not reduced to the first best fare (3.10) even if  $B_{ab} = 0$  (noting that this does not mean that the second best fare decreases with  $c_{ab}$  because  $c_{ab}$  increases the first best fare  $\tau_b$  in equation 3.18, as shown in equation 3.10).

Second, equation (3.17) can be used to formally assess Kerin (1992)'s claim that second best fares obtained by considering car and public transport only are likely to be lower than optimal if the analysis is extended to walking and cycling. A comparison between (3.17) and (3.18), assuming for illustrative purposes that demand and congestion levels are the same, indicates that the second best bus fare will be larger when considering non-motorised transport if:

$$\tau_b^{SB} > \tau_{b0}^{SB} \Leftrightarrow \frac{B_{be}}{B_{ae}} > \frac{c_{ab} - B_{ab}}{c_{aa} - B_{aa}} \quad (3.19)$$

i.e., the larger the value of  $B_{be}$  and the lower  $B_{ae}$  and  $B_{ab}$  (in absolute values), the more likely is (3.17) to be greater than (3.18). The intuition behind this result is that *if* the modal substitution between public transport and non-motorised modes ( $B_{be}$ ) is large relative to the substitution between car and public transport ( $B_{ab}$ ) and car and non-motorised modes ( $B_{aa}$ ), a lower public transport fare attracts more passengers that would otherwise be walking or cycling than driving, at least in relative terms. However, note that the change could be in either direction ( $\tau_b^{SB} < \tau_{b0}^{SB}$ ), if the modal substitution between automobile and non-motorised transport is stronger than between public transport and non-motorised modes (low value of  $B_{be}/B_{ae}$ ), the opposite result will ensue. Certainly, the final outcome depends on trip distance, since for long trips cycling



and walking are unlikely to be an option (as discussed in Section 3.2 when we analysed modal split per trip distance), which means  $B_{ae} = B_{be} = 0$  and the analysis can be reduced to motorised modes only.

Optimal frequency and bus capacity are the solution of equations (3.20).

$$(q_a - \gamma_a) \frac{\partial c_a}{\partial f_b} + q_b \frac{\partial c_b}{\partial f_b} = 0 \quad (3.20a)$$

$$(q_a - \gamma_a) \frac{\partial c_a}{\partial K_b} + q_b \frac{\partial c_b}{\partial K_b} = 0 \quad (3.20b)$$

with  $\gamma_a = \frac{q_a c_{aa} + q_b c_{ba}}{c_{aa} - B_{aa} + \frac{B_{ae}^2}{B_{ee}}}$

That is, in the second best case the congestion externality of buses to cars is less internalised because  $\gamma_a > 0$ , as commented by Ahn (2009), the intuition being that due to the underpricing of cars, the negative effect of buses on car travel time should be weighted less. If there is no congestion interaction, i.e.,  $\partial c_a / \partial f_b = 0$ , the rules for first-best and second-best frequency and bus capacity are the same (equations 3.11 and 3.20), then a higher bus demand  $q_b$  in the first best (due to the pricing of cars) would make the first best frequency higher than the second best one, which is not a straightforward result with cross congestion, due to the presence of  $\gamma_a$  in equation (3.20a).

#### *Case 2: Capacity constraint is binding*

Analogously to the first best case, the second best bus fare is obtained as (3.21), in which the marginal welfare benefit of capacity has the term  $(q_a - \gamma_a)$  discussed in the previous paragraph:

$$\tau_b^{SB} = \tau_b - (q_a c_{aa} + q_b c_{ba}) \frac{c_{ab} - B_{ab} + \frac{B_{ae} B_{be}}{B_{ee}}}{c_{aa} - B_{aa} + \frac{B_{ae}^2}{B_{ee}}} + f_b \left( \frac{q_a - \gamma_a}{q_b} \frac{\partial c_a}{\partial f_b} + \frac{\partial c_b}{\partial f_b} \right) \quad (3.21)$$

### 3.5 Extensions: Other External Costs and Collection Costs

In this section, the preceding approach is extended by including more cost components, namely toll collection costs and external costs such as accidents, pollution and noise. The toll collection and operator costs are usually disregarded from the formal analysis of pricing policies, even though current road pricing schemes show that they are not negligible; operating costs account for 7 percent of the revenues in Singapore, 25 percent in Stockholm, and 48 percent in London (May *et al.*, 2010), mostly influenced by the choice of technology for charging and enforcement<sup>18</sup>. A simple way to include operating costs  $OC$  is proposed in (3.22).

$$OC(q_a) = oc_0 + oc_1 q_a \quad (3.22)$$

$oc_0$  is a fixed cost and  $oc_1$  is the marginal cost per transaction. The fare collection cost for public transport is partially included in the bus or rail operator cost  $c_o$  (equation 3.4), which may include the fixed collection cost due to software requirements plus fare payment devices at stations or vehicles (an in-depth analysis is provided in Chapters 4 and 7). The cost per transaction, if not negligible, can be incorporated in the same way as (3.22).

---

<sup>18</sup> In the case of London, other authors present higher estimates of operating costs. Prud'homme and Bocarejo (2005) estimate that in 2003 the London congestion charging scheme's operating costs were 85 percent of toll revenue, and net revenue would not be enough to cover the annualised capital cost. Mackie (2005), Santos and Schafer (2004), and Santos (2005) are more optimistic; they conclude that the operating cost was respectively 75 percent, 72 percent, and 53–60 percent of the net revenue.

External costs  $EC$  other than congestion (Section 2.3.2) can be expressed as follows:

$$EC(q_a, q_b, f_b, K_b) = \nu_a q_a EC_a(q_a, q_b, f_b, K_b) + f_b EC_b(q_a, q_b, f_b, K_b) \quad (3.23)$$

$EC_a$  and  $EC_b$  are the external cost rate per vehicle for car and public transport (assuming the external costs of walking or cycling as zero), and the car flow is  $f_a = \nu_a q_a$  where  $\nu_a$  is the inverse of the average occupancy rate per car, as previously defined. Expressions (3.22) and (3.23) can be subtracted from the social welfare formula (3.5) to derive first best and second best pricing results. Denoting  $EC_{ij} \equiv \partial EC_i / \partial q_j$  the result for the first best prices are:

$$\tau_a = q_a c_{aa} + q_b c_{ba} + oc_1 + \nu_a q_a EC_{aa} + \nu_a EC_a + f_b EC_{ba} \quad (3.24)$$

$$\tau_b = c_o + q_a c_{ab} + q_b c_{bb} + \nu_a q_a EC_{ab} + f_b EC_{bb} \quad (3.25)$$

Since external costs other than congestion are assumed positive for car and bus users, it is likely that the result of expressions (3.24) and (3.25) will be greater than the optimal prices when considering only congestion externalities (equations 3.8 and 3.10), and therefore, the internalisation of accidents, noise or pollution costs would increase the generalised cost of motorised transport modes compared with non-motorised modes (although the final result depends on the sensitivity of demands  $q_a$  and  $q_b$  to price), and reduce the amount of subsidy for public transport on first best grounds. The second best analysis can be undertaken in the same fashion. Regarding the toll collection costs, only the marginal cost per transaction  $oc_1$  shows up in the optimal toll (3.24), however, the fixed cost of collection  $oc_0$  in (3.22) is accounted for in the calculation of social welfare; furthermore,  $oc_0$  may be so high that the total collection cost is larger than the welfare gain from the internalisation of the external cost, in which case tolling is not welfare improving unless a more cost effective way of collecting tolls is implemented.

### 3.6 Summary and Conclusions

This chapter presents a three-mode pricing model that reveals the effect of considering non-motorised transport alternatives on optimal public transport fares. Specifically, it shows that the change in the optimal fare due to the inclusion of non-motorised modes depends on the demand substitution between modes; the stronger the demand substitution between public transport and non-motorised modes is (relative to the substitution between automobile and public transport, and automobile and non-motorised modes), the more likely it is that a higher optimal public transport fare would result when considering walking or cycling on fare setting.

We revisited the role of a capacity constraint in public transport service provision, which suggested that a capacity constraint plays a role in optimal pricing only when the transport capacity cannot be set at its optimal level. We also presented a way to include externalities other than congestion and toll collection costs into the analysis of optimal pricing under first and second best rules, which showed that the internalisation of externalities other than congestion is likely to increase optimal fares and road charges, therefore increasing the generalised price of motorised transport modes relative to a non-motorised alternative.

## **Chapter 4**

# **Fare Collection Systems and Bus Boarding Time: Operational and Economic Effects**

### **4.1 Introduction**

A significant part of the total running time of buses is spent at stops and stations in the process of boarding and alighting of passengers. Understanding the nature of this process has potential benefits for both users and operators; if after a detailed characterisation of the time that a bus is stopped transferring passengers, recommendations can be made to reduce it. A possible reduction in this time can be translated into cost savings for the operator, if the total running time is reduced by a noticeable margin, and benefits for users as well, perceived as a reduction in their overall travel time, a benefit that can be monetised using the users' value of travel time savings.

The existing economic literature on bus transport considers that bus travel or in-vehicle time is either fixed or increases with the dwell time, i.e., time spent transferring passengers at bus stops (e.g., Mohring, 1972, Jansson, 1980), which in its simplest formulation is presented as the passengers' boarding and alighting time plus the time necessary to open and close doors. When the dwell time is considered as a variable, and consequently the total travel time depends on demand, analysts have assumed the average boarding and alighting time per passenger is exogenously given, thus ignoring

that currently there are several alternative boarding and alighting policies and technological options for fare collection, which have an impact on travel times, operator costs, the complexity and image of the public transport service.

Today, agencies face the challenge to decide on a fare payment system to operate new services, or whether it is worth upgrading from a slow old-fashioned method to a quicker, more efficient technology. Observed practice shows an evolution from cash payments to the driver on-board buses, to the use of paper based coupons or tickets, magnetic strip cards, and the latest smart cards and payment by SMS text messages. On the other hand, high-standard Bus Rapid Transit (BRT) systems like that implemented in Curitiba and Bogotá have shown that, in relation to fare payment, an efficient method of operating a bus service is to collect the fare at bus stations instead of on the vehicles themselves, allowing passengers to enter buses at all available doors, leading to considerable savings in dwell time and cost. Consequently, the boarding and alighting process and the fare payment technology can be regarded as decision variables from the design stage of a bus route or network, as proposed and implemented in this thesis. This will have a significant effect on the optimal level of outputs like frequency, bus size and fare, as shown and discussed in Chapter 7.

The effect of upgrading the fare payment system on the performance and cost of urban bus services has received limited attention in the literature. Bertini and El-Geneidy (2004) estimate time savings from reducing the passenger boarding time by one second, but there is no calculation of how that one-second saving may be achieved. In this chapter, we estimate the boarding time savings that are achievable by upgrading the fare collection system; therefore calculations of total travel time savings or changes in operating speed have an empirical basis<sup>19</sup>.

The rest of the chapter is organised as follows. Section 4.2 presents the fare collection technologies and boarding and alighting rules that will be analysed. Boarding and

---

<sup>19</sup> Other works that provide bus boarding times with alternative fare collection systems (including on-board and off-board payment options with cash and prepaid cards) are York (1993), TRB (2003), Balcombe *et al.* (2004) and Wright and Hook (2007).

alighting times are estimated in Section 4.3. In Section 4.4 functions for the average boarding and alighting time are estimated, which depend on the technology of payment, boarding and alighting policy and number of doors per bus. An empirical study of travel time savings and benefits that are achievable by upgrading the fare collection system is performed in Section 4.5. Conclusions are given in 4.6.

## 4.2 Definition of Alternatives

Alternative fare payment systems and bus boarding rules differ in requirements of infrastructure, ability to integrate fares across routes and modes, security, operating cost including transaction costs and evasion control, capacity to handle different fare structures (e.g., flat, zonal, distance-based or time-based fares), level of institutional arrangement, complexity of use, level of detail on demand information recorded, image of the public transport service and capacity to attract new users, among other factors<sup>20</sup> All these features should be weighed up when deciding on a fare payment system for a specific bus service or network. In this study, we focus on differences on travel time and operator costs. To this end, we estimate the average boarding time of several alternative fare payment systems and boarding policies, including on-board payment with cash, magnetic strip, contactless card, and off-board payment.

Following Wright and Hook (2007), we distinguish between *fare collection* (i.e., payment of the fare) and *fare verification* (i.e., confirmation that the fare has been actually paid). Four fare payment systems are analysed:

- i. On-board fare collection, cash payment to the bus driver (hereafter, “cash”)
- ii. Off-board fare collection and on-board fare verification using a magnetic strip (hereafter, “magnetic strip”). This ticket must be inserted in a verification device inside buses. See example in Figure 4.1a
- iii. Off-board fare collection and on-board fare verification using a contactless smart card (hereafter, “contactless card”). See example in Figure 4.1b.

---

<sup>20</sup> For a detailed discussion on factors to consider when choosing a fare collection system for a bus service, see Section 12.2 of Wright and Hook (2007).

- iv. Off-board fare collection and off-board fare verification using a contactless smart card (hereafter, “off-board payment”). This fare payment policy has been introduced in Bus Rapid Transit (BRT) systems like Bogotá and Curitiba.



(a) Magnetic strip

(b) Contactless card

**Figure 4.1: Examples of on-board fare verification systems**

(Source: left photo by the author; right photo from Wright and Hook, 2007)

We consider buses of several sizes, with one, two, three or four doors. Buses with more than one door can be operated with two alternative boarding and alighting policies regarding the use of doors, as follows:

- Boarding is allowed only at the front door while alighting takes place simultaneously through the back door(s). If on-board card payment is required (alternatives i, ii and iii above), two card readers are provided next to the front door only. Denoting the number of doors per bus as  $n$ , this fare payment policy is referred to as  $TnB1$  (Total number of doors= $n$ , Boarding=1)
- All doors are available for both boarding and alighting in a sequential way (regardless of the number of doors). In such cases, it is necessary to have card readers next to each door when payment is inside buses. This payment policy is referred to as  $TnBn$ .

For the analysis of cash payment (case i above), we assume that boarding is performed at the front door only, regardless of the number of doors of a bus (i.e., policy  $TnB1$ ). The case of on-board cash payment with boarding at multiple doors is not considered



because the data used to estimate the average boarding time in this case are obtained from buses in Sydney, in which cash payments are handled by the driver. For on-board payment with magnetic strip or contactless card (cases ii and iii), both  $TnB1$  and  $TnBn$  boarding policies are available. Finally, for off-board payment (case iv), only boarding at multiple doors ( $TnBn$ ) is considered, as the purpose of off-board fare collection is using all available bus doors to board and alight. The scenarios to be considered are summarised in Table 4.1.

**Table 4.1: Summary of fare payment and bus boarding alternatives**

Payment Technology	Boarding front door	Boarding all doors
On-board, cash	Yes	
On-board, magnetic strip	Yes	Yes
On-board, contactless card	Yes	Yes
Off-board, contactless card		Yes

The average boarding and alighting times per passenger are estimated using data from Sydney for the cases with cash, magnetic strip and off-board payment, whereas the boarding time with contactless card is obtained from a study from Santiago (Fernández *et al.*, 2009).

### 4.3 Estimation of Boarding and Alighting Times with Alternative Fare Payment Technologies and Boarding Policies<sup>21</sup>

#### 4.3.1 Background: dwell time models

In this section we introduce the fundamentals of dwell time models, which will be used on the estimation of boarding and alighting times with the alternative fare payment systems and boarding and alighting policies summarised in Table 4.1.

There are a number of studies that have analysed the determinants of the time that a public transport vehicle spends at stops or stations. The standard procedure is to use

---

<sup>21</sup> This section is partially reproduced in Tirachini (2011).

multiple regression models estimated through a series of observations which record the time a bus is stopped at stops and the number of passengers boarding and alighting, with different levels of detail regarding payment method, door used or type of bus. While some authors model the total time a bus is stopped at stops<sup>22</sup> (York, 1993), others use only the time in which the doors are open, that is, the passenger service time (boarding and alighting) plus the time necessary to open and close doors (Lin and Wilson, 1992; Gibson *et al.*, 1997; Dueker *et al.*, 2004; Fernández *et al.*, 2009, among others). In this study we use this second concept, referred to as bus dwell time. The simplest model of dwell time is a linear function of the number of passengers boarding and alighting. The specification depends on whether the processes of boarding and alighting are sequential (the same door is used to board and alight) or simultaneous (different doors to board and alight). Dwell time  $t_d$  can be expressed as (4.1) for a sequential process, or (4.2) for a simultaneous process:

$$t_d = c_{oc} + \sum_{i=1}^{\lambda^-} t_i^a + \sum_{j=1}^{\lambda^+} t_j^b \quad (4.1)$$

$$t_d = c_{oc} + \max \left\{ \sum_{i=1}^{\lambda^-} t_i^a, \sum_{j=1}^{\lambda^+} t_j^b \right\} \quad (4.2)$$

where  $c_{oc}$  is the dead time,  $\lambda^-$  the number of passengers alighting,  $\lambda^+$  the number of passengers boarding, and  $t_i^a$  and  $t_j^b$  the time that each passenger takes to alight and board, respectively. Expression (4.1) takes place in buses with one door, while for (4.2) it is assumed the existence of one door to board and another door to alight<sup>23</sup>. The dead time accounts for the time necessary to open and close doors, plus any other time lost due to the nature of the process, for example, the time after the transfer of passengers has finished, in which the driver checks that everything is safe before closing the doors and the time lost in between boarding and alighting when the process is sequential. By denoting  $a$  and  $b$  as the *average* alighting and boarding times per passenger, respectively, (4.1) and (4.2) can be simplified to:

---

<sup>22</sup> Time between when the wheels stop and the moment they start moving off again.

<sup>23</sup> In general, dwell time in buses with two or more doors is given by the dwell time in the busiest door.

$$t_d = c_{oc} + a\lambda^- + b\lambda^+ \quad (4.3)$$

$$t_d = c_{oc} + \max\{a\lambda^-, b\lambda^+\} \quad (4.4)$$

In some cases there are noticeable differences in boarding and alighting times depending on passengers' age (young students, adults, seniors) and/or payment method and technology (prepay versus cash). To capture these differences, the expressions (4.3) and (4.4) can be further generalised to (York, 1993):

$$t_d = c_{oc} + \sum_{k=1}^{m_1} a_k \lambda_k^- + \sum_{l=1}^{m_2} b_l \lambda_l^+ \quad (4.5)$$

$$t_d = c_{oc} + \max\left\{\sum_{k=1}^{m_1} a_k \lambda_k^-, \sum_{l=1}^{m_2} b_l \lambda_l^+\right\} \quad (4.6)$$

where  $m_1$  and  $m_2$  denote the number of categories for passengers alighting and boarding (including different payment methods), respectively.

### 4.3.2 Data collection

Dwell time surveys were conducted in two areas of Sydney: the city centre and the Blacktown area in the western suburbs, approximately 25 kilometres from central Sydney. Surveys were conducted on weekdays between May and October 2009. Three types of service were surveyed:

- Commercial services in the Blacktown area: These are bus services in a low density residential suburb to the west of Sydney, run by a private operator (Busways). In these services, passengers only pay directly with cash to the driver. This service will be referred to as "Blacktown".
- Commercial services in the inner Sydney area: These are services run by the State Transit Authority (STA) of New South Wales. The area surveyed comprises the city centre plus the inner west and eastern suburbs of Sydney. On these services, passengers are able to pay the fare either with a prepaid magnetic strip or with cash to the driver. In every bus, two devices are set next to the front door (one at the right and one at the left, close to the driver) for passengers to validate their

card, which has to be introduced in a slot in order to be read (case ii in Section 4.2, Figure 4.1a). Boarding is only possible through the front door, whereas alighting takes place at both front and back doors. This service will be referred to as “inner Sydney”.

- Free CBD shuttle: This is a service operated by the STA around the city centre, introduced during 2009, which is free for passengers. Boarding is allowed only through the front door while alighting is possible at both front and rear doors. Free buses are used as a proxy for obtaining boarding time rates on public transport systems with off-board payment.

Buses are 12 metre long, have two doors and capacity for 60-70 passengers (seated and standing). A summary of the characteristics of each service is presented in Table 4.2.

**Table 4.2: Characteristics of the services surveyed**

<b>Characteristic</b>	<b>Blacktown</b>	<b>Inner Sydney</b>	<b>Free shuttle</b>
<b>Payment method</b>	Cash	Cash Magnetic strip	Free
<b>Boarding</b>	Front door	Front door	Front door
<b>Alighting</b>	Front door	Front door Back door	Front door Back door

Surveys were conducted on weekdays between May and October 2009 by the author, equipped with a stopwatch. For every bus stop observation, the following items are recorded:

- Time in which doors are open, plus door opening and closing times. Any extra time in which the bus is stopped but has the doors closed is not recorded.
- Number of passengers boarding, distinguishing:
  - Passengers that pay a fare, separated by payment method: magnetic strip and cash payment.
  - Student pass holders, who do not pay any fare on board, and only have to show their pass to the driver. These are school students (observed in Blacktown).

- Passengers that buy a daily or second ticket on board the bus, whose ticket is given and stamped by the driver, which makes the process slower (observed in Blacktown).
- Number of passengers alighting

Any unusual events are noted, such as particularly slow or disabled passengers boarding or alighting, passengers with prams, shopping trolleys, bags or other luggage, bus stopped with the doors open after the boarding and alighting processes have ended, etc. Observations with an extraordinary long dwell time due to exceptional events have been disregarded. Stops at termini are not considered as drivers sometimes keep buses stopped longer than the time necessary for the service of passengers.

In general, there are several factors that may influence the duration of dwell times, such as the location of bus stops and headway control actions. If a stop is located before a traffic light, drivers may keep the door open to allow boarding while the light is red. In addition, control strategies at stops such as a bus holding to regularise headways or maintain adherence to a predefined timetable, could also enlarge dwell times. In our study, none of these cases have an influence on the results, as the bus drivers always closed doors when the boarding and alighting of passengers concluded, and secondly, among the routes surveyed, extra delays on control points were rare and deleted from the sample. Boarding and alighting times also depend on the internal design of a bus, for example, the height of the bus floor and the door width. In this sample, we only consider data from low floor buses with doors that are approximately 1.1 metre wide<sup>24</sup>.

In this section we present the estimation of the average boarding and alighting time per passenger as a function of two variables: the technology of payment and the number of doors in which boarding and alighting is possible, the latter related to vehicle size. Dwell time models for the three services previously described are estimated with the program SPSS.

---

<sup>24</sup> Dwell time models including differences in boarding and alighting times due to the age of passengers and the existence of steps at the entrance of a bus (compared with the case of low floor buses) are reported in Tirachini (2011).

### 4.3.3 Model 1: Blacktown

The following categories of passengers are defined, according to what is observed in the field:

$\lambda_c^+$  : number of passengers boarding, cash payment

$\lambda_s^+$  : number of students boarding, free fare

$\lambda_t^+$  : number of passengers boarding who are given a daily ticket by the driver

$\lambda^-$  : number of passengers alighting

The transfer of passengers is sequential; boarding and alighting is made through the front door, therefore, using equation (4.5) the dwell time is estimated as:

$$t_d = c_{oc} + b_c \lambda_c^+ + b_s \lambda_s^+ + b_t \lambda_t^+ + a \lambda^- + \varepsilon \quad (4.7)$$

where  $\varepsilon$  is the residual or unexplained variance. The estimated parameters  $a$ ,  $b_c$ ,  $b_s$  and  $b_t$  and t-ratios for this model are presented in Table 4.3 (Model 1). The time in which a bus is stopped, not serving passengers, is 5.46 seconds on average, which is the sum of the time between when the bus doors are open and the first passenger boards (or alights), the time in between when the last passenger alights (or boards) and the doors are closed, and the time lost in between the boarding and alighting sequences. The second value, 9.94 s/pax is the average time for a passenger (other than school students) to board a bus, pay the fare and receive a ticket from the driver, considering the case in which passengers pay the exact fare or require change. Finally, the average time for a passenger to alight a bus is 1.56 seconds (the boarding time for students and passengers that require a second ticket are not relevant for the comparison of fare payment methods).

**Table 4.3: Estimation of parameters for dwell time models**

Parameter	Unit	Model 1 (Black-town)	Model 2 (Inner Sydney)	Model 3 (Free shuttle)
Dead time ( $c_{oc}$ )	s	5.46 (10.71)	6.41 (14.20)	6.47 (14.46)
Boarding time cash ( $b_c$ )	s/pax	9.94 (39.39)	11.54 (27.84)	-
Boarding time magnetic strip ( $b_m$ )	s/pax	-	2.94 (33.53)	-
Boarding time school students ( $b_s$ )	s/pax	1.50 (12.27)	-	-
Boarding time daily ticket ( $b_t$ )	s/pax	15.93 (14.95)	-	-
Boarding time free shuttle ( $b_1$ )	s/pax	-	-	1.46 (32.94)
Alighting time front door ( $a_1$ )	s/pax	1.56 (12.65)	2.53 (7.87)	1.64 (6.31)
Alighting time back door ( $a_2$ )	s/pax	-	1.06 (8.96)	1.18 (15.01)
Sample size		404	394	101
Adjusted R-squared		0.872	0.843	0.925

Note: t-ratios in brackets

#### 4.3.4 Model 2: Inner Sydney

Buses in this area have two devices at the front door for passengers to pay the fare with a prepaid card, which is inserted in a slot; one device is at the right of the door and the other one is at the left, close to the driver. In some sections of the network, outside the city centre, passengers are able to pay with cash to the driver as well. Alighting is allowed at both the front door and the rear door. The following categories of passengers are included in the analysis:

$\lambda_c^+$  : number of passengers boarding, cash payment

$\lambda_m^+$  : number of passengers boarding, prepaid magnetic strip

$\lambda_1^-$  : number of passengers alighting, front door

$\lambda_2^-$  : number of passengers alighting, back door

Then, dwell time is estimated by extending equation (4.6) as:

$$t_d = c_{oc} + \max \{b_c \lambda_c^+ + b_m \lambda_m^+ + a_1 \lambda_1^-, a_2 \lambda_2^-\} + \varepsilon \quad (4.8)$$

Results in Table 4.3 (Model 2) show that the average boarding time for passengers paying with a prepaid magnetic strip is 2.94 s/pax, whilst users that pay with cash take on average 11.54 seconds, value that is 16 percent larger than the figure found for Model 1. This is explained to a large degree by the fact that all passengers in Model 1 (Blacktown) pay with cash and are commuters and regular users of the service that live nearby; whereas users that pay with cash on the city centre services are those that do not have a prepaid card, i.e. mainly occasional users, visitors and tourists that are less familiar with the way local bus services work, and are more likely to ask questions to the driver.

The alighting times are considerably different depending on the door chosen to alight; for the front door passengers it takes 2.53 seconds on average, for the back door this figure is 1.06 seconds, which is a consequence of three issues observed in the field. First, alighting at the front door is allowed but discouraged because the front door is where boarding takes place. Second, passengers approach the front door from one direction only, whereas to alight through the back door they approach it from the front and back of the bus simultaneously (as the back door is in the middle of buses), therefore there can be two lines and two users may get off at the same time. The third reason is related to the composition of passengers that use the front and rear doors. Considering all passengers, it was observed that 74 (26) percent of users alight through the back (front) door, but there is a clear difference in terms of age groups: 79 percent of adults considered under 65 years old used the back door, whilst this figure is only 45 percent for passengers considered over 65, which indicates that older passengers tend to sit closer to the front door and, consequently, are more likely to use that door to alight. As shown by Tirachini (2011), older passengers are found to be slower at both boarding and alighting, which partly explains the difference in alighting times between the front and rear doors.

#### **4.3.5 Model 3: Free CBD shuttle**

Boarding time in a free bus service is an approximation to the time obtained in public transport systems with off-board fare collection, as done in most rail and some Bus Rapid Transit systems. In the case of the free CBD shuttle in Sydney, boarding is only through



the front door, but alighting is allowed at both the front and back doors. Using expression (4.6), the dwell time is estimated as:

$$t_d = c_{oc} + \max \{ a_1 \lambda_1^- + b_1 \lambda_1^+, a_2 \lambda_2^- \} + \varepsilon \quad (4.9)$$

where

$\lambda_1^+$  : number of passengers boarding, front door

$\lambda_1^-$  : number of passengers alighting, front door

$\lambda_2^-$  : number of passengers alighting, back door

The results in Table 4.3 (Model 3) suggest that the average boarding time for passengers without paying any fare is 1.46 s/pax; i.e., roughly half of the time required for payment with magnetic strip, and between one seventh and one eighth of the time to board and pay with cash to the driver. There are at least two aspects that may push this figure down in a commercial service with payment outside of buses, in closed stations like the BRT systems of Curitiba and Bogotá. First, a proportion of users of this shuttle service are tourists, unfamiliar with the transport system and the city, that are generally slower than regular commuters. Second, in this model as in Model 2, alighting time is shorter at the back door than at the front door, which suggests that if boarding is also allowed through the rear door(s), boarding time would be also shorter as passengers have more room to form two lines and, once inside the bus, two directions in the aisle to distribute themselves. In fact, Wright and Hook (2007, p. 262) suggest a value of 1.1 s/pax for boarding time with off-board payment on 1.1 metre wide doors with stairways at the door entrance, the same figure being as short as 0.75 s/pax with at-level boarding in the Transmilenio system in Bogotá. This latter value is likely influenced by cultural and behavioural characteristics of passengers that are specific to the local context of Bogotá and South America, and are unlikely to be met in Australia (e.g., higher tolerance to walk or stand closer to other passengers).

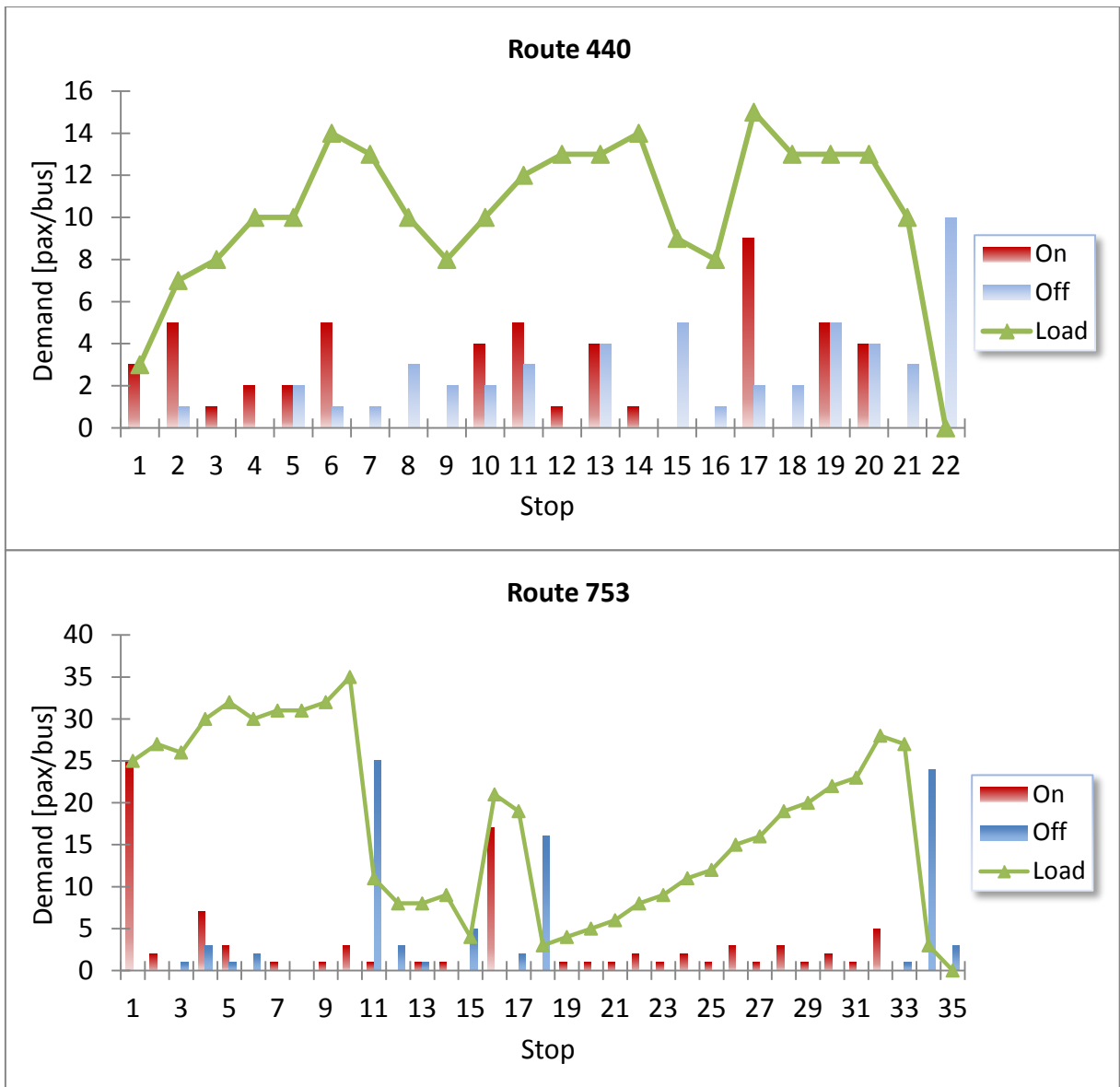
### 4.3.6 Average passenger service times

Next, it is necessary to derive a single value for each of the different categories of boarding and alighting times. For cash payment, we use the average between the two values in Table 4.3, i.e., 10.74 s/pax. In the case of alighting, for simplicity we calculate a single value as the average between the alighting times from Models 1 and 3 (1.56, 1.64 and 1.18 s/pax), i.e., 1.46 s/pax. (Model 2 is disregarded for alighting because of the particular considerations that influence the considerable difference between  $a_1$  and  $a_2$ , as discussed in Section 4.3.4). A dead time of 6.11 seconds is used, as the average of the three values on Table 4.3. Finally, a boarding time of 2.05 s/pax is used for the case with a contactless card, obtained by Fernández *et al.* (2009) for trunk services in Santiago, Chile. In summary, the values for boarding and alighting with different fare payment technologies are presented in Table 4.4.

**Table 4.4: Boarding and alighting times per passenger**

Parameter	Time [s/pax]
Boarding time cash	10.74
Boarding time magnetic strip	2.94
Boarding time contactless card	2.05
Boarding time off-board payment	1.46
Alighting time	1.46

The boarding and alighting values shown in Table 4.4 can be used to estimate average passenger service times *APST* (boarding plus alighting) for the alternative fare payment technologies and boarding and alighting policies introduced in Section 4.2 (Table 4.1). The average passenger service time is the delay that one extra passenger imposes on a bus ride. Observed demand profiles of two bus routes in Sydney are used to estimate *APST* for all cases in Table 4.1. The boarding profile, alighting profile and bus load for each route is presented in Figure 4.2, in which the total demand along a trip is 51 passengers per bus for route 440, whilst bus 753 carries 87 passengers.



**Figure 4.2: Load profile routes 440 and 753**

These demand profiles are used as a seed to generate other profiles, by uniformly reducing and amplifying observed boarding and alighting numbers, in order to make the total demand along the routes move from 20 pax/bus to 100 pax/bus. *APST* is calculated as the total boarding and alighting time divided by the total number of passengers per bus (which is different from the time that each passenger takes to board and alight, as more than one passenger may be boarding and alighting at the same time). It should be noted that as the number of passengers inside a bus increases, *APST* may increase as well due to crowding and friction effects between passengers boarding, alighting and standing in the bus aisle or next to the doors, as theoretically proposed by Jara-Díaz and Gschwender (2003a) and empirically shown by Lin and Wilson (1992). At this point we

assume that the *APST* is independent of the demand level; the impact of crowding on travel times is analysed in depth in Chapter 6.

Using the values of Table 4.4 and demand profiles of Figure 4.2, *APST* are estimated for buses with one, two, three and doors. Note that the parameters in Table 4.4 were obtained from services with boarding allowed at the front door only, and the extension of these results to a case with boarding at multiple doors is not trivial, since when there are multiple doors to board and alight, passengers can choose a door to get on and off buses, and the spatial dispersion of their decision will determine the length of the boarding and alighting times per door. The best scenario is a more and less even distribution of passengers across all doors, whereas the worst case is when they all concentrate in one or two doors, slowing down the entire boarding and alighting process.

Unfortunately, there does not seem to exist studies on the behaviour of passengers that can choose a door to board buses, therefore an assumption has to be made for the derivation of average passengers service times for cases boarding at all available bus doors (denoted *TnBn*). It seems unreasonable to suppose that passengers will distribute uniformly across doors if middle or back doors have closer access to more seats than, say, the front door. For example, in buses with two doors in Sydney it was found that 75 percent of passengers alight through the back door when alighting is also possible—but not encouraged— through the front door (Tirachini, 2011). In this analysis we assume that the middle doors would attract a number of passengers that is 50 percent higher than that of the front or back doors<sup>25</sup>. The same assumption is made regarding alighting. Figure 4.3 and Table 4.5 show the estimated average boarding and alighting time per passenger, as a function of the fare payment method and the number of doors per bus, “*TnB1*” means “Total *n* doors, boarding at front door only”, whilst “*TnBn*” stands for “Total *n* doors, boarding at all doors”, where *n* is the number of doors at a bus.

---

<sup>25</sup> For example, for buses with two doors, the rear door is placed towards the centre of the bus, and is therefore assumed to attract 60 percent of the boarding demand, leaving 40 percent boarding through the front door, next to the driver.

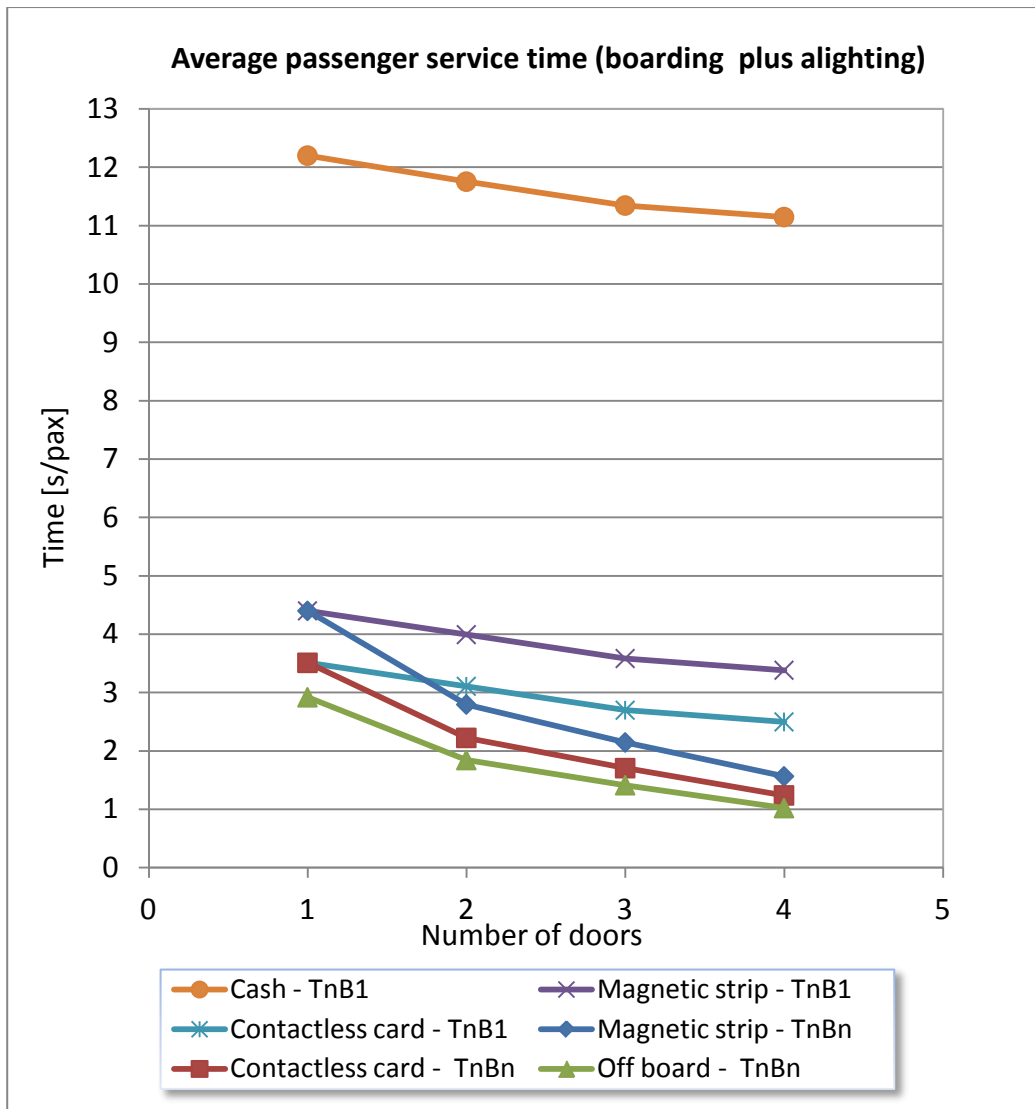


Figure 4.3: Average passenger service time

Table 4.5: Average passenger service time (s/pax)

Doors	Cash <i>TnB1</i>	Magnetic strip <i>TnB1</i>	Contactless card <i>TnB1</i>	Magnetic strip <i>TnBn</i>	Contactless card <i>TnBn</i>	Off- board <i>TnBn</i>
1	12.20	4.40	3.51	4.40	3.51	2.92
2	11.75	3.99	3.11	2.80	2.22	1.84
3	11.34	3.58	2.70	2.15	1.71	1.41
4	11.15	3.38	2.50	1.57	1.24	1.02

Figure 4.3 shows the impact of technology and the number of doors on the average time to board and alight per passenger, *APST*. On the one hand, there is a *technology effect* interpreted as for any size of bus (number of doors), it is more efficient to provide an off-board payment system than on-board payment with contactless card, which in turn is faster than paying with a magnetic strip. On the other hand, the *door effect* shows how

*APST* decreases with the number of doors, regardless of the fare payment technology. In the next two sections we analyse in detail the impact of the alternative fare payment technologies and boarding and alighting policies on the performance of a bus route, first by estimating functions for the curves in Figure 4.3 (Section 4.4), and secondly by estimating a bus travel time model to assess differences on the total running speed achievable with the fare payment and boarding options under study (Section 4.5).

#### 4.4 Technology Effect and Door Effect: Non-linearity and Interdependency<sup>26</sup>

Focusing on the scenarios in which boarding is allowed at all doors (*TnBn* in Figure 4.3), given a payment method, the evolution of the curves reveals a decreasing importance of the number of doors on lowering *APST*, i.e., the time savings due to increasing the number of doors by one unit decrease with the number of doors, something that has been observed in Bus Rapid Transit systems with off-board payment in Brazil (Wright and Hook, 2007). Figure 4.3 reveals that payment technology as a tool to reduce travel times becomes less powerful as the number of doors increases, which is shown by the reduction in the vertical difference between the *TnBn* lines as the number of doors increase. This means that the time savings due to upgrading the payment method and due to increasing the number of doors to board are not independent, i.e., there is a non-linear relationship between the boarding and alighting time, the number of doors to board and the payment technology. Ignoring cash payment, a simple model to explain the relationship observed in Figure 4.3 is presented in equation (4.10).

$$t_{TnBn} = 2.96n^{-0.72} + 1.71\delta_{mag} + 0.69\delta_{con} - 0.31\delta_{mag}n - 0.13\delta_{con}n \quad (R^2 = 0.997) \quad (4.10)$$

where  $n$  is the number of doors per bus and  $\delta_{mag}$  and  $\delta_{con}$  are dummy variables to distinguish on-board from off-board payment as follows:

$$\delta_{mag} = \begin{cases} 1 & \text{if payment with magnetic strip} \\ 0 & \text{otherwise} \end{cases}$$

---

<sup>26</sup> This section is partially reproduced in Section 2 of Jara-Díaz and Tirachini (2012).

$$\delta_{con} = \begin{cases} 1 & \text{if payment with contactless card} \\ 0 & \text{otherwise} \end{cases}$$

With off-board payment,  $\delta_{mag} = \delta_{con} = 0$ . The last two terms of equation (4.10) show that the time savings of an off-board payment system are reduced by 0.31 and 0.13 seconds per passenger as the number of doors grows by one unit. This non-linear effect is not observable when boarding is permitted only at the front door (*TnB1* lines in Figure 4.3), in which case the boarding and alighting time per passenger can be simply approximated as (magnetic strip and contactless card only):

$$t_{TnB1} = 3.55n^{-0.24} + 0.88\delta_{mag} \quad (R^2 = 0.989) \quad (4.11)$$

The scenarios with boarding through all doors are always more time efficient than their counterparts with boarding at the front door only, for a given number of doors, a non-surprising result as boarding is more time onerous than alighting. However, permitting boarding through all doors has the extra cost of installing card readers on every door (for on-board payment) or at stations (for off-board payment), relative to scenarios with boarding at the front door only. All these considerations need to be taken into account to assess the convenience of one system over the other, as it is done in Chapter 7.

#### 4.5 Technology Effect and Door Effect: Assessment with an Empirical Bus Travel Time Model<sup>27</sup>

##### 4.5.1 Background: bus travel time models

Figure 4.3 is a graphic representation of the time savings attached to upgrading the fare payment technology (*technology effect*) and/or allowing boarding at all available bus doors (*door effect*). However, Figure 4.3 says nothing about the relative weight of that saving with regard to the total travel time of buses, which would provide an insight on the size of the benefits achievable by upgrading the boarding policy and/or fare payment technology. To address this issue, in this section we assess the effect of alternative

---

<sup>27</sup> This section is partially reproduced in Tirachini (2012).

payment and boarding options on the operating speed and total travel time of a bus route, by means of estimating a bus travel time model which uses as an input parameter the average passenger service time.

The traditional approach to analyse bus travel times is the use of linear regression models. Early research by Abkowitz and Engelstein (1983) showed that the main drivers of bus travel time are route length, the period of time in which a trip is made (peak or off-peak), the number of passengers boarding and alighting and the number of signalised intersections. Outbound trips (from the *Central Business District, CBD*) were found to be slower than their inbound counterparts, whereas on-street parking also increases bus travel time. At the same time, Levinson (1983) studied a set of field data observations from several U.S. cities to characterise differences in bus travel time as a function of the number of stops per mile, location of the route (*CBD*, city or suburbs), acceleration and deceleration times and dwell time per stop. Interestingly, In Levinson's study it was suggested that generally the most effective way to reduce bus travel time is decreasing the number of stops and dwell times through changes in fare collection policies and door configuration, rather than providing bus priority lanes or reducing traffic-related congestion. In Section 4.5.5 we provide a simple comparative assessment of bus speed gains with two policies – providing dedicated busways and upgrading the fare collection system – and show that the latter can yield greater time savings if bus demand is high enough.

Following Abkowitz and Engelstein (1983), multivariate regression models to analyse bus travel time have been subsequently estimated and utilised with several purposes, such as analysing the appropriateness of scheduled service recovery times (Strathman *et al.*, 2002), the estimation of time savings by means of limited-stop services (Tétreault and El-Geneidy, 2010) and the examination of variables describing service reliability and schedule adherence (Strathman *et al.*, 1999; 2000; El-Geneidy *et al.*, 2008). Apart from the key factors identified in the 1980s, other variables found to influence travel times are departure delays (Strathman *et al.*, 2000; El-Geneidy *et al.*, 2008; Tétreault and El-Geneidy, 2010), scheduled headway (Strathman *et al.*, 2000), driver-related effects (Strathman *et al.*, 2002; El-Geneidy *et al.*, 2008), type of route service (whether the route



is cross-town or feeder, Strathman *et al.*, 2002) and weather conditions (presence of rain and snow, Tétreault and El-Geneidy, 2010). Car travel time (McKnight *et al.*, 2003) and car traffic counts (Mazloumi *et al.*, 2011) have also been used as an explanatory variable for bus travel time, as a way to quantify the effects of traffic congestion on bus operations<sup>28</sup>.

#### **4.5.2 Data collection and travel time model estimation**

On-board travel time surveys are used for the estimation of travel time models, collected by a bus operator (Busways) on weekdays from November 2007 to March 2009 in the Blacktown area, which is characterised by having a low residential density with a relatively low demand for public transport (2.1 pax/bus-km on average) and bus operating speed of 25.9 km/h on average. Figure 4.4 shows the bus network in the study area, which has been divided in four zones for the estimation of zone-specific factors; the zones are South-East, South-West, North-East and North-West. A fifth zone, referred to as Transversal-North is also defined, comprising routes that run across both the North-East and North-West zones.

The travel time surveys are manually collected by a single observer onboard buses, on either one-way or round trips. The following information is recorded:

- Bus route number.
- Time of the day.
- Scheduled arrival and departure times at the beginning of the route, end of the route and at every bus stop along the route.
- Actual arrival and departure times at the beginning of the route, end of the route and at every bus stop along the route.
- Number of passengers boarding and alighting per bus stop. Only actual stops are recorded, i.e., designated bus stops with no boarding and alighting demand (where the bus does not stop) are not noted.

---

<sup>28</sup> Beyond multiple regression models, more sophisticated techniques are usually proposed in the literature of real time estimation of bus travel time, such as artificial neural networks (Chien *et al.*, 2002; Chen *et al.*, 2007) and Kalman filter algorithms (Shalaby and Farhan, 2004; Padmanaban *et al.*, 2009) developed to predict arrival times at bus stops. Jeong and Rilett (2005) found that artificial neural network models outperform regression models and historical data-based models when predicting bus arrival in real time, using automatic vehicle location data on a route in Houston, Texas.

In total there are 316 surveys corresponding to 23 different bus routes spread out across the entire network<sup>29</sup>.

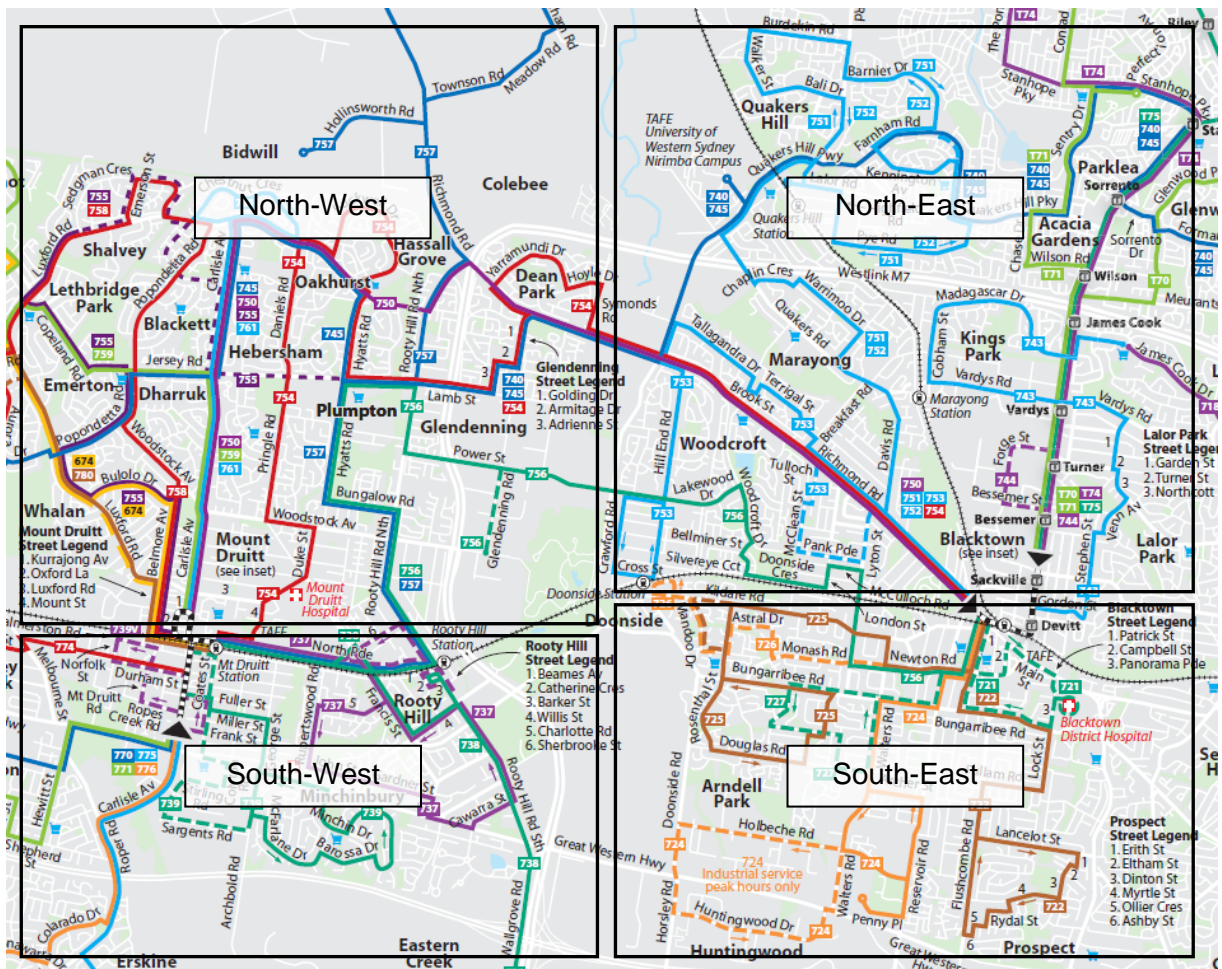


Figure 4.4: Bus network in the study area  
(Source: [www.busways.com.au](http://www.busways.com.au))

In order to assess the impact of the road configuration on bus travel times, the total number of traffic light intersections, give-way intersections, roundabouts and speed humps per route are added as potential explanatory variables, to complement the information from the travel time surveys. Finally, the route length for each observation (which may vary along the day for a specific bus line) is obtained from the schedule program of the bus operator.

<sup>29</sup> Routes 718, 721, 722, 724, 725, 726, 737, 738, 739, 743, 745, 751, 752, 753, 753W, 754, 755, 756, 756G, 757, 761, T70 and T71.

Dummy variables are assigned to each zone to analyse zone-specific differences in travel times. The best estimated model, including only statistically significant variables, is given by expression (4.12). Estimations of the models are made using the program SPSS.

$$T_c = c + (t_L + t_{47}\delta_{47} + t_{89}\delta_{89} + t_{912}\delta_{912} + t_{se}\delta_{se} + t_{dep}n_{dep})L + t_i I + t_r R + t_s S + (t_N + t_{N912}\delta_{912})N_f + \varepsilon \quad (4.12)$$

where the dependent variable is the total travel time  $T_c$ , and the independent variables and parameters are defined as follows:

$c$ : Constant [s]

$L$ : Length of the route [km]

$t_L$ : One-kilometre non-stop travel time [s/km]

$t_{47}$ : Extra one-kilometre non-stop travel time is trip is between 4 and 7 AM [s/km]

$$\delta_{47} = \begin{cases} 1 & \text{if trip is between 4 and 7 AM} \\ 0 & \text{otherwise} \end{cases}$$

$t_{89}$ : Extra one-kilometre non-stop travel time is trip is between 8 and 9 AM [s/km]

$$\delta_{89} = \begin{cases} 1 & \text{if trip is between 8 and 9 AM} \\ 0 & \text{otherwise} \end{cases}$$

$t_{912}$ : Extra one-kilometre non-stop travel time is trip is between 9 AM and 12 PM [s/km]

$$\delta_{912} = \begin{cases} 1 & \text{if trip is between 9 and 12 AM} \\ 0 & \text{otherwise} \end{cases}$$

$t_{se}$ : Extra one-kilometre non-stop travel time is trip is in South-East zone [s/km]

$$\delta_{se} = \begin{cases} 1 & \text{if trip is in South-East zone} \\ 0 & \text{otherwise} \end{cases}$$

$t_{dep}$ : Extra non-stop travel time per minute of late departure, per kilometre [s/min-km]

$n_{dep}$ : Late departure time [min]

$t_i$ : Average delay per traffic light intersection [s/intersection]

$I$ : Number of intersections

$t_r$ : Average delay per roundabout [s/roundabout]

$R$ : Number of roundabouts

$t_s$  : Average stopping delay per bus stop [s/stop]

$S$  : Number of bus stops

$t_N$  : Average boarding and alighting time per passenger (passenger service time) [s/pax]

$t_{N912}$  : Extra boarding and alighting time per passenger if trip is between 9 AM and 12 PM  
[s/km]

$N_f$  : Total demand per bus [pax]

$\varepsilon$  : Residual or unexplained variation

The results of the estimation of parameters for model (4.12) are shown in Table 4.6.

**Table 4.6: Estimation of travel time model**

Parameter	Unit	Value
Constant $c$ <sup>30</sup>	s	-55.4 (-1.55)
One-kilometre travel time $t_L$	s/km	92.6 (22.46)
Extra time 4-7 AM $t_{47}$	s/km	-12.8 (-4.60)
Extra time 8-9 AM $t_{89}$	s/km	11.9 (5.89)
Extra time 9 AM-12 PM $t_{912}$	s/km	-10.2 (-3.19)
Extra time South-East zone $t_{se}$	s/km	15.8 (6.43)
Extra time late departure $t_{dep}$	s/min-km	-0.9 (-3.09)
Delay per traffic light $t_i$	s/intersection	23.3 (6.80)
Delay per roundabout $t_r$	s/roundabout	7.7 (1.78)
Delay per bus stop $t_s$	s/stop	16.8 (5.91)
Average passenger service time $t_D$	s/pax	6.7 (7.07)
Extra delay passenger 9 AM-12 PM $t_{D912}$	s/pax	9.9 (4.88)
Adjusted R-squared		0.955

Note: t-ratios in brackets

All variables are statistically significant at the 95 percent level of confidence, except for  $t_r$ , the delay due to roundabouts, which is significant at the 90 percent level. Table 4.6 shows that buses take on average 92.6 seconds to travel along one kilometre, without interruptions of any sort (equivalent to a non-stop speed of 38.9 km/h). This base travel time has some variations depending on the period and zone where a route is. For example, if a trip is between 4 AM and 7 AM, the uninterrupted travel time is 12.8 s/km shorter (equivalent to 45.1 km/h), whereas if the trip is on the morning peak, between 8

---

<sup>30</sup> Constant  $c$  is negative which means that, theoretically, the estimated model could predict a negative travel time. However, that is not physically possible with the data used for the estimation (or on any realistic bus route), because given the estimated parameter values in Table 4.6, for  $T_c$  to be negative the route would have to be shorter than one kilometre and, for example, have no passengers boarding buses.

AM and 9 AM, travel time is 11.9 s/km longer on average. In terms of spatial distinctions, the only zone with a statistically significant difference in travel time with respect to the others is the South-East zone, whose non-stop travel time is 15.8 s/km longer on average than in the others (equivalent to 34.4 km/h). This is a measurement of the increased traffic congestion in this zone with respect to the rest of the network, which makes buses to be slower. The parameter  $t_{dep} = -0.9$  s/min-km suggests that drivers drive slightly faster when they depart late at the beginning of the route. The average delay related to traffic lights is 23.3 seconds per intersection, which accounts for accelerating, decelerating and waiting time at intersections. Roundabouts, on the other hand, yield an average delay of 7.7 seconds. The average stopping delay per bus stop is 16.8 s/bus (including the time necessary to open and close doors), while the extra delay per passenger— the average *passenger service time APST*— is 6.7 seconds, which include both passengers paying the fare with cash to the driver and free concessions (pooled together as only total number of passengers getting on and off is available in this sample), therefore it is not representative of any fare payment alternative. Then, APST from Table 4.5 will be used to analyse differences on operating speed with alternative fare payment and boarding policies.

#### **4.5.3 Percentage of time spent at each stage**

The parameters of the travel time model in Table 4.6 are useful to benchmark the performance of bus routes. In this section we show how a travel time model, once estimated, can be used back to calculate the amount of time spent at each stage of a trip. Five stages are considered:

- Non-stop running time
- Delay due to traffic lights
- Delay due to roundabouts
- Time lost at bus stops— acceleration, deceleration, opening and closing of doors (*AD*)
- Time lost at bus stops— passenger service time (*PST*).

In previous research, the percentage of time spent at different stages of a bus trip has been directly measured in the field (Levinson, 1983; Maloney and Boyle, 1999; Bertini and El-Geneidy, 2004). In this study we take a different approach by using a travel time model to indirectly estimate the time spent on-route, at intersections and at bus stops. These estimates are used to predict the impact of changes due to proposed modifications on a particular route or the entire network, such as the addition or removal of bus stops, traffic lights, roundabouts, or upgrading the fare collection system, something that is hard to do when running times or delays are directly measured with no account of the underlying factors behind them.

**Table 4.7: Percentage of time spent at each stage**

<b>Time period</b>	<b>Non-stop running</b>	<b>Traffic lights</b>	<b>Roundabouts</b>	<b>Bus stops (AD)</b>	<b>Bus stops (PST)</b>
04:00-07:00	66.0%	13.1%	3.2%	11.7%	5.9%
07:00-08:00	60.7%	10.9%	2.8%	14.1%	11.4%
08:00-09:00	62.4%	9.5%	2.2%	12.9%	12.9%
09:00-12:00	60.9%	12.5%	2.6%	10.2%	13.8%
12:00-14:00	67.9%	10.4%	2.9%	12.0%	6.8%
14:00-16:00	62.0%	10.0%	2.4%	14.2%	11.4%
16:00-18:00	63.5%	12.3%	2.7%	12.4%	9.0%
18:00-20:00	71.4%	10.9%	3.0%	10.1%	4.6%
Day average	63.3%	11.3%	2.7%	12.3%	10.3%

We estimate the amount of time spent in each of the five categories previously defined, for all recorded bus trips which are then averaged per time period in Table 4.7. It is shown that if buses did not have to stop or decelerate because of traffic lights, roundabouts or bus stops, the travel time would be, on average, 63.3 percent of the total current time, whereas average delays in traffic lights and roundabouts account for 11.3 and 2.7 percent of the total travel time, respectively. The time spent at bus stops is 22.6 percent of the total, decomposed in 12.3 percent spent on accelerating, decelerating and opening and closing doors, while 10.3 percent of the time is lost when boarding and alighting passengers. There are some differences between time periods, as for example, in the periods 04:00-07:00, 12:00-14:00 and 18:00-20:00, buses actually spend shorter times serving passengers (*PST*) relative to the total travel time, which reveals that these periods are off-peak in terms of demand.

In summary, Table 4.7 illustrates how this type of travel time model is useful to disaggregate bus travel times in stages, and therefore, to provide a first idea of the main sources of delays along a route. In turn, this can be used as background information to propose control strategies or corrections policies if necessary, such as adjustments in the number of bus stops and improvements in the efficiency of the boarding and alighting process (see Section 4.5.5 for a comparison of having busways versus upgrading the fare collection system). Changing the location and number of bus stops is a strategy that has received considerable attention in the literature (e.g., Levinson, 1983; Kuah and Perl, 1988; Furth and Rahbee, 2000; Saka, 2001; Ibeas *et al.*, 2010; Tétreault and El-Geneidy, 2010), whereas the implications of upgrading the fare payment technology are analysed in Sections 4.5.4, 4.5.5 and 4.5.6.

#### **4.5.4 Bus operating speed**

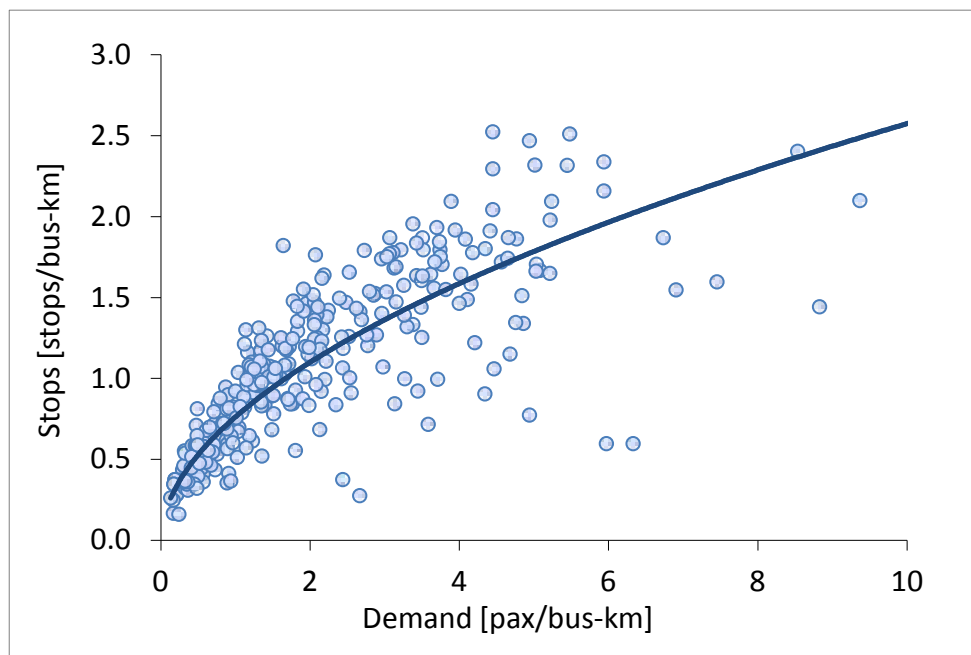
Using the values of Table 4.5, it is possible to estimate the potential time and cost savings due to upgrading the fare collection system and/or boarding policy. However, before attempting this exercise a couple of considerations need to be made. First, the full realisation of potential time savings depends on how timetables can be adjusted after changing the fare collection system; for instance, the need to synchronise transfers between two or more lines (Ceder *et al.*, 2001; Fleurent *et al.*, 2004) or to publish timetables at bus stops with departure times rounded to entire minutes would set constraints for the translation of potential savings into actual savings. In what follows, potential benefits from having a quicker fare payment method are estimated, with no concern about the adaptation of timetables; in this respect the approach is directly applicable to bus services that are not based on timetables for passengers (e.g., a high frequency route with one bus every five minutes or less).

Second, the composition of the patronage is relevant because if there are passengers exempted from paying a fare (e.g., school students, senior pensioners), the effectiveness of upgrading the fare payment technology as a tool to decrease travel times is reduced. On the other hand there might be differences within the group that pays a fare; for example senior passengers might be slower to board and alight buses than younger



passengers, although the time differences with alternative fare payment methods exist regardless of the age of the passengers (Tirachini, 2011)<sup>31</sup>.

In order to apply the estimated bus travel time model (Equation 4.12), an estimation of the actual number of stops that a bus makes to serve passengers is required. In formal urban public transport systems that face high demand, such as BRT services, it is common that buses have to stop at every designated station along the line, regardless of whether or not there are passengers that actually want to board or alight. Nevertheless, in the case of low-demand bus services, it is usual that buses stop only when they are required to by passengers on board that need to get off at the next stop, or by passengers who signal to the driver while waiting on a bus stop. Intuitively, the number of times that a bus actually stops along a route depends on ridership, as with a low total demand per bus, it is less likely that buses are required to stop at every scheduled bus stop. The relation between the actual number of stops per bus-kilometre and demand per bus-kilometre found in the travel time surveys is shown in Figure 4.5. It is evident that a high proportion of the variation in the actual number of stops is explained by the average bus demand.



**Figure 4.5: Actual number of bus stops as a function of passengers per bus**

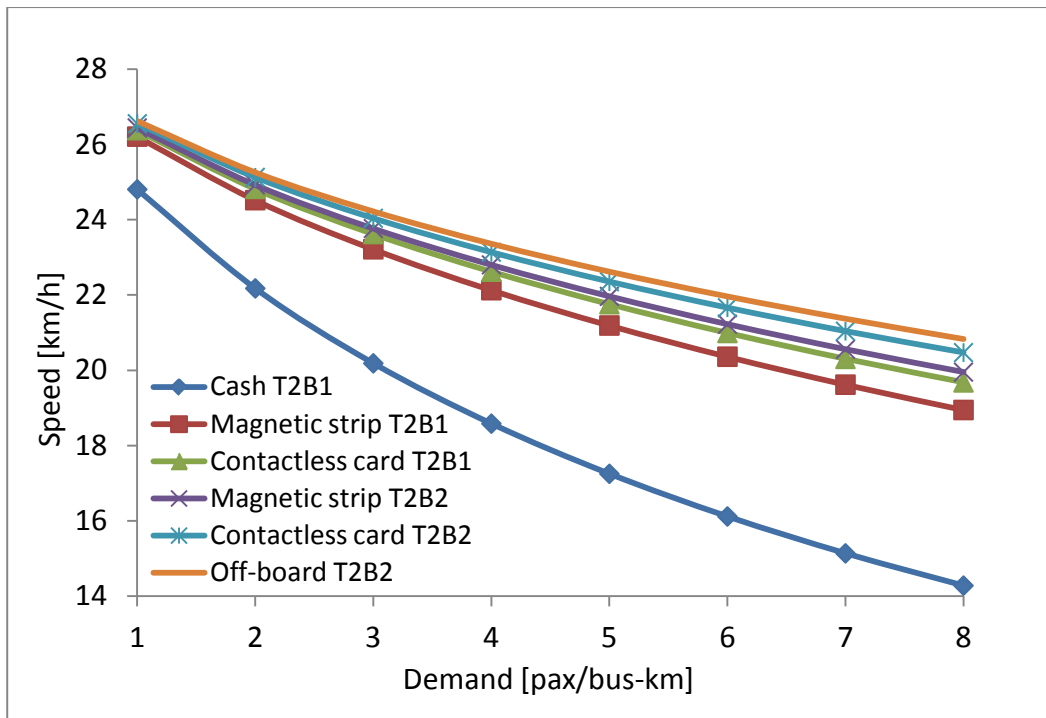
---

<sup>31</sup> The figures in Table 4.5 are average boarding and alighting times for all age groups.

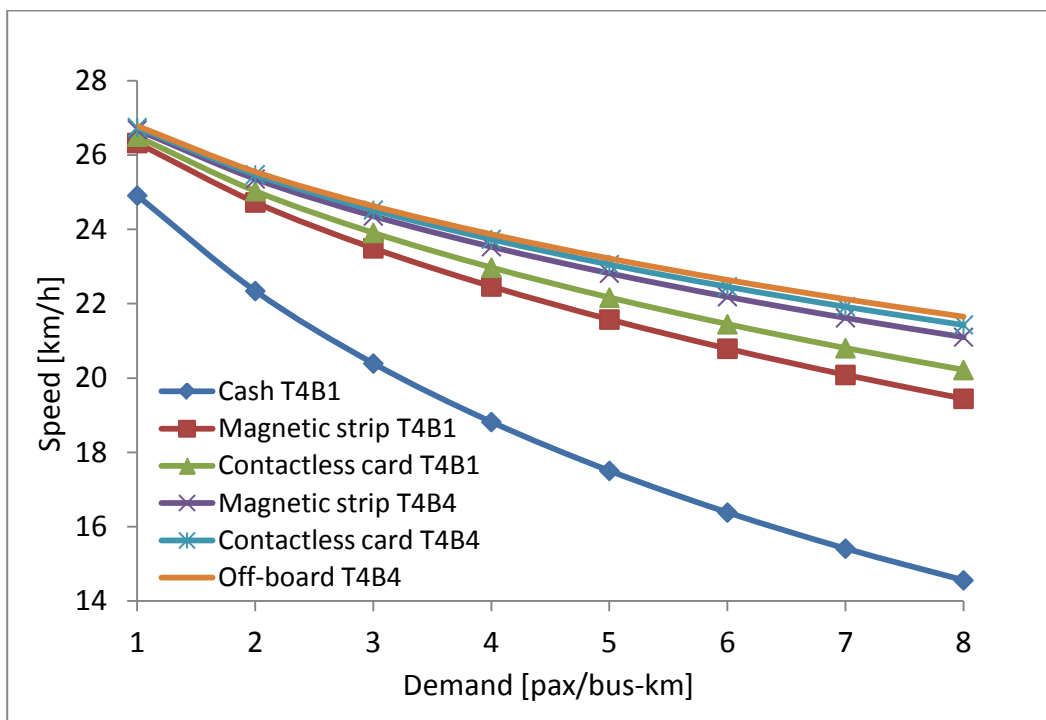
Numerically, we find that the curve that better fits the scatter plot depicted in Figure 4.5 is the power function shown in equation (4.13), where  $S_{km}$  is the actual number of stops per kilometre and  $D_{km}$  is the average number of passengers per bus-kilometre. The power 0.528 implies that the actual number of stops per kilometre roughly varies with the square root of demand.

$$S_{km} = 0.763D_{km}^{0.528} \quad (R^2=0.736) \quad (4.13)$$

The operating (or commercial) speed is the average speed along a route including both running time and stops of any sort. We simulate the circulation of buses with two and four doors, assuming that travel time model (4.12) is valid for the two types of buses despite their difference in size, on a route of 16 kilometres of length, with 11 traffic lights and 8 roundabouts (these figures are the average values of the variables in the sample) during the morning peak (8-9 AM). Average demand varies between 1 and 8 pax/bus-km, which in turn determines the actual number of bus stops per kilometre, as given by Equation (4.12). The APST per payment system are obtained from Table 4.5. Figure 4.6 shows how operating speed decreases with demand, and the loss of speed is stronger the more inefficient the fare payment system is. For example, on 2-door buses speed drops from 24.8 to 14.3 km/h for cash payment, while in the same demand range the drop is 26.4 to 19.7 km/h for contactless card with boarding at the front door only. This is a quantification of an expected result, that the benefits of having an efficient fare payment system, with prepaid cards or off-board payment, are greater the larger demand is.



(a) 2-door buses



(b) 4-door buses

Figure 4.6: Bus operating speed as a function of demand and fare payment and boarding policy

When boarding is allowed at the front door only (cases *T2B1* and *T4B1*), in both plots there is a noticeable gain in speed when upgrading the fare collection method from cash to magnetic strip, and from magnetic strip to contactless card; nevertheless, the

technology effect on increasing speed (reducing travel time) is weaker when boarding is allowed at all doors (cases *T2B2* and *T4B4*). This is particularly evident in buses with four doors, as the vertical difference between the three *T4B4* curves is almost marginal in Figure 4.6b. In other words, upgrading the fare payment technology has a major impact on performance when boarding is allowed at the front door only, but this technology effect diminishes when boarding is allowed at all doors, especially on bigger buses. Note that the *TnBn* boarding policy is superior to *TnB1* in all cases, even if the fare is paid with magnetic strip (with contact) in the former case and with the faster contactless card in the latter case. This indicates that if a bus service is provided with on-board magnetic strip payment and boarding is allowed at the front door only, in order to save travel time it is more effective to allow boarding at the back doors (installing card reading devices) than to upgrade the technology of payment to contactless card keeping the one-door boarding policy. In other words, the door effect can be more powerful than the technology effect.

#### **4.5.5 Going cashless or creating busways?**

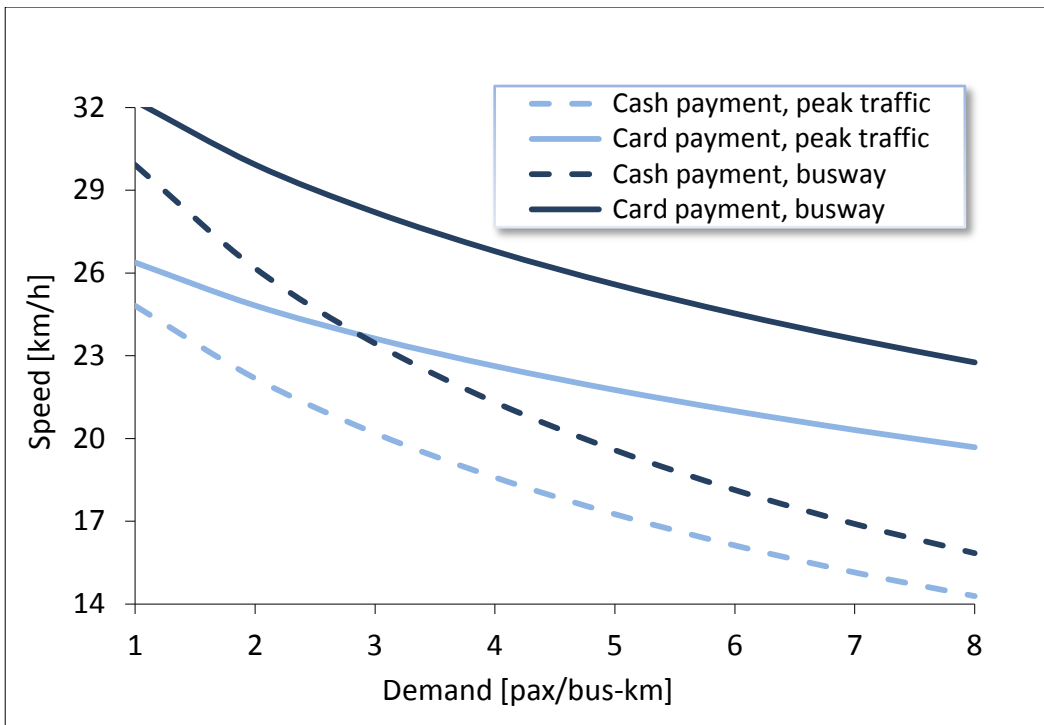
The bus travel time model and the passenger service times with alternative fare collection systems are useful as a starting point to compare bus speed gains, achieved by upgrading the fare collection system and/or providing dedicated bus lanes or busways. A busway aims at separating buses from cars in order to reduce bus travel times and improve service reliability; in this sense it is a measure to reduce travel time in links (the non-stop time in Table 4.7), as opposed to improving the boarding and alighting process which decreases the time spent at bus stops. Table 4.7 reveals that the non-stop running time for buses along the network amounts to 63.3 percent (daily average) of the total running time, whilst the passenger service time is only 10.3 percent of the total time. In this section we analyse if this evidence is sufficient to suggest the provision of on-road priority schemes for buses -instead of more efficient fare collection techniques- with the objective of speeding up buses.

The result depends on the relative impact of traffic congestion and bus demand levels on bus travel times. Analytically, if the introduction of a busway reduces non-stop bus travel time from  $t_{L1}$  to  $t_{L2}$  s/km and, on the other hand, upgrading the fare payment system

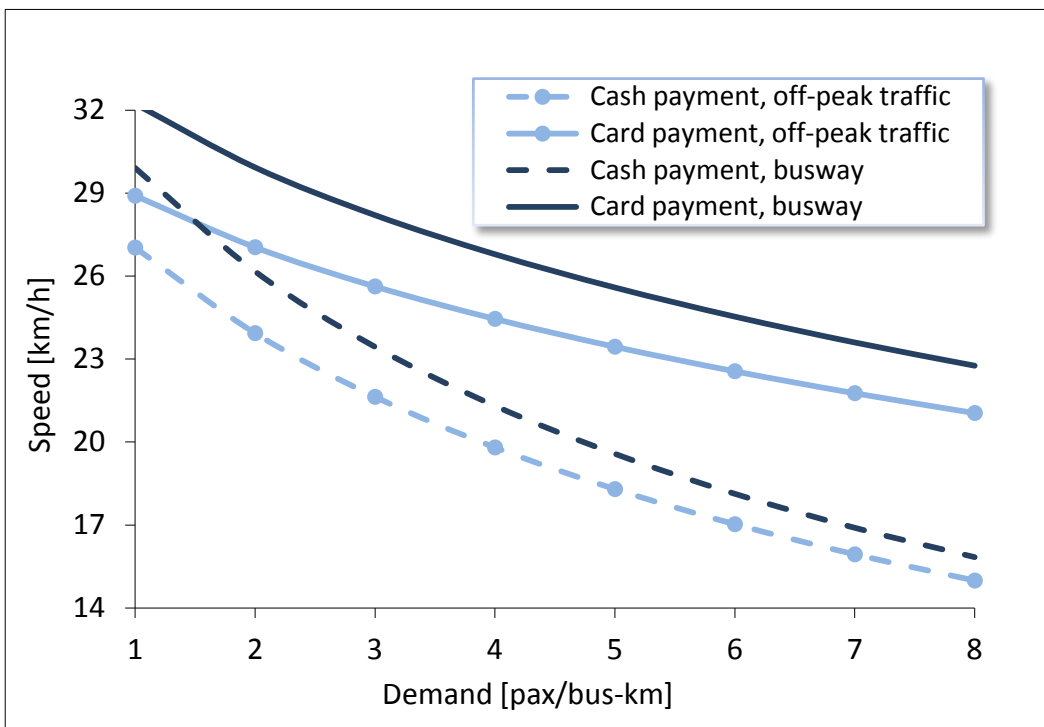
decreases passenger service time from  $t_{N1}$  to  $t_{N2}$  s/pax, the relevant comparison is between time savings  $(t_{L1} - t_{L2})L$  and  $(t_{N1} - t_{N2})N_f$ , where  $L$  is the route length and  $N_f$  is the total demand per bus ride. Let us define the average demand per bus kilometre as  $N_{km} = N_f / L$ , then for a bus demand greater (lower) than  $N_{km}^* = (t_{L1} - t_{L2}) / (t_{N1} - t_{N2})$  pax/bus-km, buses will be faster (slower) with a policy that reduces boarding and alighting times, relative to a policy that increases the non-stop bus speed. For a numerical comparison, three scenarios are defined using parameters of the travel time model given in Table 4.6. These are:

- *Segregated bus operation*: we assume that the period between 4 and 7 AM presents free-flow conditions and can be used as a proxy to an operation of buses on dedicated busways. From Table 4.6, the non-stop travel time in this case is  $92.6 - 12.8 = 79.8$  s/km (equivalent to a non-stop speed of 45.1 km/h).
- *Mixed-traffic operation, peak congestion*: in this case we use the morning peak to represent congested operation with buses and cars sharing the right-of-way, therefore the non-stop travel time is  $92.6 + 11.9 = 104.5$  s/km (34.4 km/h).
- *Mixed-traffic operation, off-peak congestion*: we use the base one-kilometre travel time to represent off-peak congestion, which is applicable to the periods 7-8 AM and after 12PM. In this case the non-stop travel time is 92.6 s/km (38.9 km/h).

Two fare payment methods are chosen in each scenario: on-board cash payment (cash T2B1) and on-board contactless card payment (contactless card T2B1). The estimated operating speed for each scenario is presented in Figure 4.7 (2-door buses), where the cases with peak and off-peak congestion are compared against the segregated bus operation separately.



(a) Peak traffic



(b) Off-peak traffic

Figure 4.7: Bus operating speed, mixed-traffic versus segregated bus operation

As expected, the faster fare collection system (contactless card) with segregated bus operation provides the highest operating speed for all demand levels, whilst the lowest speed is obtained with cash payment in mixed traffic (for both peak and off-peak traffic).

A more interesting outcome comes from the comparison of slow boarding on fast buses (cash payment on segregated busways) against a quick bus boarding system combined with a low bus speed (contactless card payment in mixed-traffic); in this case the superiority of one or the other is given by the demand level as shown by the curves that intersect each other in Figures 4.7a and 4.7b. Under congested conditions, a busway provides a higher bus operating speed for demand up to  $N_{km}^* = 3$  pax/bus-km (Figure 4.7a, intersection of curves “Card payment, peak traffic” and “Cash payment busway”), whereas with mild congestion the threshold is  $N_{km}^* = 1.6$  pax/bus-km (Figure 4.7b). Implementing a prepaid fare collection system outperforms busways beyond demand  $N_{km}^*$ . In summary, the fact that buses spend most of the travel time running between stops does not alone suggest a preference for on-road bus priority schemes over tools aimed at reducing dwell times at bus stops; the bottom line is identifying if the bottleneck is on the road or at bus stops in the first place, in order and to apply a corrective measure accordingly.

#### **4.5.6 Estimation of benefits from upgrading the fare collection system: fleet size, travel time, operator and environmental cost savings**

In this section we estimate a number of potential benefits associated with improving the boarding and alighting process. The approach outlined can be used as an input for a wider cost benefit analysis that needs to consider transaction, implementation and operating costs of upgrading the fare payment system (Wright and Hook, 2007), the implications for demand (Balcombe *et al.*, 2004), type of passengers (number of fare paying passengers, students, elderly, etc.), capacity to handle different fare structures, etc.

First, we examine the fleet size requirement and bus in-service time as a function of frequency, passenger demand and fare payment strategy. The number of buses  $B$  required to provide a service on a single route is given by the product of the cycle time  $T$  (expression 4.12) and the frequency  $f$  [bus/h]:

$$B = [T \cdot f]^+ \quad (4.13)$$

where  $[\cdot]^+$  denotes the upper integer. In Table 4.8 we estimate the number of buses (Fleet size - FS) required for 5-minute headway (high frequency – 12 veh/h) and 20-minute headway (low frequency – 3 veh/h) services for different levels of demand. The results show that with a 20-minute headway the number of buses needed does not change in some demand ranges, e.g., between 4 and 6 pax/bus-km, three buses are required to provide the service regardless of the fare collection technique, whereas for a demand between 7 and 8 pax/bus-km, one bus is saved by having a prepaid fare collection system, relative to on-board cash payment. However, on a high frequency service, between one and four buses could be saved by upgrading the fare payment system and/or boarding rule, depending on the demand level.

Even when the same number of buses is required to provide the service with slow or fast fare payment methods, the travel time of buses is shorter if the fare payment technology is quicker, as shown by the resulting operating speeds in Figure 4.6, therefore there will be savings in terms of fuel consumption, labour and travel time cost for users. A measure of the operating cost saving is presented in Table 4.8 as the in-service time ratio STR (percentage of time in which buses are running the service), which demonstrate that even when the number of required buses is the same, more efficient fare payment systems require a lower utilization time for buses, which may be translated in further fleet size savings if buses are sequentially used in several routes that share a terminal.



**Table 4.8: Fleet size (FS – buses) and in-service time ratio (STR)**

20-minute headway												
pax/bus-km	Cash T2B1		Magnetic strip T2B1		Contactless card T2B1		Magnetic strip T2B2		Contactless card T2B2		Off- board T2B2	
	FS	STR	FS	STR	FS	STR	FS	STR	FS	STR	FS	STR
1	2	96.7%	2	91.6%	2	91.0%	2	90.9%	2	90.1%	2	89.9%
2	3	72.1%	2	97.9%	2	96.7%	2	96.4%	2	95.0%	2	94.4%
3	3	79.3%	3	68.9%	3	67.7%	3	67.5%	2	99.1%	2	98.2%
4	3	86.1%	3	72.3%	3	70.7%	3	70.4%	3	68.5%	3	67.7%
5	3	92.7%	3	75.5%	3	73.5%	3	73.1%	3	70.7%	3	69.8%
6	3	99.3%	3	78.6%	3	76.2%	3	75.7%	3	72.8%	3	71.7%
7	4	79.3%	3	81.5%	3	78.8%	3	78.2%	3	74.9%	3	73.5%
8	4	84.0%	3	84.4%	3	81.3%	3	80.6%	3	76.8%	3	75.3%
5-minute headway												
pax/bus-km	Cash T2B1		Magnetic strip T2B1		Contactless card T2B1		Magnetic strip T2B2		Contactless card T2B2		Off- board T2B2	
	FS	STR	FS	STR	FS	STR	FS	STR	FS	STR	FS	STR
1	8	96.7%	8	91.6%	8	91.0%	8	90.9%	8	90.1%	8	89.9%
2	9	96.2%	8	97.9%	8	96.7%	8	96.4%	8	95.0%	8	94.4%
3	10	95.1%	9	91.9%	9	90.3%	9	90.0%	8	99.1%	8	98.2%
4	11	93.9%	9	96.4%	9	94.3%	9	93.9%	9	91.3%	9	90.3%
5	12	92.7%	10	90.6%	9	98.1%	9	97.5%	9	94.3%	9	93.0%
6	12	99.3%	10	94.3%	10	91.5%	10	90.9%	9	97.1%	9	95.6%
7	13	97.6%	10	97.9%	10	94.6%	10	93.9%	9	99.8%	9	98.0%
8	14	96.0%	11	92.1%	10	97.6%	10	96.8%	10	92.2%	10	90.3%

If  $c$  is a unit of bus cost per hour [\$/bus-h], the cost saving  $\Delta C$  [\$/h] from speeding up the passenger service time from  $t_{D1}$  to  $t_{D2} < t_{D1}$  is given by (4.14).

$$\Delta C = c f P L (t_{D1} - t_{D2}) \quad (4.14)$$

where  $P$  is the bus demand [pax/bus-km] and  $L$  is the length of the route [km]. Both operating and external costs can be expressed as a cost  $c$  per bus-hour of operation. For instance, travel time savings due to upgrading the fare payment method can be translated into fuel and (potentially) labour cost savings for operators, and environmental benefits as reductions of air pollution. These effects are monetised next.

Based on Frey *et al.* (2007), we estimate a fuel consumption rate of 2.52 litres per hour in idle time (while bus is at bus stop) for a diesel technology, which is purchased at \$1.20

per litre (2011 Australian dollar, AUD)<sup>32</sup>, therefore, the hourly cost of fuel in idle time is estimated as  $c_1 = 3.02$  \$/bus-h. On the other hand, the average labour cost of bus drivers in Australia is  $c_2 = \$29.90$  \$/bus-h (Hensher, 2010). Thirdly, we can estimate benefits due to the reduction of air pollution. Attaching a social or external cost to gas emissions is highly variable depending on several factors such that fuel technology, vehicle technology and population density. Watkiss (2002) estimates a marginal environmental cost of \$0.58 (2002 AUD) per litre of fuel consumed by diesel buses in Australia<sup>33</sup>.

Assuming an average fare paying demand of 4 pax/bus-km, the fuel, labour and environmental cost savings of upgrading from an on-board cash payment to one of the other payment systems are given in Tables 4.9, 4.10 and 4.11, respectively, as a function of the service headway, i.e., the inverse of frequency.

**Table 4.9: Fuel cost savings with respect to on-board cash payment**

Headway [min]	Magnetic strip <i>T2B1</i> [\$/h]	Contactless card <i>T2B1</i> [\$/h]	Magnetic strip <i>T2B2</i> [\$/h]	Contactless card <i>T2B2</i> [\$/h]	Off- board <i>T2B2</i> [\$/h]
30	0.83	0.93	0.95	1.07	1.11
20	1.25	1.39	1.42	1.60	1.67
15	1.67	1.86	1.90	2.13	2.22
10	2.50	2.79	2.85	3.20	3.34
5	5.01	5.57	5.70	6.39	6.67
2	12.52	13.93	14.24	15.98	16.68

<sup>32</sup> AUD 1 ≈ USD 1 ≈ EUR 0.76 on December 2011.

<sup>33</sup> For buses manufactured between 1996 and 1999. Value is representative of inner areas of large capital cities (Melbourne, Sydney, Brisbane, Adelaide and Perth). Pollutants considered are NO<sub>x</sub> (oxides of nitrogen), PM (particulates), HCs (hydrocarbons), SO<sub>2</sub> (sulfur dioxide), CO (carbon monoxide) and CO<sub>2</sub> (carbon dioxide). This value is given only with illustration purposes and might not be representative of the current 2011 situation as gas emission rates have likely changed due to technological improvements in bus emissions.

**Table 4.10: Potential labour cost savings with respect to on-board cash payment**

Headway [min]	Magnetic strip <i>T2B1</i> [\$/h]	Contactless card <i>T2B1</i> [\$/h]	Magnetic strip <i>T2B2</i> [\$/h]	Contactless card <i>T2B2</i> [\$/h]	Off- board <i>T2B2</i> [\$/h]
30	8.25	9.19	9.39	10.54	10.99
20	12.37	13.78	14.08	15.80	16.49
15	16.50	18.37	18.77	21.07	21.99
10	24.75	27.56	28.16	31.61	32.98
5	49.50	55.11	56.32	63.21	65.96
2	123.75	137.78	140.81	158.03	164.89

**Table 4.11: Environmental cost savings with respect to on-board cash payment**

Headway [min]	Magnetic strip <i>T2B1</i> [\$/h]	Contactless card <i>T2B1</i> [\$/h]	Magnetic strip <i>T2B2</i> [\$/h]	Contactless card <i>T2B2</i> [\$/h]	Off- board <i>T2B2</i> [\$/h]
30	0.40	0.45	0.46	0.51	0.54
20	0.60	0.67	0.69	0.77	0.81
15	0.81	0.90	0.92	1.03	1.07
10	1.21	1.35	1.38	1.54	1.61
5	2.42	2.69	2.75	3.09	3.22
2	6.05	6.73	6.88	7.72	8.06

As an example, upgrading from cash to on-board contactless card payment with boarding at the front door only (*T2B1*), keeping the headway constant at 10 minutes, would yield savings of \$2.79 per hour on fuel, and up to \$27.56 on labour (depending on the percentage of total time savings that can actually be translated into a reduction of driver work hours), plus \$1.35 per hour on environmental benefits. As expected, larger cost savings are accruable for high frequency (short headway) services. Note that for a different value of average demand  $N_{km2}$  [pax/bus-km], savings from Table 4.9 to 4.11 simply need to be amplified by  $N_{km2}/4$ , being 4 pax/bus-km the demand used for the calculations. Also, the cost savings of moving between any of the other fare passenger boarding alternatives are obtained by subtracting the respective values in Tables 4.9 to 4.11.

Finally, the benefits for users due to reductions in bus travel time are estimated, which depend on their travel distance and willingness to pay for time savings. Using 18.3 \$/h as the value of in-vehicle time savings (estimated in Section 6.3) and assuming that the

average travel distance for passengers is 6.4 km (which is the average travel distance for bus passengers in Sydney, TDC, 2010), the user cost savings per passenger, with respect to the on-board cash payment system, are presented in Table 4.12. Again, as the saving of time is proportional to demand, so is the benefit perceived by users.

**Table 4.12: Average user cost savings with respect to on-board cash payment**

Demand [pax/bus-km]	Magnetic strip <i>T2B1</i> [\$/h]	Contactless card <i>T2B1</i> [\$/h]	Magnetic strip <i>T2B2</i> [\$/h]	Contactless card <i>T2B2</i> [\$/h]	Off- board <i>T2B2</i> [\$/h]
1	0.26	0.29	0.29	0.33	0.33
2	0.50	0.57	0.57	0.64	0.67
3	0.76	0.83	0.86	0.98	1.00
4	1.00	1.12	1.14	1.29	1.33
5	1.26	1.41	1.43	1.62	1.69
6	1.53	1.69	1.72	1.93	2.03
7	1.76	1.98	2.00	2.26	2.36
8	2.03	2.24	2.29	2.57	2.69

The generalisability of the quantitative results obtained in this section depends on the accuracy of the passenger service time estimations used for the calculations (Table 4.5). These figures are obtained based on average boarding times estimated in Sydney and Santiago. In general, passenger boarding and alighting times depend on several factors, namely the existence of steps at the entrance of doors (York, 1993), width of doors (Fernández *et al.*, 2010), proportion of seniors and students among passengers (Tirachini, 2011) and crowding and friction effects (Milkovits, 2008), which are specific to each local situation, and consequently, transferability of the results to other places is not guaranteed. However, the analysis in this section is parametric on *PST* and therefore, the estimation of benefits with other values for boarding and alighting times (from other countries or obtained under different assumptions) is straightforward.

#### 4.6 Conclusions

In this chapter it has been argued that with the current availability of several technological options for the choice of a fare payment system on bus services, together with the assignment of doors to the processes of boarding and/or alighting, the time that

buses spend at bus stops boarding and alighting passengers can be manipulated; hence the boarding and alighting time per passenger becomes an important decision variable instead of being treated as exogenous parameters, as traditionally assumed in the microeconomic literature of public transport operations.

Based on a series of dwell times surveys in Sydney and a study in Santiago, we estimated average passenger service time (including boarding and alighting) for four fare payment methods (off board payment and on board with cash, magnetic strip and contactless card), when bus boarding is allowed only at the front door or at all available doors (Section 4.3). This information is later used to analyse, on the one hand, the effect of upgrading the fare collection technology, and on the other hand, the effect of increasing the number of doors in which boarding is permitted. In Section 4.4 we find that the time savings due to upgrading the payment technology (*technology effect*) and due to increasing the number of doors to board (*door effect*) are not independent, and rather depend on the bus size. The scenarios with boarding through all doors are always more time efficient than their counterparts with boarding at the front door only, for a given number of doors.

In Section 4.5, the impact of alternative fare collection systems and boarding policies is analysed over an empirically estimated bus travel time function in Sydney, which revealed that upgrading the fare payment technology has a major impact on performance when boarding is allowed at the front door only; but this technology effect diminishes when boarding is allowed at all doors, especially if big buses with four doors are used. We estimate savings on fleet size requirements, fuel and labour cost, travel time for users and air pollution. This is a novel application in the literature on bus travel times, and shows how the benefits from upgrading the fare payment system from slower to quicker techniques increase with demand and bus frequency. Thus, the analysis performed in this study can be applied to other bus systems by transit agencies and transport policy makers, to make a more informed decision on bus service provision alternatives when considering several alternatives referring to fare collection policy, for the implementation of new services or the enhancement of existing systems.

The relative merits of upgrading the fare collection system as a measure to reduce travel times are compared with the commonly suggested policy of segregating buses from car traffic by implementing bus lanes or busways. It is shown that the superiority of one or the other, as stand-alone policies, is given by the demand level, as implementing a faster fare payment technology can provide lower bus travel times for high demand.

## Chapter 5

# Determinants of Bus Congestion and its Inclusion in the Economic Analysis of Transport Policies

### 5.1 Introduction

Traditionally, microeconomic models for the operation of urban bus services assume that the travel time in between bus stops of buses is fixed, i.e., the bus running time is not influenced by bus frequency (e.g., Mohring, 1972; Jansson, 1980; Kuah and Perl, 1988; Jara-Díaz and Gschwender, 2003a) but may be influenced by car flow, as empirically found in regression models where car flow plays a role in determining bus travel time<sup>34</sup>. The assumption of no delays due to bus frequency is plausible for services in which the frequency is relatively low, such that there is no noticeable bus interaction due to bunching or queuing delays behind bus stops. Nevertheless, as frequency grows, it is more likely that buses will arrive at bus stops when there are other buses transferring passengers, therefore bus queues may arise before bus stops. This is a relevant issue for pricing analysis and frequency setting, as the existence of frequency-induced congestion increases bus travel time for users and operators, in contrast to the economies of scale effect on reducing waiting times (Kerin, 1992).

---

<sup>34</sup> The traffic flow influence on buses is considered either implicitly -through time-of-day specific dummy variables to account for peak and off-peak periods, as in the model of Section 4.5- or explicitly -using car flow as an explanatory variable of bus travel time (McKnight *et al.*, 2003).

The existence of delays due to the interaction of buses with each other -and with cars in the case of mixed traffic operation- is referred to as “bus congestion”. In practice, we find that bus congestion is present even in cities that have a very low total bus usage. For example, in the city of Sydney the number of trips by bus is between 5.5 and 5.8 percent of the total on the period 2001-2009 (TDC, 2010), however if we focus on the CBD only, the modal split of public transport (including bus and rail) for commuting trips is between 73 and 76 percent, with very high frequencies of bus services in the morning and afternoon peak periods (e.g., more than 40 buses per hour in George Street, the main CBD road) that cause bus queues before stops and intersections.

In this chapter, a method to include bus congestion in the economic analysis of bus service provision is proposed and operationalised, which has effects on the optimal design of the system and pricing level. Section 5.2 presents a review of bus congestion in the literature. The estimation of queuing delays is presented in Section 5.3, while the link between bus congestion, fare collection technology and bus boarding policy is discussed in Section 5.4. Finally, conclusions are provided in Section 5.5.

## **5.2 Bus Congestion in the Literature**

In general, the treatment of bus congestion is limited in the existing literature. A technical problem for the introduction of bus congestion in formal microeconomic analysis is that the bus congestion technology has not been realistically understood and defined, because of the myriad number of factors that intervene on how buses interact with each other, with other modes (e.g., cars, trucks, motorcycles, bicycles) and with passengers in an urban environment.

Jara-Díaz and Gschwender (2003a) postulate a general model in which the bus running time  $t_b$  is a function of frequency, but no functional form for the relationship is provided, while the few authors that assess bus congestion do so by applying to buses flow-delay functions borrowed from car traffic models, such as the linear function (5.1) implemented by Ahn (2009) in his analysis of bus services and road pricing, and the



Bureau of Public Roads (BPR) function (5.2) used by Fernández *et al.* (2005) in their analysis of bus cost structure, and by Wichinsin *et al.* (2007) in their analysis of car, bus and rail pricing.

$$t_b(f_b) = t_{b0} + t_{b1}(f_a + \varphi f_b) \quad (5.1)$$

$$t_b(f_b) = t_{b0} \left[ 1 + \alpha_0 \left( \frac{f_a + \varphi f_b}{K_r} \right)^{\alpha_1} \right] \quad (5.2)$$

where  $f_a$  is the car flow,  $f_b$  is the bus frequency,  $t_{b0}$ ,  $t_{b1}$ ,  $\alpha_0$  and  $\alpha_1$  are parameters ( $t_{b0}$  is the free-flow travel time),  $\varphi \geq 1$  is the passenger car equivalency factor of a bus, and  $K_r$  is the capacity of the road. Expressions (5.1) and (5.2) assume that cars and buses share the right of way. Pels and Verhoef (2007) also choose a linear function to account for congestion in (independent) cars and trains. BPR or linear delay functions are commonly used to represent traffic congestion in traffic assignment models; and as an extension could be used to represent delays that buses face on the road, for example, due to intersections. However, travel time functions that depend only on flow measures (frequency in the case of buses) such as (5.1) and (5.2) do not explicitly account for the fact that buses have to stop to transfer passengers, an issue that is sometimes implicitly internalised by applying to buses a large passenger car-equivalency factor, e.g., to assume that a standard bus is equivalent to  $\varphi=4$  or 5 cars (Parry and Small, 2009). Therefore, the application of functions inspired in car traffic models as the only measure of congestion is incomplete for public transport, as, among other things, cars do not have to stop at bus stops and do not interact with passengers getting on and off. Bus stop congestion implies that frequency is not the only variable that triggers bus delays. A poor operation of bus stops or an inefficient boarding and alighting process can impact on the time that buses are stopped, possibly imposing delays on other buses even for relatively low frequencies.

In this work, two sources of bus congestion are considered:

- (i) On the road and intersections, commonly expressed as a static congestion function such as (5.1) and (5.2) above.
- (ii) At bus stops, in the form of queuing delays that arise behind bus stops when a bus arrives and all berths are being used by other buses. Note that the dwell

time, i.e., the time necessary to board and alight passengers, is not a form of congestion.

There exist a few studies that have advanced the understanding of the congestion associated with bus stops. When buses stop to transfer passengers, part of that delay may be transferred to cars if both modes run in shared lanes; and given that the dwell time of buses depends on the number of passengers transferred, car travel time would also be affected by the number of passengers boarding and alighting a bus. Koshy and Arasan (2005) analyse the influence of two types of bus stops -curbside and bus bays- on the running speed of other modes that share a road with buses in India (cars, trucks, motorised two-wheelers, autorickshaws and bicycles); it is found that curbside bus stops cause more congestion on other modes than bus bays, and the impact increases with the dwell time of buses. Zhao *et al.* (2007) show that road capacity reductions due to the operation of a bus stop depend on the location of the stop with respect to a signalised intersection (nearside, farside and stop-intersection distance<sup>35</sup>). Finally, Basso and Silva (2010) propose a non-linear function for bus frequency that accounts for the delay that cars experience when buses stop at a bus stop, in a way that the mean delay transferred to cars is small when bus frequency is low, and equals bus dwell time when bus frequency is high, and therefore, it is assumed that cars sharing the road have no option but to behave like buses.

In summary, the existing models are far from a realistic characterisation of the phenomenon of congestion when urban buses are involved. The inclusion of engineering or simulation models that deal with bus dynamics at bus stops (Fernández and Tyler, 2005; Fernández, 2010) into economic pricing analysis is a possible way forward to

---

<sup>35</sup> Bus stops are usually classified into three groups: (i) before an intersection or *nearside*, (ii) after an intersection or *farside*, and (iii) isolated from intersections or *midblock*. Each location has advantages and disadvantages that make impossible to give general recommendations over which one is superior without taking into account myriad local considerations like the programming of signalised intersections, the number of vehicles turning left or right at intersections, the geometry of bus access to the curb, the size of the bus stop, the distance between the bus stop and the nearest intersection, traffic safety, pedestrian interference with bus movements at bus stops and with general traffic at intersections, etc. (TRB, 1996, 2003). However, when bus stops are analysed in isolation from other bus stops upstream or downstream, authors tend to agree that generally farside stops yield shorter delays than midblock and nearside stops (TRB, 1996; Furth and SanClemente, 2006). On the other hand, when traffic signals are synchronised to facilitate car flow, buses can reduce overall delays by alternating nearside and farside bus stop locations (TRB, 2003; Vuchic, 2005).

improve our understanding of bus delays at bus stops and of congestion interactions in mixed systems, and its implications for pricing policy. In this chapter, a bus stop simulator is used to estimate queuing delays at bus stops as a function of bus frequency and size, number of berths and average dwell time. As dwell time depends on the number of passengers being transferred, the fare collection system and the boarding and alighting policy, we will be able to link bus congestion to the technological choice of a fare payment method and boarding regime, analysed in Chapter 4. Consequently, the proposed congestion function is more comprehensive than traffic borrowed formulae commonly used in the economic literature of urban transport.

### 5.3 Estimation of Queuing Delays at Bus Stops

The total time delay per bus stop consists of the (i) acceleration and deceleration delay  $t_{ac}$ , (ii) the average queuing time  $t_q$ , (iii) the dwell time  $t_d$  and (iv) an internal waiting delay  $t_{iw}$ . The queuing time is a measure of the external congestion caused by a bus stop, observed when a bus arrives at a stop and all berths are occupied. This delay is commonly present in high frequency services, but it may also occur in poorly controlled low frequency services where buses tend to bunch.

If  $f_b$  represents the bus frequency,  $N_b$  denotes passenger demand and  $\Delta$  represents a fare payment technology and boarding and alighting policy (which determines boarding and alighting times), we have:

$$t_q = t_q(f_b, t_d) \quad (5.3)$$

$$t_d = t_d(N_b, \Delta) \quad (5.4)$$

There is little research on the empirical estimation of  $t_q$ ; after analysing bus stop operations with the simulation model IRENE, Fernández *et al.* (2000) found that  $t_q$  grows exponentially with the frequency of buses that enter a bus stop. On the other hand, Lu *et al.* (2010) apply a Cellular Automaton model to simulate  $t_q$  on bus stops with multiple berths and multiple bus routes arriving. In this chapter, these previous works are

extended upon by estimating a queuing delay function that depends on the design of the bus stop (number and length of berths), bus length, bus frequency and average dwell time (the latter given by the number of passengers getting on and off, the fare collection system and the number of doors to board and alight).

As argued by Fernández and Planzer (2002), a simulation approach is well suited to analyse key performance measures of bus stops (like the queuing delay), because the processes involved in the arrival of buses, passengers and the interaction between them are very complex and usually random, which suggests that analytical steady-state approaches like the Highway Capacity Manual (HCM) formula to calculate bus stop capacity (TRB, 2000) have a limited real world applicability. Consequently, we use IRENE to estimate delay-frequency functions  $t_q$  that depend on frequency and average dwell time  $t_d$ . IRENE is a bus stop simulator that calculates the capacity, queuing delay, dwell time, berth usage and other indicators of the performance of a bus stop, as a function of a number of inputs such as the boarding and alighting demand, number of berths, stochasticity of both user and bus arrivals, etc. For a detailed description of the program see Gibson *et al.* (1989) and Fernández and Planzer (2002).

We consider linear bus stops with one, two or three linear berths, and four possible bus sizes: mini (8 m), standard (12 m), rigid long (15 m) and articulated (18 m). A description of the assumptions regarding bus stop location, berth length and bus saturation flow is presented in the Appendix A2. A total of 265 simulations were run encompassing all bus sizes and bus stop designs previously described for a range of frequencies from 20 to 220 bus/h and dwell times between 10 and 65 seconds. Buses are assumed to arrive at a constant rate at stops (no bus bunching) and bus stops are isolated from traffic lights.

The estimated model for the queuing delay  $t_q$  [s/bus] as a function of the bus length  $s_b$  [m], dwell time  $t_d$  [s/bus], frequency  $f_b$  [veh/h] and number of berths per bus stop is shown in expression (5.5)

$$t_q = 0.001 \left[ \beta_0 + \beta_{11} s_b + (\beta_{d1} + \beta_{d2} Z_2 + \beta_{d3} Z_3) t_d \right] e^{0.001 f_b [\beta_f + \beta_{12} s_b + (\beta_{d4} + \beta_{d5} Z_2 + \beta_{d6} Z_3) t_d]} \quad (5.5)$$

where  $\beta_0, \beta_{l1}, \beta_{l2}, \beta_{d1}, \beta_{d2}, \beta_{d3}, \beta_{d4}, \beta_{d5}, \beta_{d6}$  and  $\beta_f$  are estimated parameters and factors 0.001 are introduced for scaling of the parameters (see Appendix A2 for further details).  $Z_2$  and  $Z_3$  are dummy variables defined as follows:

$$Z_2 = \begin{cases} 1 & \text{if bus stop has two berths} \\ 0 & \text{otherwise} \end{cases}$$

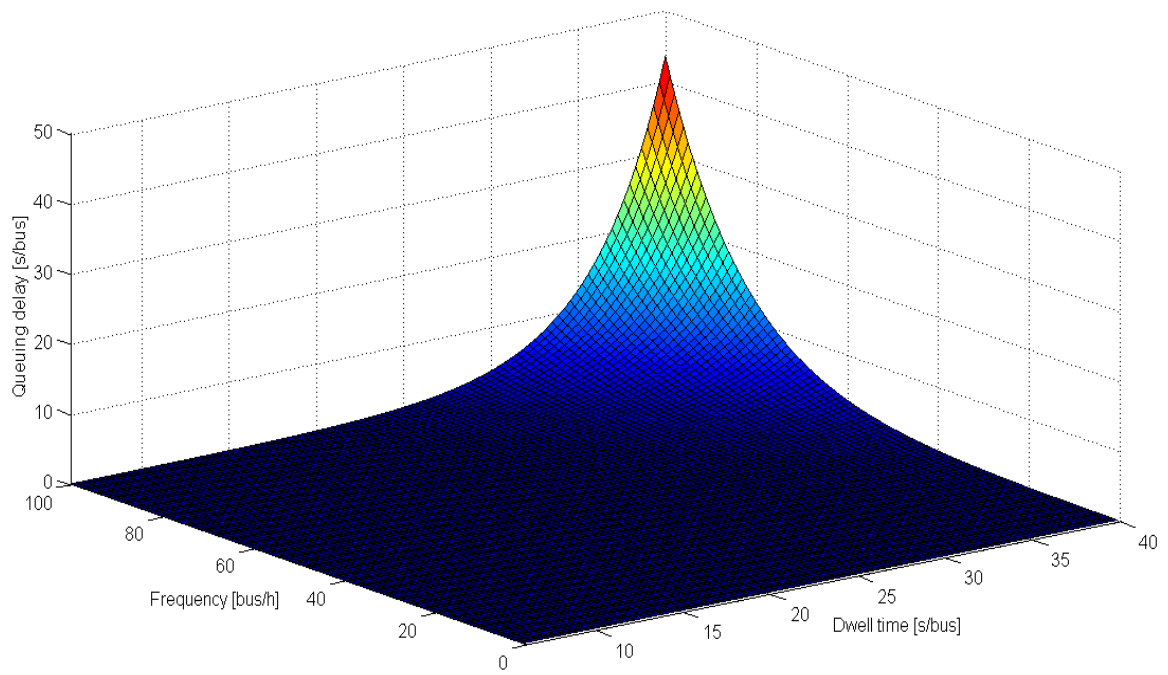
$$Z_3 = \begin{cases} 1 & \text{if bus stop has three berths} \\ 0 & \text{otherwise} \end{cases}$$

The case of split bus stops (a large stopping area consisting of two subgroups with one, two, or three berths each) can be accommodated by setting a rule for the assignment of buses to the stopping areas (e.g., 50 percent of buses to each stopping area). A similar expression to (5.5) was first proposed by Fernández *et al.* (2000), but with the boarding demand instead of the total dwell time as an explanatory variable, and therefore, the function from Fernández *et al.* (2000) is linked to a particular fare collection system (cash payment on board buses in Santiago, Chile). However, expressing the queuing delay (5.5) as a function of the dwell time makes it more general to accommodate the influence of different fare collection systems and boarding and alighting policies. The parameters estimated for equation (5.5) are presented in Table 5.1.

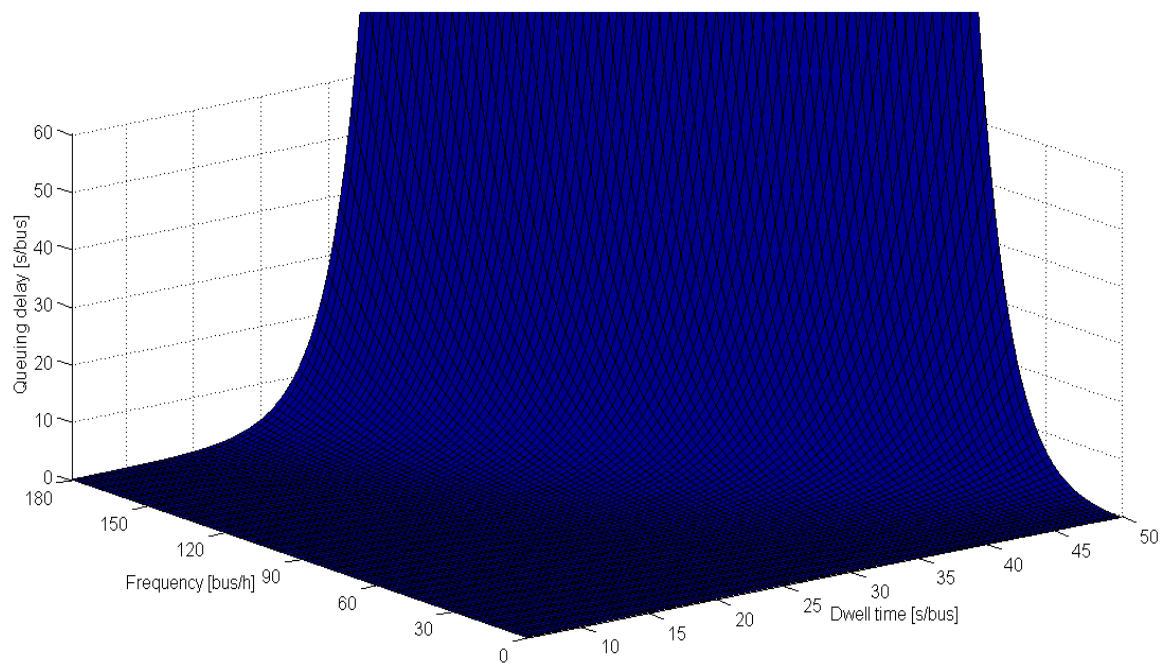
**Table 5.1: Queuing delay parameters**

<b>Parameter</b>	<b>Estimate</b>	<b>t-ratio</b>
$\beta_0$	-2.952	-3.328
$\beta_{l1}$	0.061	3.050
$\beta_{d1}$	2.185	4.123
$\beta_{d2}$	-1.903	-3.844
$\beta_{d3}$	-2.044	-4.008
$\beta_f$	23.089	31.935
$\beta_{l2}$	0.361	7.848
$\beta_{d4}$	1.807	19.857
$\beta_{d5}$	-0.374	-4.022
$\beta_{d6}$	-0.627	-7.207
$R^2$	0.921	
Sample size	265	

Figure 5.1 shows the estimated growth of queuing delay (5.5) as a function of frequency and dwell time, for bus stops of two berths and buses of 12 metres; both Figures 5.2a and 5.2b represent the same estimated function (5.5), but Figure 5.1b has a larger domain (frequency and dwell time) and its z-axis (queuing delay) has been cut at 60 s/bus in order to reveal the differences in queuing delays in the middle range of frequency and dwell time.



(a)



(b)

**Figure 5.1: Queuing delay as a function of bus frequency and queuing delay, two-berth bus stops**

The exponential nature of (5.5) is clear in Figure 5.1a, as queuing delay is negligible for low frequencies and dwell times, but it explodes quickly once a threshold is reached, up to 45 s/bus if frequency is 100 bus/h and dwell time is 40 seconds. The threshold depends on how high bus frequency is or how long dwell time is. For example, Figure 5.1b reveals

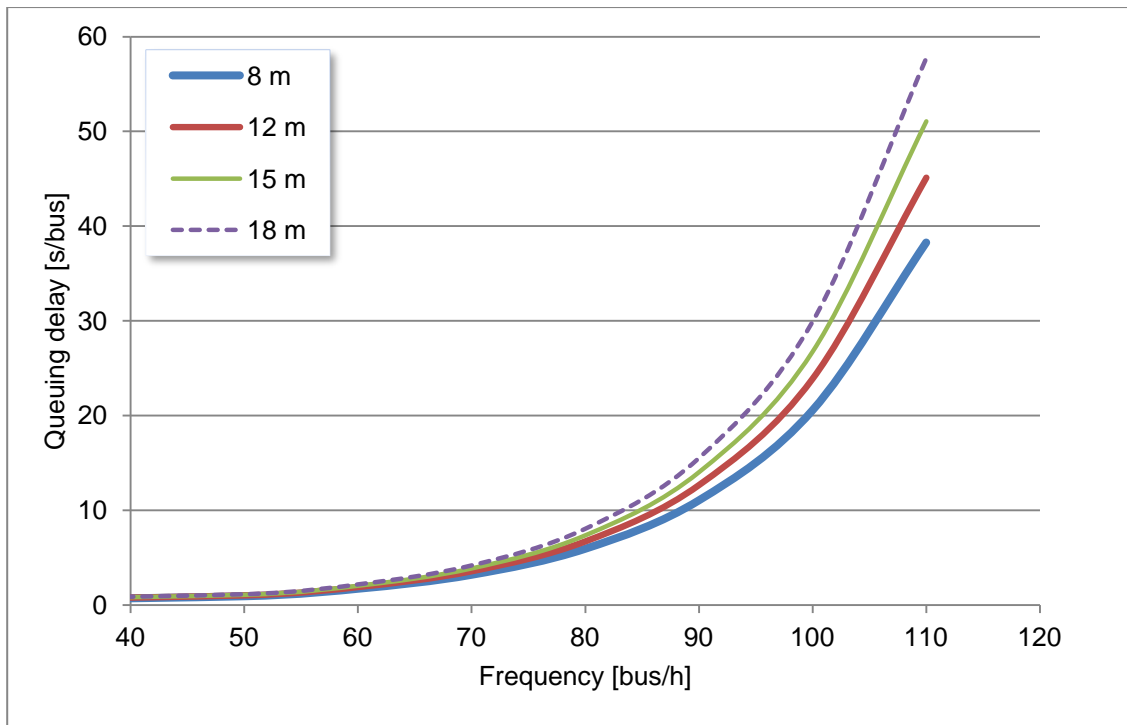
that an average dwell time of 40 seconds yields noticeable queuing delays for a frequency of 60 veh/h, whereas a dwell time of 20 seconds can operate with negligible average queuing delays for frequencies up to 100 veh/h<sup>36</sup>.

The particular influence of the number of berths and size of vehicles is illustrated in Figure 5.2. For a given frequency,  $t_q$  increases with bus size (Figure 5.2a), a difference that is amplified the more berths are provided on the bus stop (Figure 5.2b).

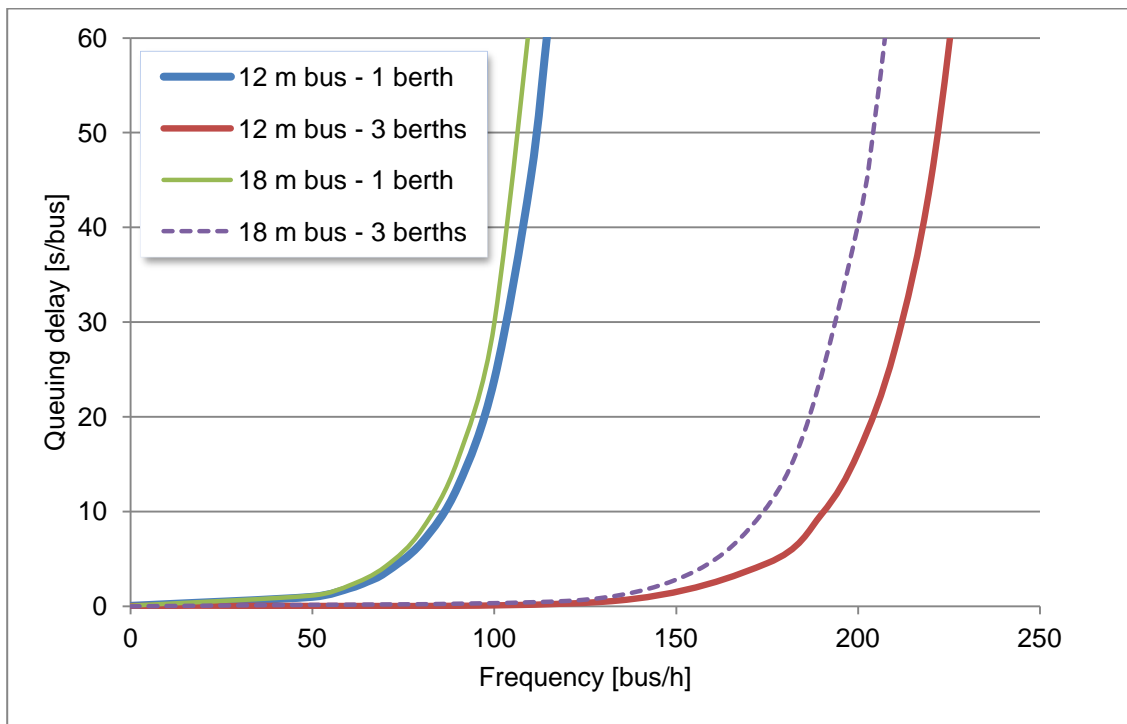
---

<sup>36</sup> This does not mean that all individual buses have no queuing delays, some vehicles actually have to wait in queue due to the randomness in dwell times on the preceding bus, but the average value of the queuing delay, over one hour of simulation, is close to zero.





(a) Average queuing delay for buses of 8, 12, 15 and 18 metres, 1 berth, dwell time=20 s



(b) Average queuing delay for buses of 12 and 18 metres, 1 and 3 berths, dwell time=20 s

Figure 5.2: Bus stop queuing delay in different configurations

#### 5.4 The Relationship between Bus Congestion, Fare Collection Technique and Bus Boarding Policy

Now that we have an estimation of the extent to which bus stop congestion (i.e., queuing delays at bus stops) is determined by the dwell time at bus stops, it is worth investigating the link between congestion and alternative fare collection systems and boarding policies, another issue that is missing in the extant literature. This is possible by embedding into dwell time  $t_d$  in queuing delay (5.5) the parameters for average boarding and alighting times with alternative fare collection technologies, estimated in Chapter 4. For example, for a bus with two doors dwell time  $t_d$  is estimated as:

$$t_d = \begin{cases} c_{oc} + b\lambda^+ + a\lambda^- & \text{if sequential boarding and alighting at all doors} \\ c_{oc} + \max\{b\lambda^+, a\lambda^-\} & \text{if boarding at front door and alighting at back door} \end{cases} \quad (5.6)$$

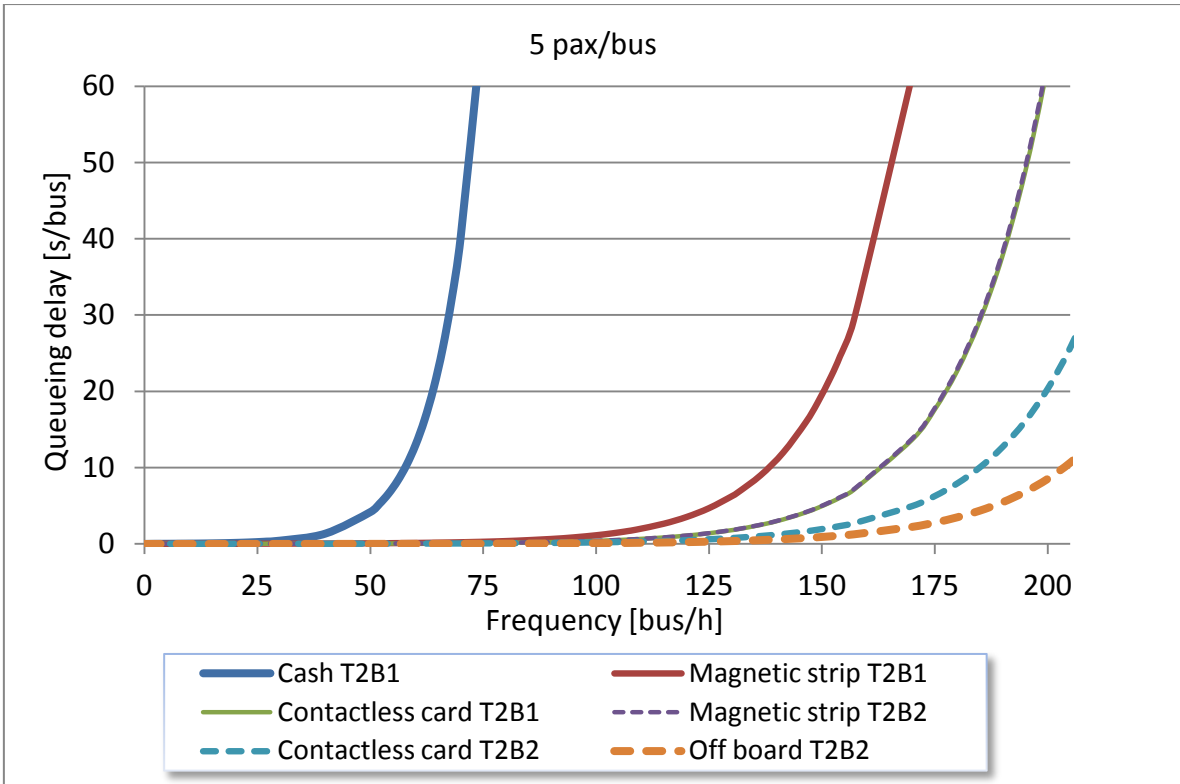
where  $c_{oc}$  is the dead time (which will be referred to as time to open and close doors),  $b$  and  $a$  are the average boarding and alighting time per passenger, respectively (which depend on the fare collection system), and  $\lambda^+$  and  $\lambda^-$  are the number of passengers getting on and off, respectively. Thus, introducing (5.6) into (5.5), queuing delays can be estimated as a function of the fare collection and boarding/alighting systems.

Existing bus routes in the urban context show that queuing delays are observed only in a subset of all bus stops, namely those stops with a high demand of passengers boarding or alighting passengers. In particular, based on the Sydney evidence, demand is usually concentrated in a few stops, and long dwell times (prone to cause queuing delays) are triggered at stations with a large number of passengers boarding rather than at stations with a large number of passengers alighting, because boarding is more time consuming than alighting. Consequently, we define as “high demand bus stops” stations with a large number of passengers boarding. For example, based on the dwell time surveys analysed in Section 4.3, we find that usually between 30 and 50 percent of the passengers boarding buses are concentrated at 10 to 20 percent of the bus stops along a route, and moreover, at those high demand bus stops, on average the number of passengers

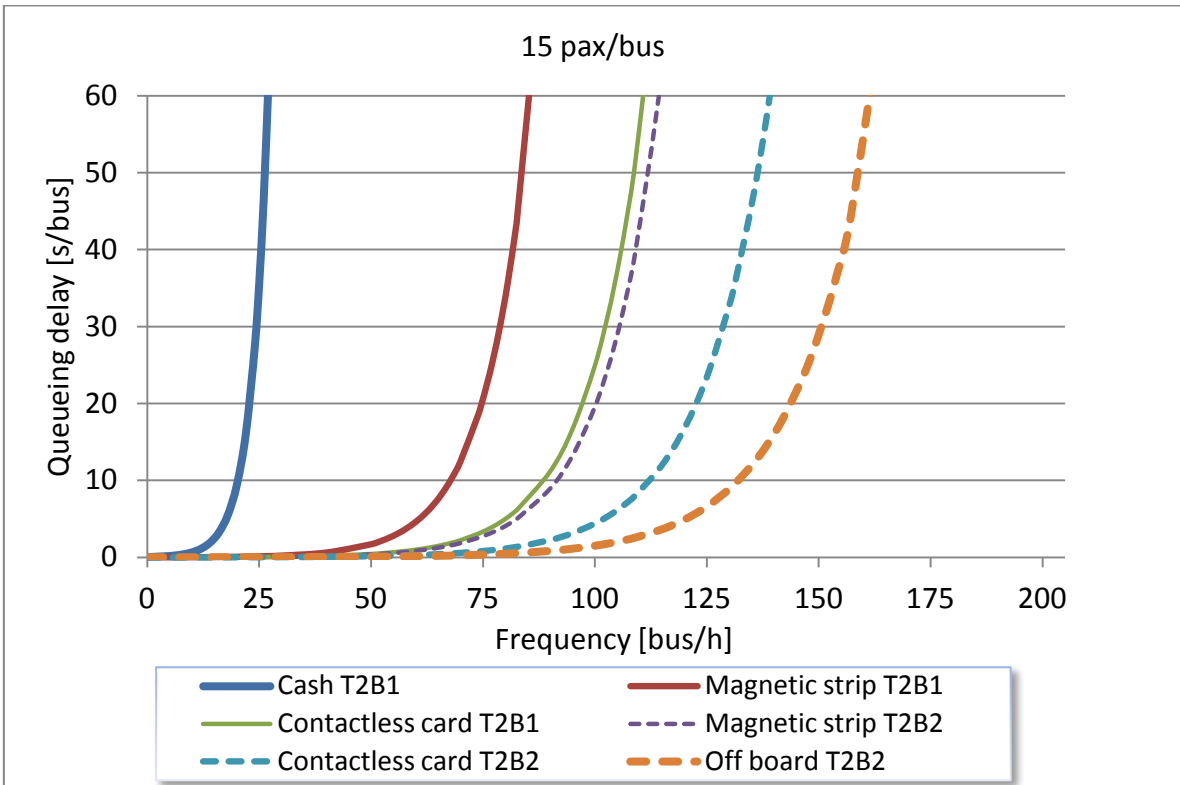
alighting buses is 20 percent of the number of passengers boarding. Therefore, we estimate queuing delays in bus stops (eq. 5.5) for two levels of boarding demand, 5 and 15 pax/bus, in which the number of alightings is 20 percent the number of boarding, i.e., 1 and 3 pax/h, respectively (using eq. 5.6 to estimate dwell time). Results for all fare payment and boarding policies defined in Section 4.2 are presented in Figure 5.3, for the case of two-door buses (T2B1 denotes simultaneous boarding at the front door and alighting at the back door, T2B2 denotes sequential boarding and alighting at both doors). The time to open and close doors ( $c_{oc}$ ) is 6.1 seconds, the average of the three values for  $c_{oc}$  in Table 4.3.

Figure 5.3 reveals that the frequency threshold that triggers queuing delays depends on the fare payment method, boarding and alighting policy, and boarding and alighting demand. As expected, the quicker a fare payment method and the lower the demand, the larger the flow of buses that can be accommodated without causing major delays, in other words, the larger the bus stop capacity. Therefore, a reform towards implementing a quicker fare collection system and/or to allow boarding buses at all available doors, not only provides shorter dwell times at bus stops or stations, but also has the potential of reducing (or eliminating) queuing delays and increase bus stop capacity, which is the main bottleneck in high-frequency bus routes.

Finally, Figure 5.3 reinforces a significant insight into the inclusion of bus congestion in the economic analysis of transport policies, in the sense that bus congestion not only depends on frequency or flow (as usually assumed for car traffic), but also on boarding demand at bus stops, the fare payment and boarding policies. Therefore, we have shown that models that use flow measures (including frequency only or frequency plus traffic flow) as the only explanatory variables for bus congestion are incomplete.



**(a) Boarding demand 5 pax/bus**



**(b) Boarding demand 15 pax/bus**

**Figure 5.3: Queuing delay for alternative fare payment and boarding policies**

## 5.5 Conclusions

In this chapter, an approach to incorporate bus congestion in the economic analysis of pricing policies and optimisation of bus services has been introduced, by estimating a function for the queuing delays that are observed behind bus stops when demand and frequency are high. We argue that this approach is more comprehensive than usual models that assume either that the travel time between two bus stops is fixed (no congestion), or that the bus congestion technology is only explained by delays borrowed from car traffic theory, like BPR or linear flow-delay functions. In fact, travel time functions such as (5.1) and (5.2) that represent travel time in-between bus stops, could be used in combination with a bus stop congestion function like (5.5) to provide a better representation of frequency induced bus delays along a route.

A bus stops simulator, IRENE, is used to estimate queuing delays at bus stops as a function of the dwell time, frequency of service, number of berths and size of vehicles. The estimated formula (5.5) is then used to investigate the influence on queuing delays of the number of passengers boarding and alighting buses, the use of alternative fare collection technologies and boarding policies. It is concluded that providing a faster boarding and alighting process not only reduces dwell times, but also has the potential of reducing bus congestion. The bus congestion approach developed in this chapter will be embedded into the public transport optimisation models of Chapters 7 and 8 in order to analyse the impact of congestion on the optimal value of variables like the bus frequency, bus size and distance between bus stops.

## **Chapter 6**

# **The Effects of Passenger Crowding on Public Transport Demand and Supply**

### **6.1 Introduction**

The empirical assessment of modal choice in transport has traditionally relied on time and cost as the main attributes influencing people's travel decisions. Nevertheless, with the improvement of both our understanding of the modal choice problem and analytical tools (e.g., advanced choice models), we have accumulated unambiguous evidence that shows how users take into account several qualitative aspects that enhance or harm the experience of travelling. In the case of public transport, this could include the number of passengers that have to share a bus or train, the quality of seats, the smoothness of the ride and the availability of air conditioning. This chapter is concerned with the first of these characteristics, i.e., the occupancy level on public transport vehicles and stations, and more specifically, the case in which there is a significant number of people sharing a limited space while travelling, which is usually referred to as crowding.

Crowding at bus stops, at rail stations, on buses and trains is becoming a major concern to service providers as they struggle to cope with increased public transport demand. Together with travel time, cost, trip time reliability and service headway (or frequency), crowding is now seen as having a significant influence on modal choice through the value

attached to reducing crowding in all its definitional variants. The cost of crowding, summed up over a whole public transport network might be substantial; for example, in the case of Sydney, it is estimated that the passenger crowding cost in the rail network is around \$82 million per annum (Wang and Legaspi, 2012). The inclusion or omission of the crowding cost is expected to influence optimal frequency, bus size and fare level, among other variables.

A technical advantage of the concept of crowding is that it can be quantitatively assessed, although there is no a single measure of the crowding phenomenon. The most common metric used in quantitative assessment is the occupancy rate or load factor, which is defined as the ratio between the actual number of passengers inside vehicles and the number of seats (Whelan and Crockett, 2009). Other authors use the nominal capacity of a vehicle (including both seating and standing) to measure the load factor (Oldfield and Bly, 1988; Jara-Díaz and Gschwender, 2003); using this definition we could suggest that, for example, if the load factor is over 80 percent a vehicle can be regarded as crowded. However, none of the load factor definitions provide a clear picture of the degree of crowding for passengers standing, which is more accurately captured by computing the density of standees per square metre (Wardman and Whelan, 2011). For example, a load factor of 150 percent, relative to the seating capacity, indicates that one out of three passengers is standing, but it does not say anything about the crowding conditions of those standing. On the other hand, a standing density of four or five passengers per square metre is an unmistakable indicator of crowding, regardless of the size or capacity of a bus or train.

This chapter presents a review of multiple dimensions of crowding effects on public transport demand and supply, including the impact of crowding on travel time, waiting time, value of travel time savings, optimal supply and pricing (Section 6.2). Next, crowding cost functions estimated using data from Sydney are presented (Section 6.3), in order to analyse how users value time under uncrowded and crowded conditions in terms of the number of available seats inside a vehicle and the density of standees. The ultimate goal is including the estimated taste parameters in the social welfare

maximisation model of Chapter 8. Finally, the main conclusions of the chapter are discussed in Section 6.4.

## 6.2 Effects of Crowding

### 6.2.1 Effect on in-vehicle time

When buses and trains circulate with a low number of passengers, everyone is able to find a seat, transfer of passengers at stations is smooth, and passenger-related disruptions that impose unexpected delays are rare. As the number of passengers increase, a threshold is reached at which not everyone is able to find a seat and some users need to stand inside vehicles. In turn, this may make more difficult the movement of other passengers that need to board to or alight from a vehicle, as shown in Figure 6.1. Therefore, riding time increases due to friction or crowding effects among passengers.



**Figure 6.1: Crowding can slow down both alighting and boarding of passengers**  
(Photos: Bogotá's Transmilenio, source: Wright and Hook, 2007)

The crowding effect on increasing boarding and alighting times has been captured by a number of authors who have estimated dwell time functions for trains and buses under uncrowded and crowded operation. Lin and Wilson (1992) estimate dwell time models for light rail trains in the Massachusetts Bay and find a statistically significant friction effect between passengers alighting and those standing at stations to board, and between passengers boarding and those that are standing inside trains. The authors estimate linear and non-linear dwell time models on crowding, with the latter providing a slight better fit to the observed data than the former. A later analysis over the same light



rail system by Puong (2000) showed that the interaction between boarding passengers and through standees is well explained by a cubic term on the number of passengers standing around a door; the average boarding time is 2.3 seconds per passenger in uncrowded conditions but raises to 2.9 and 4.4 seconds per passenger with 10 and 15 through standees per door, respectively.

In the case of buses, using data from Chicago Mikovits (2008) finds that dwell time increases with the square of the number of standees inside a bus, multiplied by the total number of passengers boarding and alighting at a bus stop. Like in the previously described rail models, this quadratic term captures the increased friction amongst passengers when the number of standees is high. Other authors have found average boarding and alighting times per passenger that depend on the number of passengers boarding and alighting (Dueker *et al.*, 2004; Fernández *et al.*, 2009), i.e., the length of the queue to board may speed up or slow down the boarding process.

The limited capacity of bus stops and train stations may also represent a problem if a large volume of passengers need to be handled at the same time, particularly in those stations in which many bus services stop. In such cases, some passengers may take longer to reach a door to board a vehicle if several other people are standing on his/her way, or obstructing his/her line of sight to sign and approach an incoming bus (TRB, 2003; Jaiswal *et al.*, 2007, 2010). Passengers inside buses may also face difficulties leaving a vehicle if the station is crowded. These station-related crowding issues have also been analysed in the literature; for example Lin and Wilson (1992) estimate the marginal friction effect between passengers alighting and those standing at stations to board, while Gibson *et al.* (1997) in Santiago and Jaiswal *et al.* (2010) in Brisbane found that the boarding time per passenger also depends on how congested is the platform on bus stations.

In summary, there is strong evidence that supports the fact that travel times increase when bus stops, train stations, buses and trains get crowded, which has a negative impact on both users and operator costs.

### **6.2.2 Effect on waiting time**

When the number of passengers is low relative to the capacity of the system, users are able to board the first vehicle that arrives at their bus stop or train station, and therefore the waiting time at stations is given by a fraction of the headway between two consecutive vehicles. Nonetheless, when the occupancy rate is high, having a limited capacity becomes an issue, as the chance of buses circulating full in some sections increases, which consequently implies that passengers waiting to board are left behind, increasing waiting time and the discomfort of travel. A formal treatment of this phenomenon was presented by Oldfield and Bly (1988) in their analysis of optimal bus size; they proposed that average waiting time is related not only to the headway (the inverse of bus frequency), but also to the occupancy rate or crowding level in an additive or multiplicative way.

The effect of high demand on increasing waiting times for passengers has received considerable attention in the literature on passengers' assignment to public transport networks. Spiess and Florian (1989) considered that the travel cost per link is a function of the passengers flow, to internalise the fact that waiting time and in-vehicle comfort may be a function of how many passengers use the service. On the other hand, Cominetti and Correa (2001) and Cepeda *et al.* (2006) model waiting time as inversely proportional to the *effective frequency*, which is a function of the actual frequency that decreases with the occupancy rate of buses upstream of a bus stop. The assignment model of Kurauchi *et al.* (2003) introduce that passengers may be risk-averse in their behaviour regarding what line or service to use, and therefore, be more prone to choose routes in which occupancy levels are lower as a way to reduce the chance of failing to board a bus (for the effect of seating and standing probabilities on route choice, see Section 6.2.4). In real-world applications, the increase in waiting time due to capacity constraints has been considered in the estimation of public transport load and demand in large scale scenarios including London (Department of Transport, 1989; Maier, 2011), Winnipeg, Stockholm and Santiago (Florian *et al.*, 2005), Los Angeles and Sydney (Davidson *et al.*, 2011) and San Francisco (Zorn *et al.*, 2012).

A second effect of high occupancy levels on waiting times is the possibility of triggering bus bunching (Abkowitz and Tozzi, 1987). When a bus is full and does not stop to pick up passengers at a bus stop (or if it stops but it is unable to load all passengers waiting), a larger number of passengers than is expected are left to wait for the next bus, which will need to stop for a longer period of time to board the increased number of passengers, presuming it too has capacity to accept the additional passenger load. As such, this second bus will likely be delayed and run late decreasing its headway relative to the next bus behind, and increasing its headway with respect to the next bus ahead, a phenomenon that is amplified as buses advance along the route if control measures like bus holding are not applied (Sun and Hickman, 2008; Daganzo, 2009; Delgado *et al.*, 2009; Sáez *et al.*, 2012). In short, bus bunching leads to variability in headways, which increases average waiting time (Welding, 1957).

### **6.2.3 Effect on travel time reliability**

We have discussed that when the occupancy of buses or trains approaches capacity, there might be an increase in both waiting and in-vehicle times. The inherent randomness of public transport demand, however, makes those delays difficult to predict. In other words, when occupancy rates are always low, users know that they will board the first bus that approaches their stops; nevertheless when the occupancy rate is high on average, passengers do not know for sure if the next bus will have spare capacity or will be full, implying having to wait for at least another bus, i.e., there is an increase in waiting time up to a probability. This is a source of unpredictability of travel times, which adds to the generalised cost of travel beyond an increase in average waiting time, because a higher variability in travel times is negatively valued by travellers as shown by the growing body of research on travel time variability and reliability (e.g., Senna, 1994; Bates *et al.*, 2001; Bhat and Sardesai, 2006; Li *et al.*, 2010).

A second issue worth of note is the likely relation between high crowding levels and the occurrence of incidents at bus stops or train stations, which is a source of unexpected delays that affects the service performance and reliability (beyond the phenomenon of bus bunching mentioned in Section 6.2.2). A common example of this situation is the case of passengers blocking the closing of doors in trains in order to enter a crowded carriage,

thereby introducing an extra delay in the process of closing doors (that might include several seconds for safety reasons).

#### **6.2.4 Effect on the valuation of travel time savings and route choice**

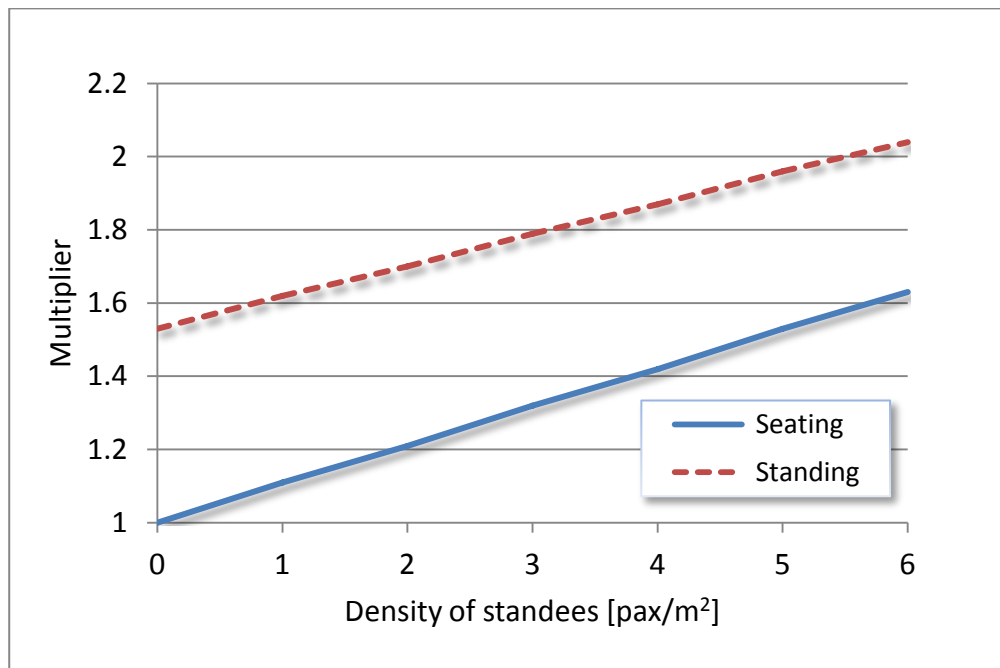
Users dislike travelling in crowded conditions due to a number of reasons, including the discomfort of sharing a limited space with several people, a feeling of ‘invasion of privacy’ (Wardman and Whelan, 2011) or a possible loss in productivity for passengers that work while sitting on a train (Fickling *et al.*, 2008). Consequently, crowding levels inside vehicles and at bus stops or train stations do have an impact on travel decisions. Thus, we can define a crowding cost, crowding externality or crowding penalty, which arises in some way as the occupancy levels of vehicles or transfer stations increase. Intuitively, we could expect users to be willing to pay more to reduce their travel time if they travel in a bus with an average occupancy of four passengers per square metre, than in the case in which a bus has a few passengers, all comfortably seated. Then, a relationship between crowding and the value of travel time savings (VTTS) is expected to exist, as empirically found by Maunsell and Macdonald (2007), Whelan and Crockett (2009) and Hensher *et al.* (2011) among others<sup>37</sup>. Moreover, this relationship may not be linear as an extra passenger per bus or train does not impose the same cost on everyone else when the occupancy level is 20 or 95 percent (measured against total capacity). The effect of crowding at train stations on increasing the discomfort of travellers has been estimated by Lam *et al.* (1999) and Douglas and Karpousis (2005).

A common methodology used within the literature to derive preference or utility functions for crowding is discrete choice models. A usual outcome of discrete choice models that include a crowding parameter on the valuation of in-vehicle travel time savings is the estimation of a “crowding multiplier”, i.e., a factor that multiplies the value of in-vehicle time savings found under uncrowded conditions. For example, Whelan and Crockett (2009) estimated the crowding multiplier for rail services in England, as a function of either the load factor (defined as the total number of passengers inside a

---

<sup>37</sup> For recent reviews of crowding valuation studies, see Wardman and Whelan (2011) and Li and Hensher (2011).

vehicle, over the seating capacity) or the number of passengers standing per square metre. The results for the latter case are quite illustrative, as shown in Figure 6.2.



**Figure 6.2: Crowding multiplier for passengers seating and standing**  
(Source: adapted from Whelan and Crockett, 2009)

For passengers seating, the crowding multiplier increases from 1.0 to 1.63 as the density of standing passengers increases from zero to six passengers per square metre, whereas for passengers standing these figures are 1.53 and 2.04, respectively. Figure 6.2 confirms intuition, as passengers standing have a higher willingness to pay to reduce travel time than passenger seating (when the former have not chosen to stand, but rather have to do it because all seats are taken), the discomfort of travelling of passengers seating and standing increases with the number of standees.

The disutility of standing may influence route choice when passengers have multiple alternatives to complete a trip. This has been recently incorporated into public transport assignment models like Sumalee *et al.* (2009), Leurent and Liu (2009), Schmöcker *et al.* (2011) and Hamdouch *et al.* (2011), who estimate the probability of getting a seat both when boarding a bus, and once on board if the passenger has to stand at the beginning of his/her trip. Passengers choose departure time and route according to their perceived travel disutility, which includes the probability of getting a seat (or failure to do so) as a

key attribute. Numerical applications show that the perceived seat availability may have a significant influence on both departure time and route choice; for example, Leurent and Liu (2009) found that the predicted passenger load in the Paris metro is reduced by around 30 percent when applying a model with different seat/stand disutilities, relative to a model that does not distinguish seating from standing.

#### **6.2.5 Impact of crowding externality on optimal supply and fare**

Crowding as a factor that affects the users' generalised cost of travelling has been recognised by several authors in the analysis of road and public transport pricing policy (Jansson, 1979; Kraus, 1989, 1991; Jansson, 1993; Arnott and Yan, 2000; Huang, 2002; Pedersen, 2003; Pels and Verhoef, 2007; Parry and Small, 2009). The basic idea is that when a person boards a bus or a train, he or she may impose a discomfort externality on everyone else on board, which is especially noticeable when there are passengers standing. Therefore, the crowding externality raises the marginal social cost of travelling, thus increasing the optimal bus fare, which is obtained as the difference between total marginal cost and average users cost on first best pricing (Section 2.2).

It is usually proposed in the literature that when users' waiting time cost is included in the total cost function of public transport services, the marginal cost pricing rule does not cover operator cost due to the positive effect of increasing frequency in reducing waiting time for users (Mohring, 1972; Turvey and Mohring, 1975; Jansson, 1979). This is a common result obtained from a number bus pricing and optimisation studies along the lines of Mohring (1972)'s square root formula (expression 2.2), which states that an increase in demand is met by a less than proportional increase in frequency. Therefore, as demand grows there is an increase in the occupancy rate or load factor inside vehicles, that is, an increase in crowding levels. Consequently, it is reasonable to analyse what would happen if a crowding disutility is considered in the frequency and fare optimisation problem.

The first answer to this problem is provided by Kraus (1991), who considers the standing externality that long-distance passengers who are able to find an empty seat, impose upon short-distance passengers that have to stand if all seats are taken by long-distance

passengers. Kraus (1991) assumes that the value of in-vehicle time savings ( $P_v$ ) is higher for standees than for passengers seating due to the discomfort caused by standing, which is shown to increase the optimal fare for long-distance travellers relative to short-distance travellers. The effect of a crowding externality on optimal bus supply is later analysed by Jara-Díaz and Gschwender (2003), who by including that  $P_v$  is a linear function of the average bus occupancy rate, demonstrate that the optimal frequency is higher than in the case in which there is no crowding externality reflected on  $P_v$ .

The inclusion of crowding externality in public transport optimisation models has also been shown to challenge frequency-related total cost savings (scale economies). When the disutility of crowding is accounted for as increasing  $P_v$ , average total cost could pass from a decreasing function of demand for low to middle demand levels, to an increasing function of demand for middle to high demand levels, as shown by Tirachini *et al.* (2010a) with a frequency optimisation model on a single public transport route. This result is due to the increase in crowding level when demand rises, which (in a model that takes crowding into account) is translated into an increase of users in-vehicle time cost. However, the result of a crowding-induced increasing total cost for a single route vanishes if the number of routes is also an optimisation variable, in which case route density is adjusted to keep total costs down (Tirachini *et al.*, 2010b).

In summary, the acknowledgement of a crowding externality on the valuation of travel time and on travel time itself might have sizeable impacts on the design of a public transport system, particularly in terms of the capacity provided to serve demand. When the crowding cost is ignored, policy makers may choose to provide a transport capacity that is just enough to meet demand, in which buses would be full (or close to full if a safety level of spare capacity is defined by design) in the most loaded sections of a route. Nevertheless, when the crowding cost is considered in the design stage of a route, it should be optimal to provide a greater service frequency and bus capacity in order to reduce the occupancy levels inside vehicles, and consequently improve the quality of travelling (Jara-Díaz and Gschwender, 2003). This issue will be revisited in the next section with new crowding cost functions that account separately for the proportion of

users seated and the density of standees inside public transport vehicles. This approach, together with the consideration of bus congestion and the election of a boarding and alighting technique and fare payment method, will be used in Chapter 8 for the optimisation of public transport services.

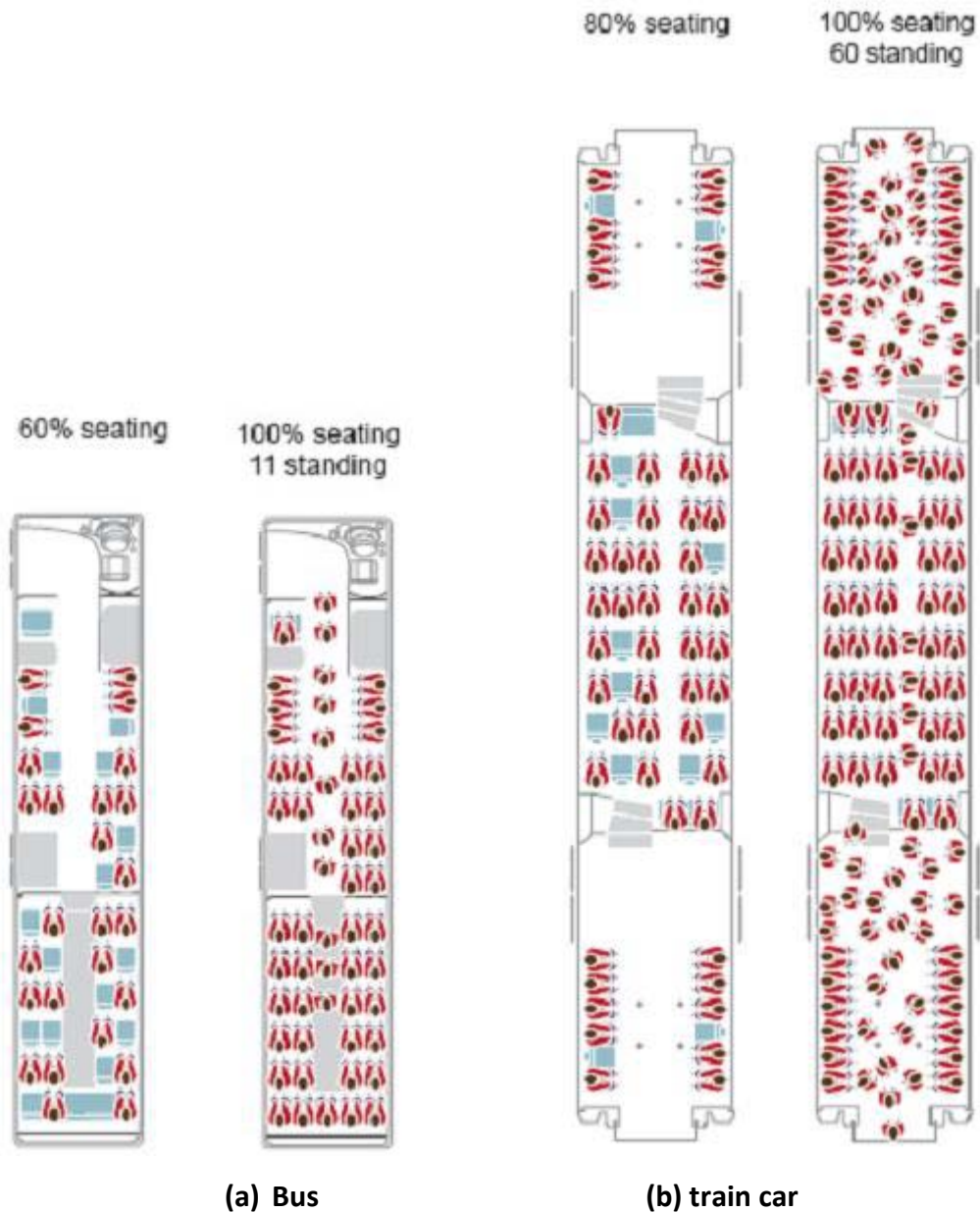
### **6.3 Estimation of Crowding and Standing Costs**

Section 6.2 discussed several dimensions of the influence of having a large number of users inside public transport vehicles and stations, which may be generically referred to as crowding cost or disutility. In particular, Section 6.2.4 reviewed previous studies that estimate crowding and standing costs as increasing the valuation of travel time savings. In this section we estimate mode choice models that include the proportion of available seats and the density of standees as attributes. The database used for the estimation of choice models is part of a feasibility study for a new metro system proposed for Sydney, conducted in 2009 at the Institute of Transport and Logistics Studies, The University of Sydney<sup>38</sup>. The modes included are car, bus, train and metro. In the stated choice experiment, respondents compare the levels of access and in-vehicle times, frequency, proportion of users seating and number of users standing, and costs (e.g., public transport fare, running cost and parking fee for cars). The experiment design, study area, sample size and socioeconomic characteristics of respondents are described at length in Hensher *et al.* (2011). Crowding levels on bus, train and metro were represented with diagrams, two examples of different levels of bus and train crowding are shown in Figure 6.3.

---

<sup>38</sup> The original study is not part of this thesis. The author thanks David Hensher and John Rose for providing access to their dataset for the development of this work.





**Figure 6.3: Different levels of occupancy in stated preference study**  
(Source: stated choice model described in Hensher *et al.*, 2011)

Hensher *et al.* (2011) estimate the crowding disutility as a function of the proportion of users seating (which affects the probability of getting a seat), and the number of users standing, in order to estimate the willingness to pay to get a seat as a function of the number of people seating and standing. In this work we use the density of standees per square metre -instead of the number of standees- to represent the disutility of crowding and standing, in order to have a common base among the three public transport modes considered, which have different sizes and proportion of area for seating and standing

(For example, in Figure 6.3 the train has proportionally more space allocated to standing than the bus).

Let  $U_m$  be the utility of mode  $m$ . In order to compare values of travel time savings and crowding multipliers, we propose three different models that incorporate attributes representing the number of passengers seating and standing, interacting with travel time; these models will be compared with a specification that ignores any crowding or standing cost. The models, named M1 to M4, are described as follows:

- M1: No crowding cost (eq. 6.1).
- M2: Only the density of standees [pax/m<sup>2</sup>] imposes an extra discomfort cost (eq. 6.2).
- M3: The density of standees and the proportion of seats occupied are sources of disutility (eq. 6.3).
- M4: The density of standees and the proportion of seats occupied are squared in the utility function (eq. 6.4).

$$U_m = \alpha_m^{M1} + \beta_a^{M1} t_{am} + \beta_h^{M1} h_m + \beta_{vm}^{M1} t_{vm} + \beta_e^{M1} t_{em} + \beta_c^{M1} \bar{c}_m \quad (6.1)$$

$$U_m = \alpha_m^{M2} + \beta_a^{M2} t_{am} + \beta_h^{M2} h_m + \beta_{vm}^{M2} t_{vm} + \beta_e^{M2} t_{em} + \beta_c^{M2} \bar{c}_m + \beta_{den}^{M2} n_{den} t_{vm} \quad (6.2)$$

$$U_m = \alpha_m^{M3} + \beta_a^{M3} t_{am} + \beta_h^{M3} h_m + \beta_{vm}^{M3} t_{vm}^{ij} + \beta_e^{M3} t_{em} + \beta_c^{M3} \bar{c}_m + \beta_{den}^{M3} n_{den} t_{vm} + \beta_{seat} p_{seat} t_{vm} \quad (6.3)$$

$$U_m = \alpha_m^{M4} + \beta_a^{M4} t_{am} + \beta_h^{M4} h_m + \beta_{vm}^{M4} t_{vm} + \beta_e^{M4} t_{em} + \beta_c^{M4} \bar{c}_m + \beta_{den2} n_{den}^2 t_{vm} + \beta_{seat2} p_{seat}^2 t_{vm} \quad (6.4)$$

In (6.1) to (6.4),  $t_{am}$  and  $t_{em}$  are the access and egress time, respectively,  $h_m$  is the headway between two consecutive vehicles (representing a proxy of the waiting time cost or scheduling delay),  $t_{vm}$  is the in-vehicle time,  $\bar{c}_m$  the money cost or fare,  $n_{den}$  the density of standees per square metre,  $p_{seat}$  the proportion of seats been used,  $\alpha_m$  is an alternative specific constant (ASC) and  $\beta_k$  are the parameters associated with the different attributes.

Multinomial logit (MNL) models are estimated in order for the parameters to be used as input into the social welfare maximisation model presented Chapter 8, which assumes a

log-sum function for the calculation of the users benefit. The estimation of parameters for commuting and specification tests are presented in Table 6.1 ( $n=1932$  observations):

**Table 6.1: Estimation of parameters, MNL models**

Parameter	M1	M2	M3	M4
Access time $\beta_a$	-0.016 (-1.22)	-0.017 (-1.33)	-0.017 (-1.33)	-0.017 (-1.33)
Headway $\beta_h$	-0.0088 (-2.71)	-0.010 (-3.06)	-0.010 (-3.07)	-0.010 (-3.06)
Travel time public transport ( $t_{vm}$ ) $\beta_v$	-0.019 (-5.09)	-0.013 (-3.45)	-0.004 (-0.58)	-0.006 (-1.25)
Egress time $\beta_e$	-0.055 (-4.31)	-0.058 (-4.54)	-0.059 (-4.59)	-0.059 (-4.61)
Travel time car ( $t_{vm}$ ) $\beta_v$	-0.016 (-3.11)	-0.018 (-3.41)	-0.018 (-3.37)	-0.018 (-3.37)
Cost $\beta_c$	-0.062 (-5.50)	-0.064 (-5.63)	-0.064 (-5.62)	-0.064 (-5.63)
ASC train $\alpha_t$	-3.393 (-5.16)	-3.455 (-5.24)	-3.473 (-5.26)	-3.476 (-5.27)
ASC bus $\alpha_b$	-4.131 (-5.67)	-4.275 (-5.82)	-4.313 (-5.86)	-4.315 (-5.86)
ASC metro $\alpha_m$	-2.526 (-4.30)	-2.444 (-4.14)	-2.465 (-4.17)	-2.460 (-4.16)
$t_{vm} \times$ den stand $\beta_{den}$		-0.004 (-4.48)	-0.003 (-2.82)	
$t_{vm} \times$ prop seat $\beta_{seat}$			-0.013 (-1.70)	
$t_{vm} \times$ (den stand) <sup>2</sup> $\beta_{den2}$				-0.0005 (-2.41)
$t_{vm} \times$ (prop seat) <sup>2</sup> $\beta_{seat2}$				-0.013 (-2.46)
Specification tests				
Log-likelihood	-1283.4	-1273.1	-1271.7	-1271.7
Adjusted $\rho^2$ (relative to ASCs)	0.107	0.113	0.113	0.113
Likelihood ratio test with respect to M1		20.53 ( $> \chi_{1,0.001}$ =10.83)	23.39 ( $> \chi_{2,0.001}$ =13.82)	23.34 ( $> \chi_{2,0.001}$ =13.82)
Likelihood ratio test with respect to M2			2.86 ( $< \chi_{1,0.05}$ =3.84)	

Note:  $t$ -ratio in bracket below parameter estimates. Time in minutes, cost in \$ (AUD).

Focusing on the goodness-of-fit measures, the log-likelihood and adjusted  $\rho^2$  statistics relative to a model with alternative specific constants (ASCs) only, demonstrate that the three crowding models (M2-M4) outperform the model with no crowding (M1), but the difference in overall fitness amongst the crowding models is not significant. In fact, M2, M3 and M4 have the same adjusted  $\rho^2$  value, and a likelihood ratio test indicates that M2, M3 and M4 are significantly superior than M1 at the 99.9 percent confidence level, however M2 and M3 are not statistically different at 95 percent confidence level<sup>39</sup>. Therefore, if we use the density of standing to characterise crowding costs, the inclusion of the availability of seats as a variable that influences modal choice is not statistically relevant, nevertheless from a behavioural perspective, the alternative crowding cost specifications do provide differences on the estimation of value of travel time savings.

Figure 6.4 shows the crowding multiplier (mark-up on the VTTs induced by crowding conditions, compared against uncrowded travel conditions) for increasing levels of occupancy of buses. The bus configuration of Figure 6.3a is used, in which there are 44 seats and a maximum of 27 standees, which corresponds to 4.4 pax/m<sup>2</sup> at crush capacity. Occupancy rate is measured against the seating capacity. In model M2, the only cause of discomfort is the density of standees, in which case the crowding multiplier grows only when the occupancy rate is over 100 percent, until reaching a value of 2.2 at crush capacity (higher than the values for seating and standing obtained by Whelan and Crocket, 2009 for rail in Britain). On the other hand, the models that are sensitive to the availability of seats (M3 and M4) present considerable differences in the VTTs relative to the number of passengers seating and standing, with crowding multipliers up to 4.6 (M4) and 7.3 (M3) at crush capacity.

The estimated value of travel time savings is graphically shown in Figure 6.5, which shows that models M3 and M4 are very sensitive to the availability of seats for the estimation of in-vehicle time savings. Importantly, not accounting for crowding differences (M1) in the valuation of time savings would imply an overestimation of the value of in-vehicle time savings for low occupancy levels and an underestimation for high occupancy rates, with a

---

<sup>39</sup> Models M2 and M4 cannot be compared with a likelihood ratio test because they are not nested.

threshold between one and two standees per square metre, depending on the specification of the crowding costs (M2, M3 and M4). Chapter 8 explores the implications of explicitly accounting for crowding in the determination of the optimal number of seats inside buses.

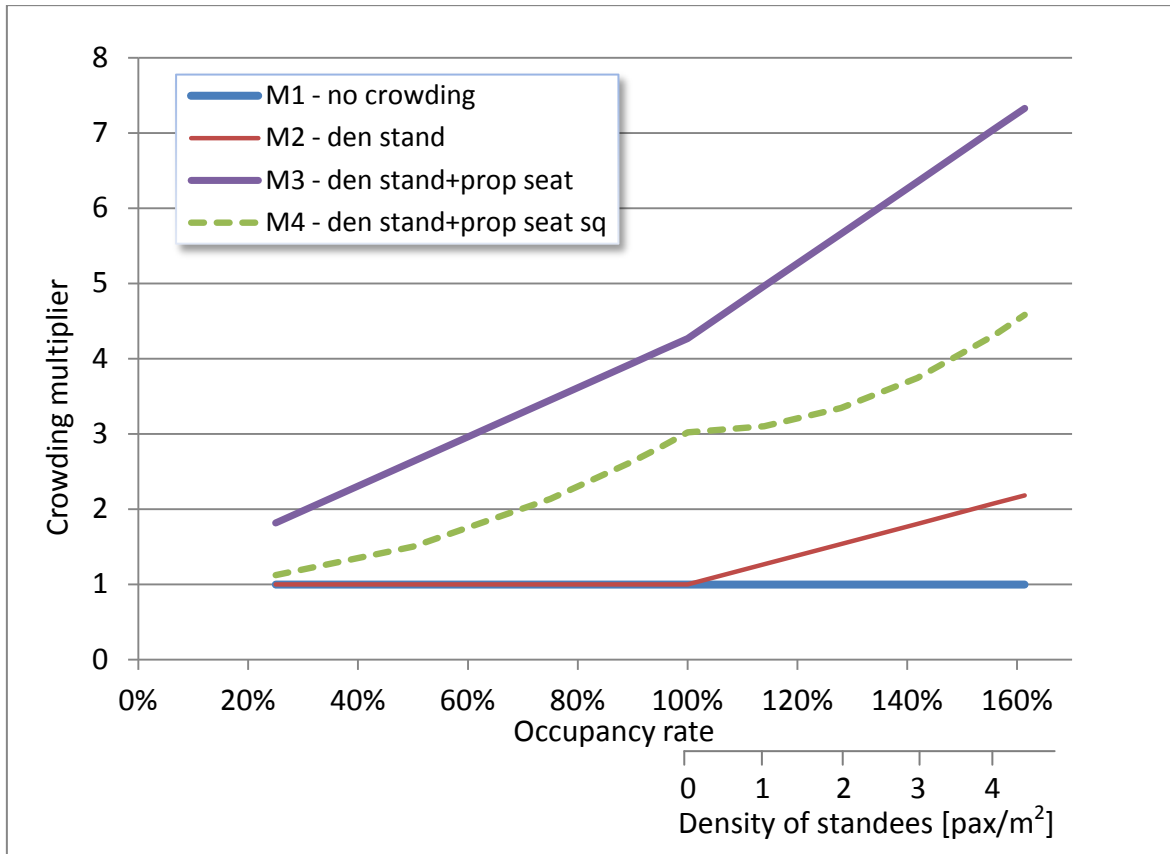


Figure 6.4: Crowding multiplier as a function of occupancy rate and density of standees

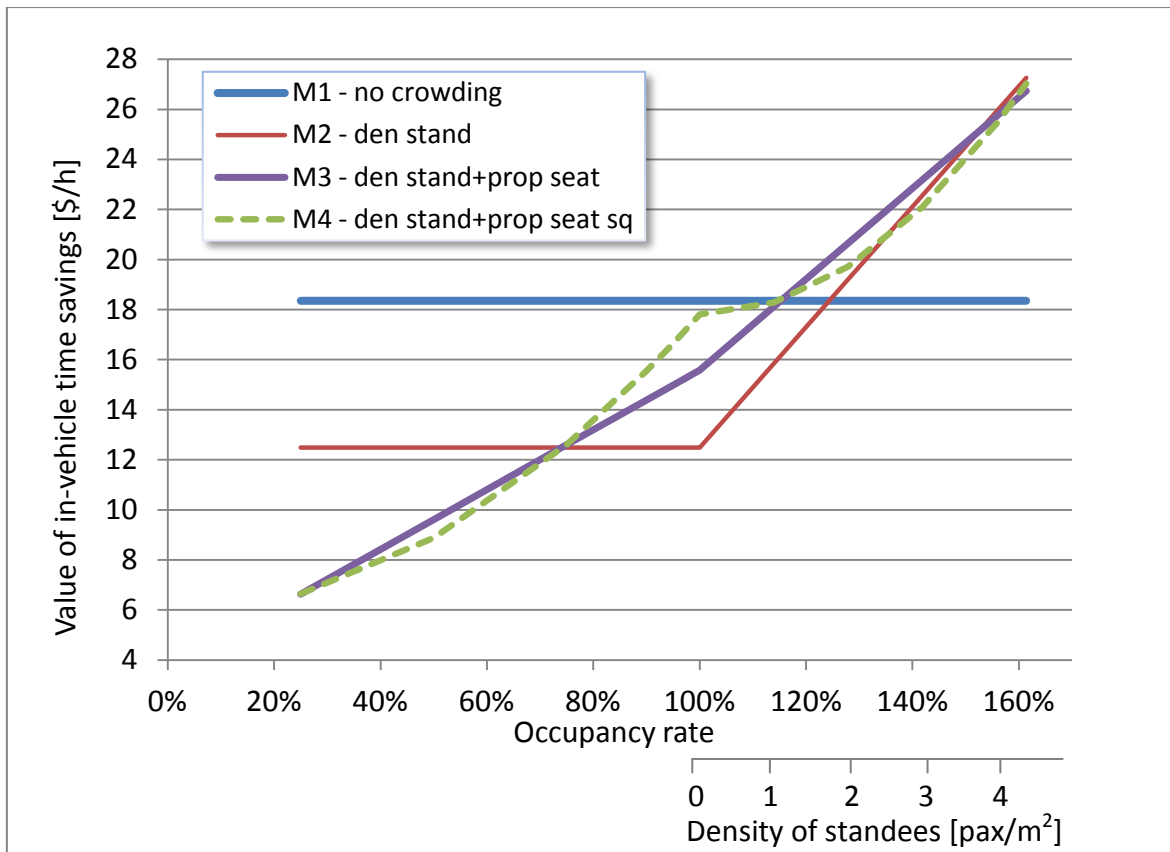


Figure 6.5: Value of in-vehicle time savings

#### 6.4 Conclusions

As the income of a population increases, improving the quality of public transport services may become more important in attracting passengers to public transport - relative to the valuation of travel time savings- which suggests that attributes like crowding, reliability and security will be increasingly relevant for public transport policy over time, in both developing and developed economies.

This chapter has provided a comprehensive review of the multiple effects that the crowding of passengers in public transport systems has on the quality and comfort of travelling, waiting and riding times, travel time variability and the determination of the service frequency, size of vehicles and optimal fare. Using data from Sydney we have estimated crowding cost functions that depend on the availability of seats and the density of standees per square metre, which shows the dependence of the valuation of travel time savings on the level of crowding inside vehicles. It is expected that the

specification of the crowding cost does have an influence in the optimal design of a bus system, including frequency, bus size and number of seats, an issue that is analysed in detail in Chapter 8.

## Chapter 7

# Bus Congestion, Optimal Infrastructure Investment and the Choice of a Fare Collection System: an Extended Total Cost Minimisation Model<sup>40</sup>

### 7.1 Introduction

The most common modelling approach in the microeconomic literature on public transport operations is the minimisation of total cost, defined as the summation of operators and users cost (Mohring, 1972; Jansson, 1980; Chang and Schonfeld, 1991; Jara-Díaz and Gschwender, 2003a). A growing number of elements of transit service provision have been progressively built within the cost minimisation framework, in order to describe dimensions that matter to operators and users, including the search for optimality conditions for service frequency [veh/h], vehicle size [pax/veh] and distance between stops, among other variables, as reviewed in Section 2.4.

In this chapter we extend the existing literature by including two key decision variables that are increasingly available to public transport policy makers: the optimal choice of a fare collection system and boarding policy (introduced in Chapter 4), and the investment in dedicated infrastructure for buses, the latter related to the speed that decision makers

---

<sup>40</sup> This chapter is an extended version of Tirachini and Hensher (2011).



want buses to achieve. In addition, we assess the influence of congestion caused by the interaction of buses in the determination of the optimal design of a system, using the queuing delay function defined in Chapter 5 (equation 5.5).

As discussed in Chapter 4, the existing economic literature on bus transport considers boarding and alighting times as given, thus ignoring the current availability of several technological options for fare payment, with different levels of investment, complexity and efficiency in the transfer of passengers. We consider the fare payment system as a policy variable, with implications for travel times and operator cost. As done with the empirical estimation of benefits from upgrading the fare collection system (Section 4.5), the performance of six alternative payment methods and bus boarding policies is compared, namely:

- i. On-board cash payment, front door boarding  $TnB1$ , referred to as “cash  $TnB1$ ”
- ii. On-board magnetic strip verification, front door boarding  $TnB1$ , referred to as “magnetic strip  $TnB1$ ”
- iii. On-board contactless card verification, front door boarding  $TnB1$ , referred to as “contactless card  $TnB1$ ”
- iv. On-board magnetic strip verification, all doors boarding  $TnBn$ , referred to as “magnetic strip  $TnBn$ ”
- v. On-board contactless card verification, all doors boarding  $TnBn$ , referred to as “contactless card  $TnBn$ ”
- vi. Off-board contactless card verification, all doors boarding  $TnBn$ , referred to as “off-board  $TnBn$ ”

Together with the analysis of fare collection systems and bus boarding rules, we take a closer look at bus running speed, defined as the cruising speed that buses attempt to maintain in between two consecutive stops. All previous studies treat running speed as an exogenous parameter, given by the physical conditions and regulations (speed limits) of the bus routes under study; however, bus running speed can be a decision variable, if an investment in infrastructure, like upgrading or building new busways, is designed to have a positive impact on the running speed of buses. A linear relationship between

infrastructure cost per kilometre and running speed is used, based on a positive correlation between infrastructure investment and commercial speed (total speed including stops), empirically identified by comparing data from a number of Bus Rapid Transit (BRT) systems. We show that a target speed increases the investment in infrastructure but also reduces the travel time between stops, and hence a compromise running speed is selected as the optimal solution.

In the present chapter, we stay within the usual total cost minimisation approach with parametric demand to optimally choose frequency, bus capacity and station spacing as previously undertaken in the literature, and incorporate decisions on running speed and the fare payment system and bus boarding rule, under uncongested and congested bus operations. The circulation of buses in a dedicated corridor is modelled in terms of three components: links, bus stations and (traffic light) intersections. In Section 7.2 the time lost at each stage of a round-trip is derived, which is then used in Section 7.3 to find expressions for the cost to users and operators. Section 7.4 presents an in-depth analysis of the results and implications of different modelling assumptions. Section 7.5 summarises the main findings of the chapter.

## 7.2 Bus Round-trip Time

We consider a linear bi-directional corridor of length  $L$  and a single period of operation. The round-trip or cycle time,  $T_c$ , is defined as the total travel time during one cycle, given both service time and slack time at termini. Let  $T_r$  be the running or movement time along the route,  $T_i$  the delay at intersections (due to traffic lights),  $T_s$  the time lost at bus stops, and  $T_k$  the layover time at the end of the route; then the round-trip time is:

$$T_c = T_r + T_i + T_s + T_k \quad (7.1)$$

Buses circulate with no interaction with other modes on a dedicated road corridor, apart from the implicit delay due to traffic lights. The running time without any delay due to stopping is given as (7.2), where  $v_0$  is the constant running (cruising) speed.

$$T_r = \frac{2L}{v_0} \quad (7.2)$$

To model the delay in the process of decelerating to stop, and accelerating to start running again (either at intersections or bus stops), we assume uniform acceleration and deceleration; thus the extra stopping delay on top of the uniform travel time given by  $v_0$ , is expressed as (7.3)<sup>41</sup>.

$$t_l = \frac{v_0}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \quad (7.3)$$

$r_a$  and  $r_d$  are the acceleration and deceleration rates of the bus [m/s/s], respectively. The mean queuing delay at intersections is modelled as (7.4).

$$T_i = d_1 + d_2 + t_{ac} \bar{h}_i \quad (7.4)$$

In (7.4),  $d_1$  is the non-random delay due to signal cycle effects, calculated assuming an average non-random arrival rate (Akçelik, 1981; Akçelik and Rouphail, 1993),  $d_2$  is the overflow delay (including the effects of random arrivals and over-saturation),  $t_{ac}$  is the acceleration and deceleration delay given by (7.3), and  $\bar{h}_i$  is the average number of stops per vehicle. The final expression for (7.4) is given in (7.5) (see Appendix A3 for details):

$$T_i = \left[ \frac{0.5C_T(1-u)^2}{1-ux_b} + \frac{v_0}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \frac{1-u}{1-ux_b} \right] I \quad (7.5)$$

where  $C_T$  is the traffic light cycle time [s],  $u = g/C_T$  is the ratio of effective green time  $g$  [s] to the cycle time  $C_T$ ,  $x_b = f_b/K_I$  is the degree of saturation, given the capacity  $K_I$  of the intersection [veh/h],  $f_b$  is the bus frequency [veh/h], and  $I$  is the number of intersections along the route.

---

<sup>41</sup> From kinematics, the total deceleration time is  $v_0/r_d$  and the deceleration delay (on top of uniform speed movement) is  $v_0/2r_d$ . The acceleration delay is obtained analogously.

Next, we will analyse the delays caused by bus stops. Let  $t_s$  be the total delay time per bus stop, which consists of the acceleration and deceleration delay  $t_{ac}$  (equation 7.3), the average queuing time  $t_q$ , the dwell time  $t_d$ , and an internal waiting delay  $t_{iw}$ , i.e.

$$t_s = t_{ac} + t_q + t_d + t_{iw} \quad (7.6)$$

As discussed in Chapter 5, the queuing time  $t_q$  is a measure of the external congestion caused by a bus stop, observed when a bus arrives at a stop and all berths are occupied. This delay is commonly present in high frequency services, but it may also occur in poorly controlled low frequency services where buses tend to bunch. We use the functional function established in Chapter 5:

$$t_q = 0.001 \left[ \beta_0 + \beta_{11} s_b + (\beta_{d1} + \beta_{d2} Z_2 + \beta_{d3} Z_3) t_d \right] e^{0.001 f_b [\beta_f + \beta_{12} s_b + (\beta_{d4} + \beta_{d5} Z_2 + \beta_{d6} Z_3) t_d]} \quad (7.7)$$

which estimates queuing delay  $t_q$  [s/bus] as a function of the bus length  $s_b$  [m], dwell time  $t_d$  [s/bus], frequency  $f_b$  [veh/h] and number of berths per bus stop. Parameters  $\beta_0$ ,  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{d1}$ ,  $\beta_{d2}$ ,  $\beta_{d3}$ ,  $\beta_{d4}$ ,  $\beta_{d5}$ ,  $\beta_{d6}$  and  $\beta_f$  are estimated in Table 5.1, and factors 0.001 are introduced for scaling of the parameters.  $Z_2$  and  $Z_3$  are dummy variables defined as follows:

$$Z_2 = \begin{cases} 1 & \text{if bus stop has two berths} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_3 = \begin{cases} 1 & \text{if bus stop has three berths} \\ 0 & \text{otherwise} \end{cases}$$

When applying equation (7.7), it is relevant to recall that in real public transport corridors, bus queues develop only on high frequency services at a subset of bus stops, namely those stops with a high passenger boarding demand, and therefore, low capacity in terms of buses per hour they can handle. In this context, and observing the exponential nature of  $t_q$  in (7.7), in order to properly model the effects of congestion in bus operations, it is inappropriate to assume that demand is uniformly distributed along the corridor (i.e., assuming that the same number of passengers board and alight at every bus stop), in which case we would likely be underestimating bus stop delays. For example, based on the Sydney bus dwell time surveys described in Section 4.3, in most cases

between 30 and 50 percent of the passengers boarding buses are concentrated at 10 to 20 percent of the bus stops along a route. Accordingly, we separate bus stops into two groups: a percentage  $p_h$  of *high demand stops* (low capacity, possible queuing delay), at which most boardings (a percentage  $p_d$ ) are concentrated, and a percentage  $1 - p_h$  of *low demand stops* (high capacity, zero or little queuing time). We define ‘high demand stops’ as those with high boarding numbers because boarding is more time consuming than alighting; therefore, as observed in Sydney, long dwell times (prone to cause queuing delays) are triggered at stations with a large number of passengers boarding rather than at stations with a large number of passengers alighting.

Based on the Sydney evidence, we assume that  $p_d = 30$  percent of the total demand boarding buses in  $p_h = 10$  percent of the stops, and on those high demand stations on average the number of passengers alighting buses is  $p_c = 20$  percent of the number of passengers boarding. Hence the average number of passengers boarding ( $\lambda^+$ ) and alighting ( $\lambda^-$ ) a bus per bus stop in high ( $h$ ) and low ( $l$ ) demand stops by direction  $j$  ( $j \in \{1, 2\}$ ) is given in (7.8), where  $S$  is the number of bus stops along the corridor.

$$\lambda_{hj}^+ = \frac{p_d N_{bi}}{p_h S f_b} \quad (7.8a)$$

$$\lambda_{lj}^+ = \frac{(1 - p_d) N_{bi}}{(1 - p_h) S f_b} \quad (7.8b)$$

$$\lambda_{hj}^- = \frac{p_c p_d N_{bi}}{p_h S f_b} \quad (7.8c)$$

$$\lambda_{lj}^- = \frac{(1 - p_c p_d) N_{bi}}{(1 - p_h) S f_b} \quad (7.8d)$$

The estimation of the dwell time per stop requires the cases with boarding allowed at all doors ( $TnBn$ ) and at the front door only ( $TnB1$ ) to be addressed separately, since in  $TnBn$  boarding and alighting is sequential at all doors, whereas in  $TnB1$  boarding at the front door occurs simultaneously with alighting at the rear doors. As discussed in Section 4.3.6, where there are multiple doors to board and alight, passengers can choose a door to get

on and off buses, and the spatial dispersion of their decision determines how long the boarding and alighting process is going to last. It seems unreasonable to suppose that passengers will distribute uniformly across doors if middle or back doors have closer access to more seats than, say, the front door. We assume that the middle doors would attract a number of passengers that is 50 percent higher than that of the front or back doors. For example, for buses with two doors, the rear door is placed towards the centre of the bus, and is therefore assumed to attract 60 percent of the boarding demand, leaving 40 percent boarding through the front door, next to the driver. The same assumption is made regarding alighting. With this, the proportion of passengers assumed to get on and off buses at each door for the case  $TnBn$  is shown in Table 7.1, as a function of the number of doors per bus (door numbers are assigned from front to back, i.e., door 1 is the front door and door 2 is its closest door).

**Table 7.1: Proportion of passengers boarding and alighting at each door, as a function of the number of doors per bus, regime  $TnBn$**

Number of doors	Door 1	Door 2	Door 3	Door 4		
1	100%					
2	40%				60%	
3	29%				43%	28%
4	20%				30%	30%

On the other hand, for the cases with boarding at the front door only ( $TnB1$ ), we assume that passengers alighting spread out through the back doors with proportions given in Table 7.2 (also assuming that middle doors get 50 percent more demand than the back door).

**Table 7.2: Proportion of passengers alighting at each door, as a function of the number of doors per bus, regime *TnB1***

Number of doors	Door 1	Door 2	Door 3	Door 4
2	0%	<b>100%</b>		
3	0%	<b>60%</b>		
4	0%	<b>38%</b>	<b>38%</b>	24%

Thus, the dwell time is obtained as expression (7.9)

$$t_d^i = \begin{cases} c_{oc} + p_b \beta_b \lambda^+ + p_a \beta_a \lambda^- & \text{if boarding at all doors (TnBn)} \\ c_{oc} + \max \{ \beta_b \lambda^+, p_a \beta_a \lambda^- \} & \text{if boarding at front door only (TnB1)} \end{cases} \quad (7.9)$$

where the number of passengers boarding ( $\lambda^+$ ) and alighting ( $\lambda^-$ ) a bus are given by equations (7.8), and factors  $p_a$  and  $p_b$  are the proportion of passengers boarding and alighting at the busiest door, respectively, given by the bold figures in Tables 7.1 and 7.2. Factors  $p_a$  and  $p_b$  are summarised in Table 7.3.

**Table 7.3: proportion of passengers boarding and alighting at the busiest door**

Number of doors	<i>TnBn</i> $p_a (= p_b)$	<i>TnB1</i> $p_a$
1	100%	-
2	60%	100%
3	43%	60%
4	30%	38%

Equations (7.8) and (7.9) and Table 7.3 conclude the derivation of the dwell time per stop on the bus corridor.

The internal waiting delay  $t_{iw}$  occurs in stops with two or more berths, when one bus blocks the movement of the bus behind if the latter wants to leave the stop, or when there is a traffic light immediately after the stop. We will consider stops isolated from

traffic lights and with a second lane to overtake buses; therefore  $t_{iw}$  is nil (Valencia and Fernández, 2007).

In consequence, the total delay along the corridor due to bus stops is given as (7.10)

$$T_s = \left[ p_h (t_s^{1h} + t_s^{2h}) + (1 - p_h) (t_s^{1l} + t_s^{2l}) \right] S \quad (7.10)$$

where  $t_s^{jh}$  and  $t_s^{jl}$  are the delay for high demand and low demand stops in direction  $j$ , respectively.

The last component of the round-trip time (equation 7.1) is the layover time  $T_k$ . Bus travel time can vary significantly between trips and days, and layover time is usually introduced into the schedule to accommodate delays and to provide a break for drivers. As such, a scheduled layover time should be determined as a function of the travel time variability (Furth, 2000). We will include layover time as a constant and exogenously defined value (e.g., 5 minutes).

In summary, the round-trip time  $T_c$  given in (7.11), is the summation of expressions (7.2), (7.5), (7.10) and the layover time  $T_k$ .

$$T_c = \frac{2L}{v_0} + 2 \left[ \frac{0.5C_T(1-u)^2}{1-ux_b} + \frac{v_0}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \frac{1-u}{1-ux_b} \right] I + \left[ p_h (t_s^{1h} + t_s^{2h}) + (1 - p_h) (t_s^{1l} + t_s^{2l}) \right] S + T_k \quad (7.11)$$

### 7.3 User Cost, Operator Cost and Problem Formulation

User cost is divided into access, waiting and in-vehicle time costs. For the formulation of the access time cost, recall that the number of low and high demand stations,  $p_h S$  and  $(1 - p_h) S$  respectively, are also a variable of the problem. We assume the existence of two geographic areas, a high demand area (where all  $p_h S$  high demand stops are



located), and a low demand area (where all  $(1 - P_h)S$  low demand stops are located), and in each of these areas, demand is uniformly distributed. This implies that around a bus stop (either with low or high demand), passengers are homogeneously distributed. Therefore, if the bus stops are equally spaced, passengers have to walk on average  $L/4S$  at both the origin and the destination. Hence, the average total walking distance is  $L/2S$ , and the access time cost  $C_a$  is given as (7.12), where  $P_a$  is the value of access time savings [\$/h], and  $v_w$  is the walking speed [km/h].

$$C_a(S) = P_a \frac{L}{2v_w S} N_b \quad (7.12)$$

The waiting time cost  $C_w$  is linked to the bus frequency. We distinguish between cases with low and high frequency; when frequency is high, passengers usually arrive at the stations randomly at a constant rate, but when frequency is low, generally a timetable of services is provided, and most of the users arrive at stations following the schedule, in order to reduce their waiting time. We assume that for frequencies greater than 5 veh/h, equivalent to an average headway up to 12 minutes, users arrive randomly at stations. Following Tirachini *et al.* (2010a), the two cases can be formulated as the single expression (7.13).

$$C_w(f_b) = P_w \left( t_0 + \frac{t_1}{2f_b} \right) N_b \quad (7.13)$$

with

$$t_0 = \begin{cases} 0 & \text{if } f_b \geq 5 \text{ veh/h} \\ t_w & \text{if } 0 < f_b < 5 \text{ veh/h} \end{cases} \quad t_1 = \begin{cases} 1 & \text{if } f_b \geq 5 \text{ veh/h} \\ \mu & \text{if } 0 < f_b < 5 \text{ veh/h} \end{cases}$$

$P_w$  is the value of waiting time savings [\$/h]. For the low frequency case ( $0 < f_b < 5 \text{ veh/h}$ ),  $t_w$  is a fixed 'safety threshold' time that passengers spend waiting at stations before the expected arrival of the next vehicle, and  $\mu = P_h/P_w$  is the ratio of the value of *home waiting time* savings  $P_h$  to the value of *station waiting time* savings  $P_w$  (for

example,  $\mu = 0.33$ )<sup>42</sup>. Implicit in (7.13) is that the capacity constraint of the vehicles is not binding; in fact the frequency will be set to avoid overloading of vehicles (as will be seen in expression 7.24a).

In-vehicle time is modelled as a fraction  $l_i/L$  of the total travel time  $T_{ti}$  in direction  $i$  [h], where  $l_i$  [km] is the average trip length in direction  $i$  and  $L$  [km] is the route length. Then, if  $P_v$  is the value of in-vehicle time savings [\$/h], the in-vehicle time cost  $C_v$  of users is given as (7.14)<sup>43</sup>

$$C_v = P_v \left( \frac{l_1}{L} T_{t1} N_{b1} + \frac{l_2}{L} T_{t2} N_{b2} \right) \quad (7.14)$$

The travel times in each direction are given as equation (7.15), derived in an analogous way to the round-trip time (equation 7.11).

$$T_{t1} = \frac{L}{v_0} + \left[ \frac{0.5C_T(1-u)^2}{1-ux_b} + \frac{v_0}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \frac{1-u}{1-ux_b} \right] I + [p_h t_s^{1h} + (1-p_h) t_s^{1l}] S \quad (7.15a)$$

$$T_{t2} = \frac{L}{v_0} + \left[ \frac{0.5C_T(1-u)^2}{1-ux_b} + \frac{v_0}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \frac{1-u}{1-ux_b} \right] I + [p_h t_s^{2h} + (1-p_h) t_s^{2l}] S \quad (7.15b)$$

Operator cost is divided into five components:

$c_1$ : Busway infrastructure and land costs [\$/km-h]

$c_2$ : Station infrastructure cost [\$/station-h]

$c_3$ : Personnel costs (crew) and vehicle capital costs [\$/bus-h]

$c_4$ : Running costs (fuel consumption, lubricants, tyres, maintenance, etc.) [\$/bus-km]

<sup>42</sup> We assume that when users arrive at stations following a timetable, there is a waiting time cost outside stations because departures are not at the time desired by users (called "schedule delay"). As this *passive waiting time* can be spent at home or another place where passengers can assign their time to a more productive use or leisure, the opportunity cost or value of *passive waiting time* savings,  $P_h$ , is lower than the value of station or *active waiting time* savings,  $P_w$ .

<sup>43</sup> In this formulation, the influence of crowding on increasing both travel time and its valuation is ignored, because only total demand and average bus load are assumed to be known. The effect of crowding on the optimal design of bus systems is analysed in Chapter 8, with a multimodal pricing model in which the number of passengers seating and standing is known stop by stop.

$c_5$ : Implementation cost related with the fare payment technology (e.g., software requirements) [\$/h]

First, we assume a positive relationship between the investment in infrastructure for buses and the speed buses achieve, specifically, that the bus running speed is to some extent a function of the land cost and infrastructure investment (for example, buses run faster if dedicated busways are built). That is,  $v_0 = v_0(c_1)$ , which has to be inverted for the estimation of  $c_1$  as a function of variable  $v_0$ , i.e.,  $c_1 = c_1(v_0)$ , the latter interpreted as the necessary investment  $c_1$  in order to achieve a target running speed  $v_0$ . We further assume that  $c_1(v_0)$  is a linear function, as justified in Appendix A4.

$$c_1(v_0) = c_{10} + c_{11}v_0 \quad (7.16)$$

Secondly, the station cost  $c_2$  (equation 7.17), consists of two components: the station infrastructure cost which depends on the bus length  $s_b$  i.e.,  $c_{20}(s_b)$ , and the cost of fare vending machines and fare collection readers (if validation is undertaken at the station and not on bus),  $c_{21}(\Delta)$ , where the dependency on  $\Delta$  denotes the fare payment method.

$$c_2(s_b, \Delta) = c_{20}(s_b) + c_{21}(\Delta) \quad (7.17)$$

Thirdly, the cost per bus-hour  $c_3$  also has two elements: the personnel cost (wages) and the capital cost of a vehicle, which includes the cost of the fare collection readers (validation devices) installed in buses. As in Chapter 5, we consider four commercial sizes for buses: mini (8 m), standard (12 m), rigid long (15 m) and articulated (18 m). Let  $c_{30}(s_b)$  be the cost associated with bus size and driving wages, and  $c_{31}(\Delta)$  the cost of the fare collection readers, then, the total cost per bus-hour  $c_3$  is expressed as (7.18).

$$c_3(s_b, \Delta) = c_{30}(s_b) + c_{31}(\Delta) \quad (7.18)$$

Finally, the fourth component of operator cost is the running cost per vehicle-kilometre  $c_4$ , which could include fuel consumption, lubricants, tyres, maintenance, etc. We assume that the running cost function depends on bus size and running speed, and is

estimated using data on fuel consumption, bus size and average speed of several bus operators in New South Wales collected by Hensher (2003) as follows:

$$c_4(s_b, v_0) = c_{40} + c_{41}s_b + c_{42}v_0 \quad (7.19)$$

Finally,  $c_5(\Delta)$  accounts for the cost of software and implementation of the alternative fare collection technologies and boarding and alighting policies. The estimation of the parameters for equations (7.16) to (7.19) is given in the Appendix A4. After deriving expressions (7.16) to (7.19), we can define the total operator cost  $C_o$  as formula (7.20).

$$C_o = c_1(v_0)L + c_2(s_b, \Delta)S + c_3(s_b, \Delta)F + c_4(s_b, v_0)VF + c_5(\Delta) \quad (7.20)$$

where  $F$  is the fleet size requirement and  $V$  is the commercial speed (operating speed including movement and stops). The fleet size requirement is given in (7.21) as the product of the frequency  $f_b$  and the round-trip time  $T_c$  (the latter given as equation 7.11).

$$F = f_b T_c \quad (7.21)$$

Rewriting  $T_c$  as  $2L/V$  and introducing this into (7.21), we see that the fourth term in (7.20) does not depend on the commercial speed and passenger demand. Thus, the final expression for operator cost is given by (7.22).

$$C_o(f_b, s_b, S, v_0, \Delta) = c_1(v_0)L + c_2(s_b, \Delta)S + c_3(s_b, \Delta)f_b T_c(f_b, s_b, S, v_0, \Delta) + 2c_4(s_b, v_0)Lf_b + c_5(\Delta) \quad (7.22)$$

Finally, after deriving expressions for operator and user costs, the total cost minimisation problem is formulated as:

$$\text{Min } C_t(f_b, s_b, S, v_0, \Delta) = C_o(f_b, s_b, S, v_0, \Delta) + C_a(S) + C_w(f_b) + C_v(f_b, s_b, S, v_0, \Delta) \quad (7.23)$$

Subject to

$$\alpha_{max} N_b \leq \kappa f_b K_b(s_b) \quad (7.24a)$$

$$f_{min} \leq f_b \leq f_{max} \quad (7.24b)$$

$$v_{min} \leq v_0 \leq v_{max} \quad (7.24c)$$

$$s_b \in \{s_{b1}, \dots, s_{b4}\} \quad (7.24d)$$

$$\Delta \in \{\Delta_1, \dots, \Delta_6\} \quad (7.24e)$$

In (7.23), total cost  $C_t$  is the sum of operator and user costs (equations 7.12, 7.13, 7.14 and 7.22). Frequency  $f_b$  and running speed  $v_0$  are assumed to be continuous variables, the number of stops  $S$  is an integer (but for the solution will be considered continuous), there are four alternatives for bus size (expression 7.24d) and the boarding and alighting policy and fare collection technology  $\Delta$  belongs to one of the six alternative systems defined in Section 7.1 (expression 7.24e). Formally, we are minimising the total cost associated with the designated bus system, evaluated at market prices. There is no consideration of deviations between market and shadow prices, nor of the fact that part of the operator costs may be shared between several routes in a network and two or more levels of government.

As for the constraints, inequality (7.24a) is a capacity constraint, where  $\alpha_{max}$  is the fraction of passengers that traverse the most loaded section of the line,  $K_b(s_b)$  is the capacity of a bus given its length, and  $\kappa$  is introduced to have spare capacity to absorb random variations in demand (for example,  $\kappa = 0.9$ ). Thus, (7.24a) states that the passenger capacity of the system (the product of bus capacity and frequency) must be sufficient to cover demand in the most loaded section of the route. Next, for (7.24b), frequencies are also constrained by a minimum policy frequency  $f_{min}$  (set to have a minimum level of service, if desired) and the maximum feasible frequency  $f_{max}$  which will be given by the capacity constraints of the corridor. Expression (7.24c) establishes minimum and maximum values for the running speed. The system (7.23)-(7.24) is solved for the six different payment methods and results are then compared. The constrained optimisation problem is solved using the optimisation toolbox of MATLAB.

## 7.4 Results and Analysis

### 7.4.1 Assumptions

We assume a corridor of length  $L = 20$  km. The minimum and maximum bus running speeds are  $v_{min} = 20$  km/h and  $v_{max} = 80$  km/h. For the acceleration and deceleration rates, a standard value is  $r_a = r_d = 1.2$  m/s<sup>2</sup> (TRB, 2000). Traffic light intersections are every 800 m, which implies that the number of traffic lights is  $I = L/0.8 = 25$ . We further assume for all traffic lights that the cycle time is  $C_T = 120$  seconds, and the ratio of effective green time is  $u = 0.6$ . For the relative distribution of demand along the line, based on the Sydney evidence, we assume that  $p_d = 30$  percent of the total demand boards buses in  $p_h = 10$  percent of the stops, and on those high demand stations on average the number of passengers alighting buses is  $p_c = 20$  percent of the number of passengers boarding (on both directions), as explained in Section 7.2. For the user cost functions, the values of travel time savings are  $P_a = 15.5$  \$/h (access),  $P_w = 17.0$  \$/h (waiting), and  $P_v = 18.4$  \$/h (in-vehicle), calculated as the ratio of the respective time and cost parameters from M1 (the model without crowding attributes) in Chapter 6, Table 6.1, i.e.,  $P_a = \beta_a^{M1} / \beta_c^{M1}$ ,  $P_w = 2\beta_h^{M1} / \beta_c^{M1}$  and  $P_v = \beta_{vb}^{M1} / \beta_c^{M1}$  respectively<sup>44</sup>. The average trip length is the same in both directions,  $l_1 = l_2 = 10$  km, and the walking speed is  $v_w = 4$  km/h. Total directional demand is the same in both directions, i.e.,  $N_{b1} = N_{b2} \equiv N$ . We assume that 8 metre long buses have two doors, 12 metre long buses have three doors and 15 and 18 metre long buses have four doors. Bus stops have two berths.

### 7.4.2 Results

The most notable results are reported in Table 7.4 and Figures 7.1 to 7.7. On the x-axis is the one-way demand  $N$ . First, in terms of total average cost (users plus operators, Figure 7.1), we observe that the system with on-board cash payment is by far the least efficient

---

<sup>44</sup>  $P_w$  assumes that the waiting time parameter is twice the headway parameter, which comes from assuming that average waiting time is half of the headway. In our results frequency is always higher than 10 veh/h, therefore this assumption is in line with the waiting time specification (equation 7.13) for high frequency operation. We ignore the egress time parameter from Table 6.1 because it yields an unrealistic value of egress time savings, 53.2 \$/h. Throughout the chapter, we use the following dollar notation, \$: Australian Dollar, US\$: U.S. Dollar.

and cannot handle more than 10,000 pax/h, at which demand level the average queuing delay on high demand stops is 35 seconds (Figure 7.2) because the large boarding time associated with cash payment increases dwell time, frequency, and consequently, queuing delay (eq. 7.7). Second, the systems with boarding at all doors (*TnBn*) provide the lowest total cost across the demand range under study; Figure 7.1 shows that the cost of the three *TnBn* systems is virtually the same and clearly lower than the scenarios with boarding at the front door only (*TnB1*). The system with off board fare collection and verification is the most cost effective for a demand greater than 600 pax/h, and only for a low demand of 500 pax/h an on-board verification system with a magnetic strip would be more effective (although up to 4,000 pax/h the difference in total cost between the three *TnBn* systems is lower than 1 percent, and the highest demand of 15,000 pax/h the difference is up to 2 percent). Therefore, we can conclude that once on-board cash payment has been ruled out, allowing boarding at all doors is more important as a tool to reduce both users' and operator's costs than technological improvements on fare collection.

Third, total average cost for the *TnBn* alternatives is always decreasing (economies of scale) in contrast to the increasing total average cost that is observed for higher demands when boarding is allowed at the front door only. For example, average cost increases beyond a demand of 4,000 pax/h, which is a result of the increasing bus stop congestion provoked by slow boarding systems (Figure 7.2).

Fourth, Figure 7.1 also shows that the *TnBn* systems have economies of scale along the whole demand range tested, whereas the *TnB1* boarding methods have diseconomies of scale from some point (3,500; 8,000 and 10,500 pax/h for cash, magnetic strip and contactless card, respectively), which is directly related to the queuing delays resulting from having high frequencies for high demands, as depicted in Figure 7.2. The slowest method (cash payment) would yield delays of up to 36 s/bus for 10,000 pax/h, whereas magnetic strip and contactless card reach delays of 34 and 25 s/bus at 15,000 pax/h, respectively. This increase in travel time due to higher demand is what pushes the average total cost up. On the other hand, for the *TnBn* alternatives, average queuing delay is under 4 seconds for the whole demand range.

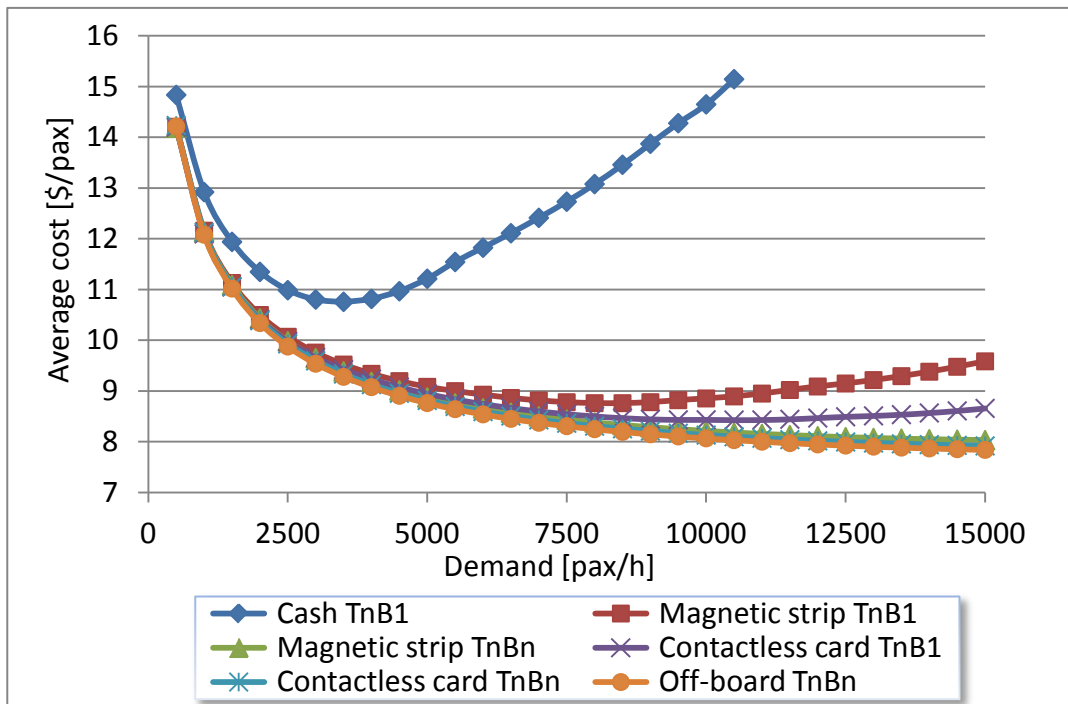


Figure 7.1: Total average cost

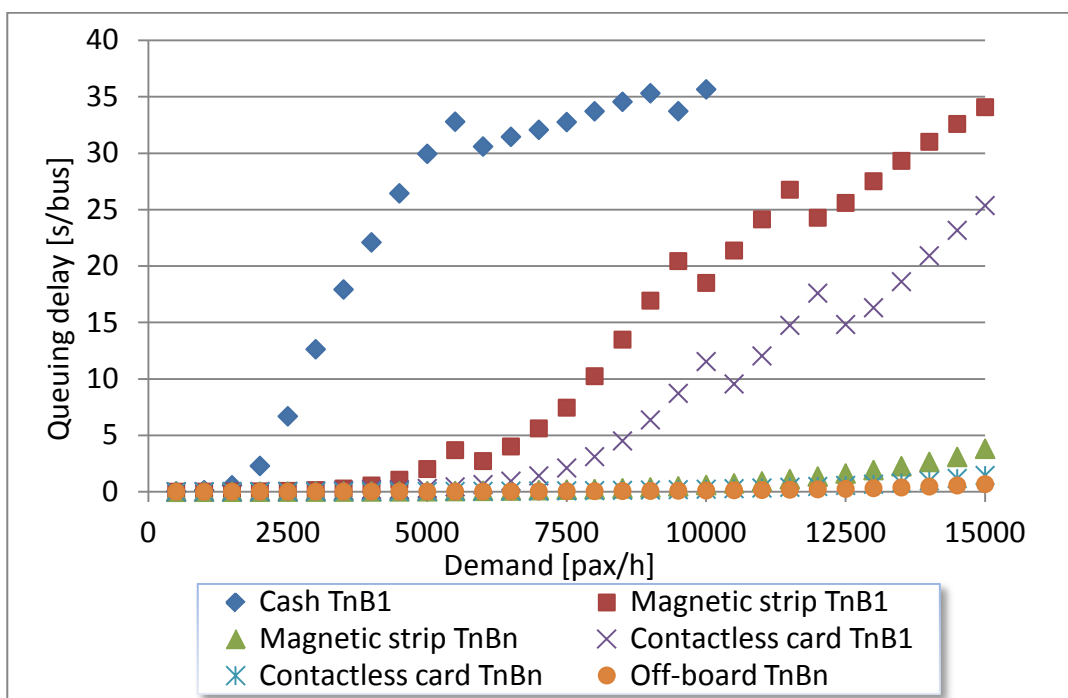


Figure 7.2: Average queuing delay

Table 7.4 presents the optimal bus size for different demand levels. For every fare collection system and boarding policy, bus size grows with demand as previously found by Jara-Díaz and Gschwender (2003a); the novelty of Table 7.4 is that reducing boarding

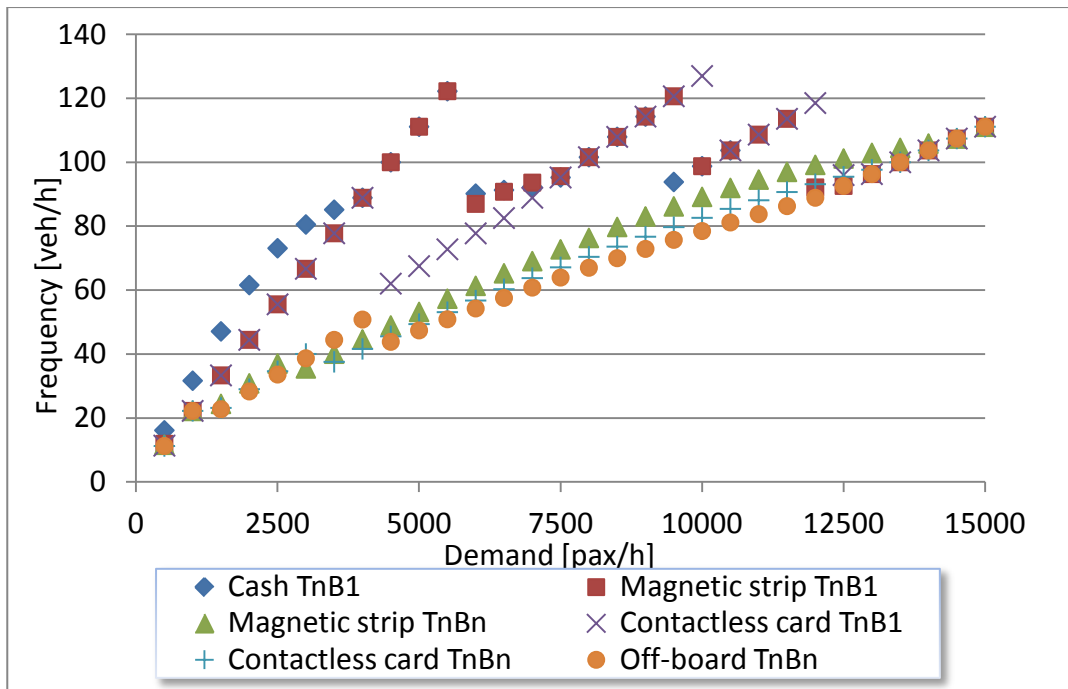


and alighting time ( $TnBn$ ) may result in the use of bigger buses for a smaller demand level than the cases with front door boarding<sup>45</sup>. The use of bigger buses when the boarding and alighting process is quicker, is matched by a lower bus frequency (Figure 7.3); this is because one of the objectives of bus frequency in this model is to reduce dwell time by reducing the number of passengers getting on and off one bus; if the time to board and alight is short, it is not necessary to provide a frequency as high as when boarding and alighting are slow.

**Table 7.4: Optimal bus size [m]**

<b>Demand [pax/h]</b>	<b>Cash <math>TnB1</math></b>	<b>Magnetic strip <math>TnB1</math></b>	<b>Magnetic strip <math>TnBn</math></b>	<b>Contactless card <math>TnB1</math></b>	<b>Contactless card <math>TnBn</math></b>	<b>Off-board <math>TnBn</math></b>
1,000	8	8	8	8	8	8
2,000	8	8	12	8	12	12
3,000	8	8	12	8	12	12
4,000	8	8	18	8	18	18
5,000	8	8	18	12	18	18
6,000	12	12	18	12	18	18
7,000	12	12	18	12	18	18
8,000	12	12	18	12	18	18
9,000	12	12	18	12	18	18
10,000	15	15	18	15	18	18
11,000		15	18	15	18	18
12,000		18	18	15	18	18
13,000		18	18	18	18	18
14,000		18	18	18	18	18
15,000		18	18	18	18	18

<sup>45</sup> For example, between 6,000 and 9,000 pax/h the service with  $TnBn$  alternatives should be provided with 18 metre long buses, whereas 12 metre buses are preferred if boarding is at the front door only.



**Figure 7.3: Optimal frequency**

The optimal running speed  $v_0$  is analysed next. Figure 7.4 shows that  $v_0$  grows with demand at a decreasing rate; it is always increasing for the *TnBn* systems, while in the cases of on-board payment with cash, magnetic strip and contactless card through the front door only, the optimal running speed reaches a maximum and then decreases. This finding is related to the results on station spacing (Figure 7.6) and the previously analysed queuing delay (Figure 7.2). For example, for demand over 6,000 pax/h for front door boarding and payment with magnetic strip and contactless card, in order to reduce bus congestion it is optimal to decrease the distance between stations, or equivalently, increase the number of stations, so that the number of passengers per station is reduced, and consequently the queuing delay (that depends on the number of passengers getting on and off buses though the dwell time) is shorter. Therefore the number of stops along the route is increased. As the delay due to acceleration and deceleration per stop (equation 7.3) increases linearly with the constant running speed,  $v_0$ , the model responds by reducing the optimal speed to be provided.

On Figure 7.5, total infrastructure cost (land, busways and stations) is shown, i.e., the summation of costs  $c_1$  and  $c_2$  per kilometre, which grows following the evolution of

optimal speed, as the investment in busways ( $c_1$ ) is large relative to the investment in stations per kilometre ( $c_2 S^*/L$ , where  $S^*$  is the optimal number of stations along the route). A numerical analysis of the curves in Figure 7.4 and 7.5 shows that the optimal speed and investment in infrastructure per kilometre grow with the natural logarithm of demand for the case with no queuing delay (the off-board payment method):

$$v_0 = 16.07 \ln(N) - 73.17 \quad (R^2 = 0.998) \quad (7.25)$$

$$c_1 + c_2 \frac{S^*}{L} = 8.55 \ln(N) - 49.28 \quad (R^2 = 0.997) \quad (7.26)$$

That is, if bus congestion is negligible, it is optimal to increase infrastructure investment as demand grows, if that investment has an increase in running speed as output. Major increases in investment are suggested when facing growth in a low range of demand (for example, from 2,500 to 5,000 pax/h), whilst growth within a high demand range (for example, from 10,000 to 15,000 pax/h) is matched by a much smaller increment in investment, since the system is already relatively fast in terms of bus speed (buses running over 70 km/h<sup>46</sup>).

---

<sup>46</sup> Such high speeds are reachable with complete segregation of busways from traffic, such as the case of Brisbane's South East Busway in which buses run at 80 km/h (FTA, 2008).

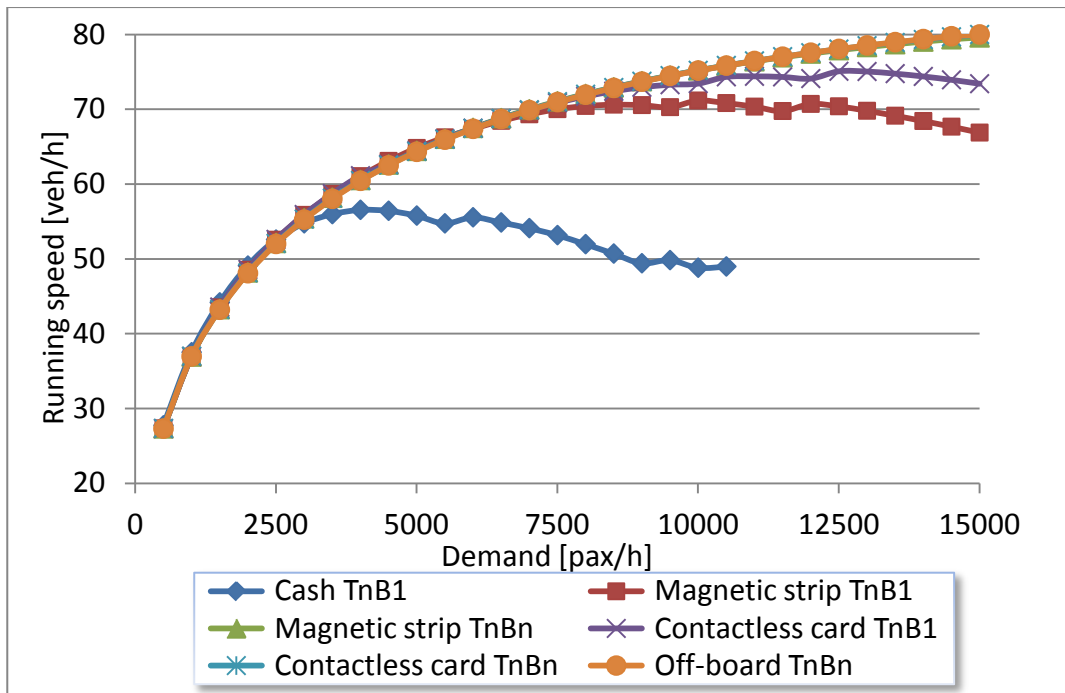


Figure 7.4: Optimal bus running speed

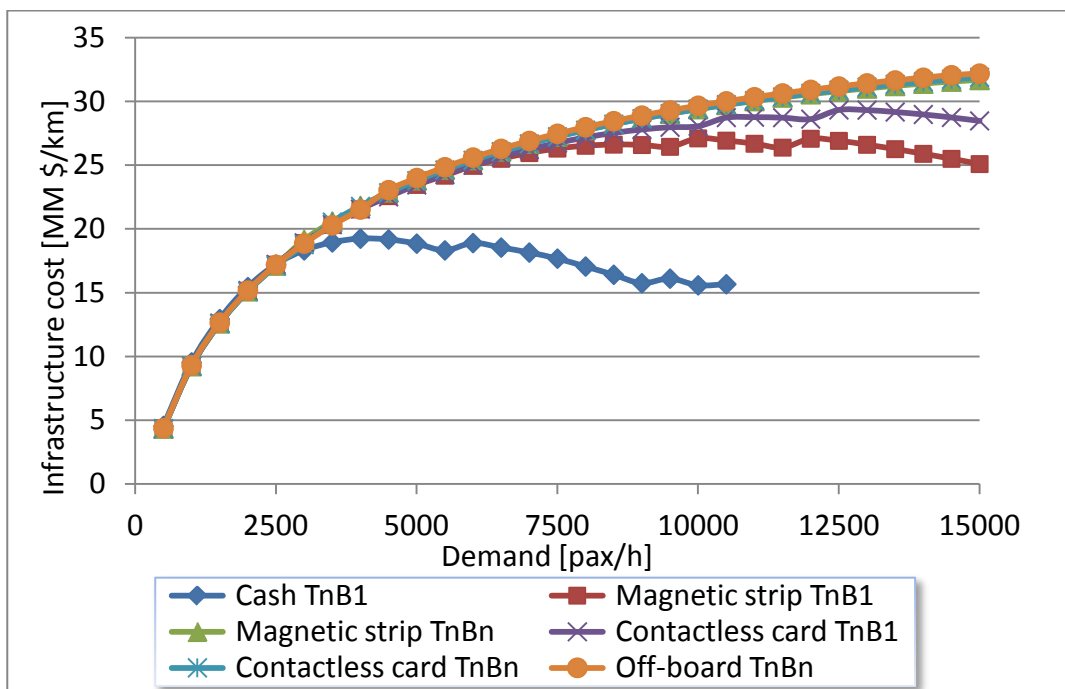
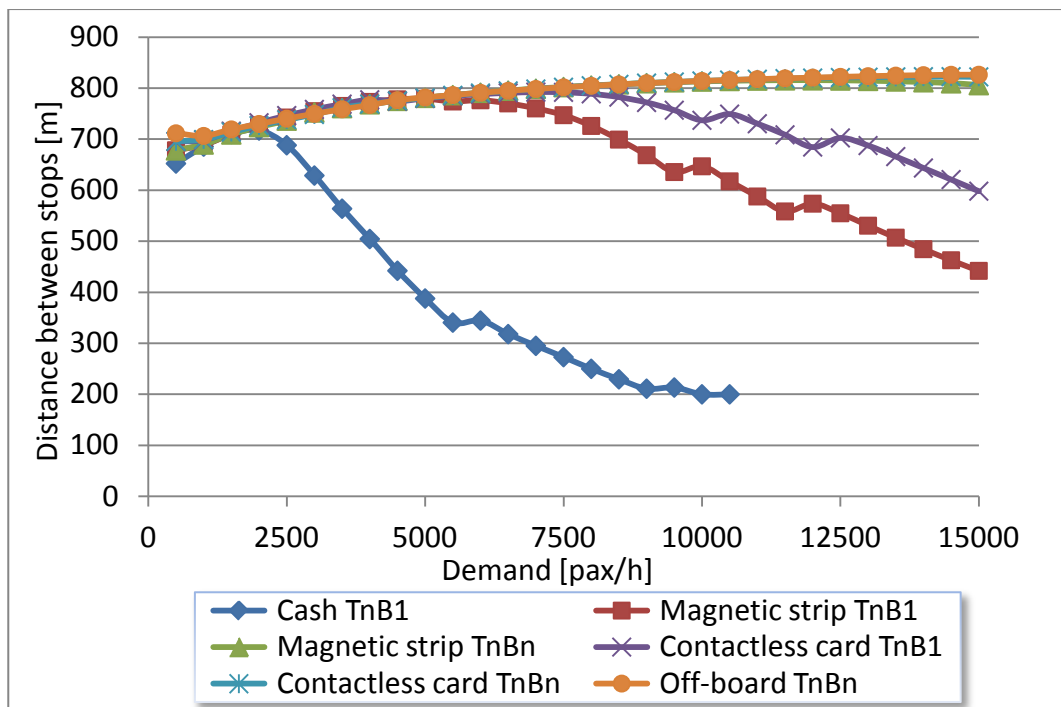


Figure 7.5: Infrastructure cost per kilometre

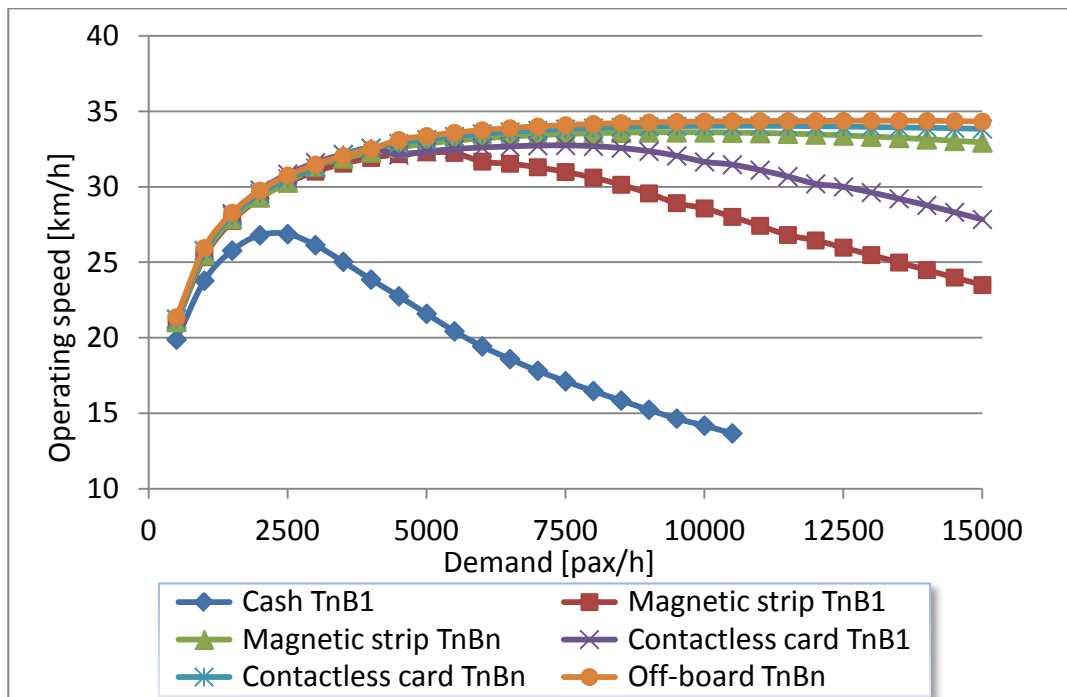


**Figure 7.6: Optimal bus stop spacing**

Given that bus running speed increases for a greater demand before bus stop congestion builds (Figure 7.4), it is interesting to analyse the resulting commercial or operating speed (Figure 7.7), which include running plus detentions due to bus stops and intersections<sup>47</sup>. For all systems, commercial speed increases with demand when demand is low due to the increase in operating speed; nevertheless, bus stop congestion reduces commercial speed for the systems with payment through the front door only (*TnB1*). A maximum speed of 34.4 km/h is reachable for 13,000 pax/h if off-board payment is provided (Figure 7.7), however, higher demand results in commercial speed slowly decreasing also for the *TnBn* alternatives. In summary, there is a middle demand range at which a bus corridor is able to realise its maximum potential in terms of speed and reduction of users costs; however, if demand is too high (over 13,000 pax/h-direction in our example) inevitably total speed is going to drop even if high investments in dedicated bus infrastructure are made and the most efficient fare collection system and boarding policy are implemented. In such cases, it might be appropriate to develop alternative operation strategies to speed up buses and increase the capacity of the corridor, for example, the combination of local and express services designed in the *Transmilenio* system in Bogotá, which

<sup>47</sup> The slack time introduced at the end of a roundtrip is not included.

records the highest transport capacity for a bus corridor around the world (45,000 pax/h, Wright and Hook, 2007).



**Figure 7.7: Operating speed (running plus detentions)**

Finally, Figure 7.8 depicts the number of buses required in each case, allowing for 5 percent of vehicles to be left as spare in depots. As expected, the implementation of efficient fare collection systems and boarding policies could imply sizeable savings in terms of fleet size requirements. This result, obtained with an optimisation model, is in line with the savings in fleet size due to upgrading the fare collection technology or speeding up the boarding process, estimated with the empirical bus travel time model of Section 4.5 (Table 4.8).

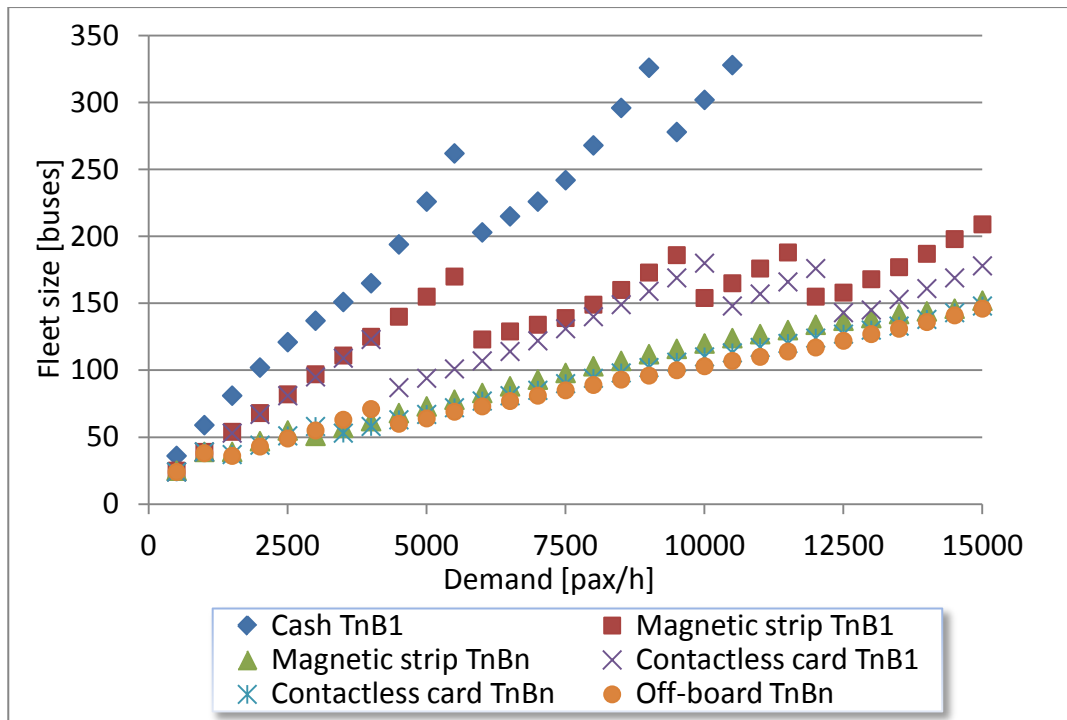


Figure 7.8: Fleet size

### 7.4.3 Analysis of other scenarios

Changes in the optimal values of the variables resulting from the application of other modelling assumptions are analysed in this section. A couple of new scenarios are compared against the *base case*, analysed in Section 7.4.2. We consider the case of on-board contactless card and front door boarding (*TnB1*), as the conclusions are the same for the other fare payment methods. Two modifications are introduced as follows:

#### A. The effect of ignoring congestion

The objective is to look at the differences in the design of the system that arise if the bus station queuing delays are disregarded in the optimisation. Such differences are observable only when there are queuing delays in the solution depicted in Figure 7.2. Results in Figure 7.9 show that ignoring the queuing delay (case labelled as *Congestion not accounted*) would yield larger frequencies for a ridership over 10,000 pax/h, given that from this point, the model with queuing delay (*base case*) starts to adjust the frequency (down) to reduce the increasing congestion in high-demand bus stops, as shown in Figure 7.2 (Curve “Contactless card TnB1”). As expected, the solution that

ignores the queuing delay in the design would yield a disproportionate average queuing time for high demands (Figure 7.10). Also, the consideration of congestion in the design yields a shorter distance between stations (Figure 7.11), with the objective of reducing the number of passengers boarding per stop, as previously discussed.

#### B. *Running speed cannot be optimised*

So far we have assumed that an increase in running speed through investment in infrastructure is possible, for example, by acquiring land and upgrading busways. However, in some cases, such improvements are not feasible due to physical, geographical or financial constraints, which prevent public transport planners and decision makers from building the dedicated busways they want for a high standard bus corridor, especially in city centres. In such a case, there is little or no room to influence running speed. In this scenario we fix the running speed  $v_0$  at 50 km/h. This scenario is referred to as *Fixed speed* in Figures 7.10 and 7.11. Optimal frequencies with fixed speed are virtually the same as in the base case (Figure 7.9). What does change is the optimal station spacing; as with fixed speed, the distance between stations is always decreasing, i.e., the higher the demand, the larger the number of stations. As already mentioned, when speed is optimised, there is a range in which the distance between stops increases with demand (up to 7,000 pax/h, *base case* curve in Figure 7.11) in order to reduce the delay due to acceleration and deceleration. On the other hand, the queuing delay with fixed speed is lower than in the base case, where for high demand, there are more stations (and consequently fewer passengers boarding on each of them) when the running speed is constant.



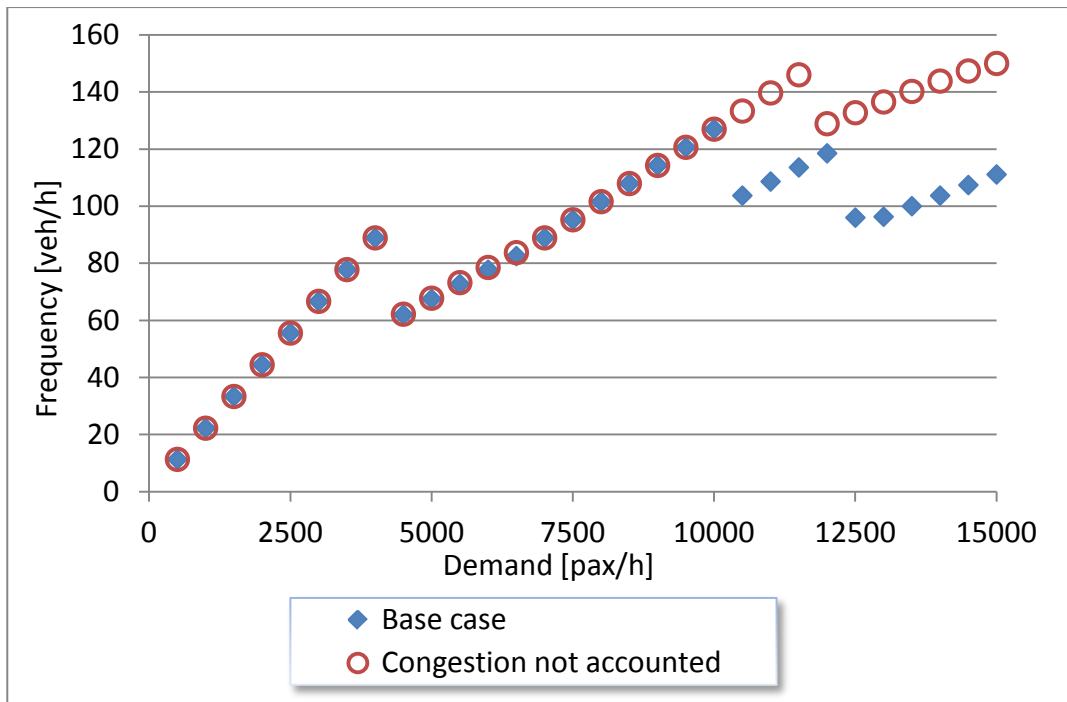


Figure 7.9: Optimal frequency, contactless card payment *TnB1*

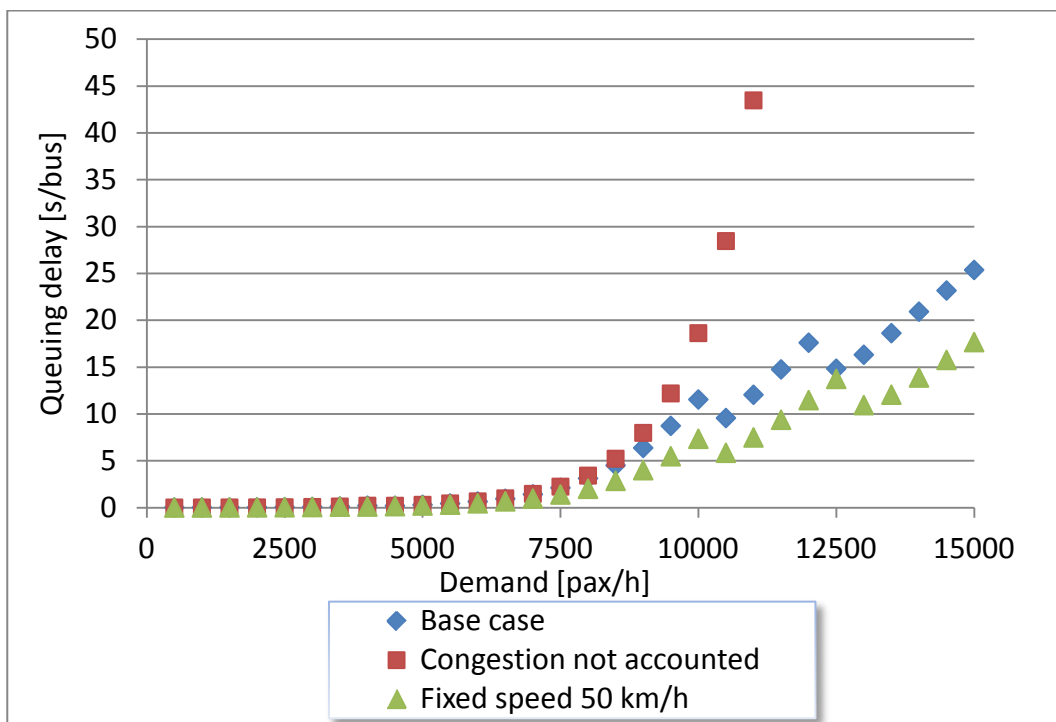
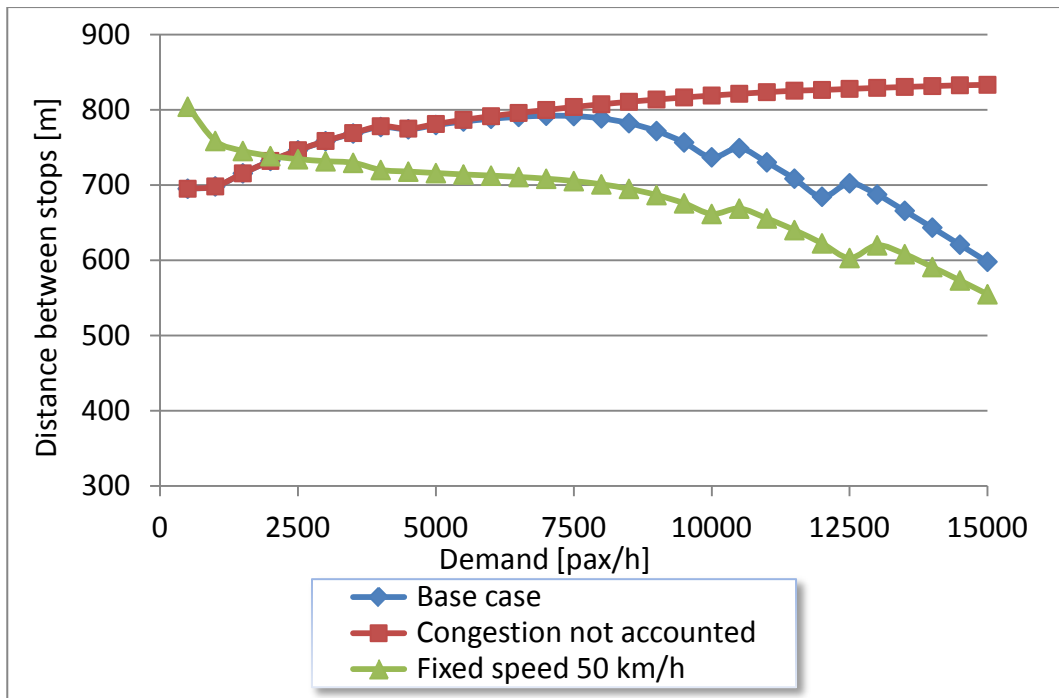


Figure 7.10: Queuing delay, analysis of scenarios with contactless card payment *TnB1*



**Figure 7.11: Optimal distance between stops, analysis of scenarios with contactless card payment *TnB1***

## 7.5 Summary and Conclusions

We have presented a model to optimise key variables in the design of bus services provided in a dedicated corridor, such as the frequency, bus capacity, station spacing, running speed and fare payment method. The former three variables have been widely analysed in the previous literature, whilst the consideration of different alternatives for the fare payment technology and the running speed as a variable (related to the investment in infrastructure for buses) are new in the microeconomic modelling of bus operations. The introduction of these two elements as decision variables in a formal model of public transport operation is not only a theoretical contribution, but also has the potential of influencing practice; given the widespread recognition of buses as crucial players in enhancing urban mobility in metropolitan areas, the selection of efficient fare collection policies and investment in dedicated infrastructure to speed up buses are two strategic instruments increasingly available to public transport policy makers to improve the quality of existing services and/or to design new high standard transit systems.

Another important feature of this approach is that congestion amongst buses is also embedded in the model, by considering that bus queues arise behind stations, if demand and frequency are high, which shows that a comprehensive treatment of bus congestion should go beyond the consideration of flow measures as explanatory variables, to also include passenger demand and the fare payment system. We provide a more detailed representation of the circulation of buses in a corridor than what is usual in microeconomic models of public transport operation, embedding relevant engineering aspects within the traditional approach of total cost minimisation (considering both operators plus users).

The main conclusions derived from the application of the model are the following:

- The systems with boarding at all doors (*TnBn*) provide the lowest total cost across the demand range under study. If all-door boarding is allowed, off-board fare collection emerges as the most efficient payment system for a demand larger than 600 pax/h; however the difference in total cost saving relative to on-board payment with magnetic strip and contactless card is no greater than 2 percent. Therefore, we can conclude that once on-board cash payment has been ruled out, allowing boarding at all doors is more important as a tool to reduce both users' and operator's costs than technological improvements on fare collection.
- For the off-board payment system and the on-board payment alternatives before congestion builds, optimal speed (and consequently the optimal investment in infrastructure) grows with the logarithm of demand. This result depends on the existence of a positive and linear relationship between investment in infrastructure and desired running speed.
- Decreasing total average cost is observed when boarding is allowed at all doors, whereas increasing average costs occur for high demand if boarding is restricted to the front door only. The highest total cost is associated with on-board cash payment, followed by payment with magnetic strip and contactless card. This is because buses spend more time in bus stops boarding passengers when payment is on-board at the front door only, which triggers bus queues that in turn increase travel time and operator costs.

- Disregarding bus congestion in the design of the service would yield greater frequencies when congestion is noticeable, i.e., for high demand.

The present model can be extended in several ways, for instance, by considering elasticity of demand, or the optimisation of two or more bus routes forming a network. This chapter has focused on dedicated bus corridors, and hence interactions with other modes were ignored. In the next chapter we include modal competition on a multimodal framework with bus, car and walking as travel alternatives, and congestion interactions between car and bus.

## **Chapter 8**

# **Optimal Design of Bus Services on an Extended Multimodal Framework**

### **8.1 Introduction**

The optimisation of a bus service on a dedicated corridor was the focus of Chapter 7, in which a total cost minimisation model was set up to analyse decisions on fare collection methods, bus boarding rules, and investment in bus road infrastructure, which is related to a target running speed. Bus congestion in the form of queuing delays was also considered, and its impact on design variables such as bus frequency and the spacing of bus stops was assessed. In this chapter, we extend the analytical model of Chapter 7 by incorporating modal choice, travel time interaction between cars and buses and the existence of bus crowding and its impact on the valuation of travel time savings. The new features of the model and contributions to the literature are introduced and highlighted in this section.

As discussed in Chapter 6, the analysis of the economic effects of crowding and standing costs inside public transport vehicles has focused, on the one hand, on estimating how the perception of travel time changes with levels of crowding, i.e., the influence of crowding on the value of travel time savings (Maunsell and Macdonald, 2007; Whelan

and Crockett, 2009; Hensher *et al.*, 2011; Wardman and Whelan, 2011), and on the other hand, on determining the effect of this crowding or comfort externality on the optimal bus fare (Kraus, 1991), and the values of bus frequency and size (Jara-Díaz and Gschwender, 2003).

The literature on crowding valuation suggests that the discomfort of travelling depends, among other things, on the number of passengers seating and standing, which can be reflected in the value of travel time savings through the estimation of in-vehicle time parameters that interact with the proportion of seats being used and the density of standees per square metre (Whelan and Crockett, 2009). The importance of people's dislike of crowding as a behavioural outcome goes beyond applications to estimate travel demand or the willingness to pay for a seat on a bus or train, as it can be used to determine the optimal distribution of space inside a vehicle, that is, the proportion of space that the designer should allocate to seating and standing. A change in the number of seats inside a bus or train has an impact on the discomfort of travelling, as it influences the number and density of passengers seating and standing. This is a key insight from the estimation of crowding and standing externalities that has been given no attention in the literature on the design and optimisation of public transport systems.

Microeconomic models that have included the level of crowding as an influence on the value of in-vehicle time savings do not distinguish between passengers seating and standing (Jara-Díaz and Gschwender, 2003; Tirachini *et al.*, 2010a; 2010b), whereas Kraus (1991) applies a premium on the value of travel time savings for passengers standing, but his work is concerned with the marginal cost and pricing of services considering the discomfort of standing, rather than with the design of vehicles. Thus, even though crowding and discomfort externalities have been analysed in the public transport economic literature, it is always assuming a given internal design or layout of the vehicles involved, i.e., a given bus or train capacity. In short, it is assumed that size implies capacity. In this chapter we develop a new approach to find the optimal bus capacity *given* bus size, on account of the fact that bus capacity can be manipulated through

different configurations of seating and standing layouts (i.e., different number of seats), given the length of a vehicle. The model is developed for buses but the same principles are applicable to rail.

A social welfare maximisation model with disaggregated origin and destination demand, and multiple travel alternatives, is proposed in this chapter, in a framework that includes bus frequency, bus size, number of bus seats, fare collection system, bus boarding policy, fare level and congestion toll as decision variables. In contrast to other social welfare maximisation models (e.g., De Borger *et al.*, 1996; Proost and Van Dender, 2004; Wichensin *et al.*, 2007; Ahn, 2009; Parry and Small, 2009; Basso and Silva, 2010; Jansson, 2010), this approach is more detailed in the characterisation of bus operations and includes a larger number of variables on the bus supply side, uncovers the trade-off between bus crowding and traffic congestion under several modelling assumptions, and shows that the inclusion of a non-motorised mode (walking) as an alternative to choosing bus and car may have a significant role when the transport system is optimised in highly congested scenarios. A numerical analysis over a transport corridor in Sydney is undertaken using the alternative specifications of multinomial logit models for mode choice, with and without crowding variables, that were estimated in Chapter 6.

The remainder of the chapter is organised as follows. The theoretical model is developed in Section 8.2, including assumptions and definitions (Section 8.2.1), demand and crowding modelling (8.2.2), travel time and congestion (8.2.3), internal bus layout (8.2.4) and operator cost items (8.2.5); the section concludes with the formulation of the social welfare maximisation problem. Section 8.3 presents the numerical application of the model to Sydney and discussion of results in several scenarios. Conclusions are provided in Section 8.4.

## 8.2. Model Set Up

### 8.2.1 Assumptions and definitions

We consider a linear bi-directional road of length  $L$  and a single period of operation with directions denoted as 1 and 2. The road is divided into  $P$  zones denoted as  $i \in \{1, \dots, P\}$ , and the total demand  $Y^{ij}$  per origin-destination pair  $(i, j)$  is fixed. The distance between zone  $i$  and zone  $i+1$  is denoted as  $L_i$  such that  $L = \sum_{i=1}^{P-1} L_i$ , as shown in Figure 8.1. Users can choose to travel by car ( $a$ ), bus ( $b$ ) or to walk ( $e$ ). Then, if  $y_m^{ij}$  is the travel demand for mode  $m$  between zones  $i$  and  $j$ , it holds that:

$$Y^{ij} = \sum_m y_m^{ij} = y_a^{ij} + y_b^{ij} + y_e^{ij} \quad (8.1)$$

Let  $f_{a1}^i$  be the traffic flow between zone  $i$  and zone  $i+1$  (direction 1) and  $f_{a2}^i$  be the traffic flow between zone  $i+1$  and zone  $i$  (direction 2). The decision variables of the problem are denoted as follows:

$f_b$  : bus frequency [bus/h]

$s_b$  : bus size (length) [m]

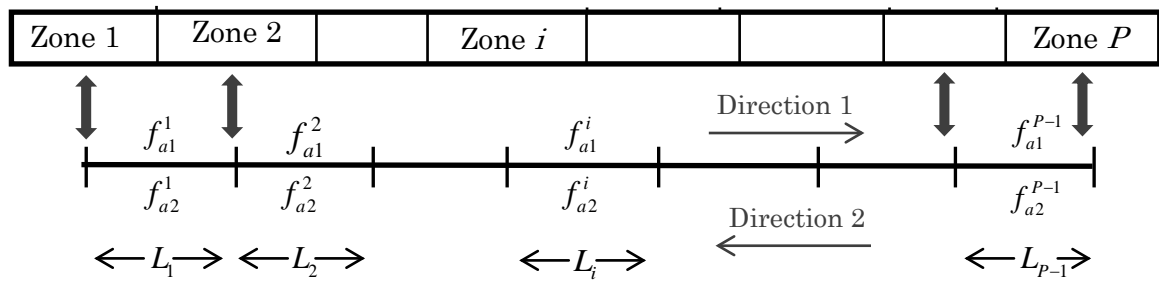
$n_{seat}$  : number of seats inside a bus

$\Delta$  : fare collection technology and boarding policy (one-door or all-door boarding)

$\tau_a$  : car toll [\$/trip]

$\tau_b$  : bus fare [\$/trip]





**Figure 8.1: Transport corridor diagram**

It is assumed that there is only one bus stop per zone<sup>48</sup> and that the travel distance between zones is the same for the three modes. Bus frequency is assumed to be continuous, whereas options on bus lengths are constrained by the size of commercial vehicles; four sizes are considered in the application of the model as defined in Section 7.3; these are mini (8 m., 1 or 2 doors), standard (12 m., 2 or 3 doors), rigid long (15 m., 3 or 4 doors) and articulated (18 m., 4 doors). A larger number of doors reduces boarding and alighting time (Chapter 4) but also reduces the capacity of the bus as the area next to a door must be left clear of seats and standees (Appendix A5). The number of seats  $n_{seat}$  can be freely chosen subject to lower and upper bounds, the former is given by a minimum number of seats per bus that is exogenously decided in order to provide a minimum level of service, whereas the latter is determined by a minimum area on a bus that must be clear of seats (i.e., aisle, doors, space for a wheelchair, area next to the driver). As defined in Section 4.2, we consider four alternative fare collection technologies: on-board payment with (i) cash, (ii) magnetic strip (with contact) and (iii) smart card (contactless), plus (iv) off-board payment (on the bus stop). The bus boarding and alighting policy can be chosen as well, two alternatives are available to implement in buses with more than one door: (a) simultaneous boarding and alighting, in which boarding is allowed at the front door only while alighting takes place at the back(s) doors (denoted  $TnB1$ ), and (b) sequential boarding and alighting, in which boarding is allowed

---

<sup>48</sup> The location of bus stops is fixed in this model, which allows us to know the number of passengers that a bus carries in each segment of the route (between two consecutive zones). For models that optimise the number of bus stops see Chapter 7, Kikuchi (1985), Chien and Schonfeld (1998) and dell’Olio *et al.* (2006) among others.

at all doors giving priority to passengers alighting (denoted  $TnBn$ ). In principle, we assume that cars and buses share the right-of-way and that bus stops do not directly affect cars, an assumption that is revised in Section 8.3.5.

### 8.2.2 Demand modelling and crowding

Mode choice models that include the proportion of available seats and the density of standees as attributes for buses are estimated. Data collected from a stated choice survey conducted in Sydney in 2009 is used to this end, as shown in Section 6.3 for public transport and car modes. Let  $U_m^{ij}$  be the utility associated with travel by mode  $m$  in OD pair  $(i, j)$ . In order to analyse differences in optimal bus service design due to alternative assumptions regarding user's valuations of seating, standing and crowding levels inside buses, we propose three different models that incorporate attributes representing the number of passengers seating and standing, interacting with travel time; these models will be compared with a specification that ignores any crowding or standing cost. The models, named M1 to M4 for the bus mode, are described as follows:

- M1: No crowding cost (eq. 8.2).
- M2: Only the density of standees [pax/m<sup>2</sup>] imposes an extra discomfort cost (eq. 8.3).
- M3: The density of standees and the proportion of seats occupied are sources of disutility (eq. 8.4).
- M4: The density of standees and the proportion of seats occupied are squared in the utility function (eq. 8.5).

$$\text{Bus - M1: } U_b^{ij} = \alpha_b^{\text{M1}} + \beta_a^{\text{M1}} t_{ab}^i + \beta_h^{\text{M1}} h_b + \beta_{vb}^{\text{M1}} t_{vb}^{ij} + \beta_c^{\text{M1}} \tau_b \quad (8.2)$$

$$\text{Bus - M2: } U_b^{ij} = \alpha_b^{\text{M2}} + \beta_a^{\text{M2}} t_{ab}^i + \beta_h^{\text{M2}} h_b + \beta_{vb}^{\text{M2}} t_{vb}^{ij} + \beta_c^{\text{M2}} \tau_b + \beta_{den}^{\text{M2}} n_{den} t_{vb}^{ij} \quad (8.3)$$

$$\text{Bus - M3: } U_b^{ij} = \alpha_b^{\text{M3}} + \beta_a^{\text{M3}} t_{ab}^i + \beta_h^{\text{M3}} h_b + \beta_{vb}^{\text{M3}} t_{vb}^{ij} + \beta_c^{\text{M3}} \tau_b + \beta_{den}^{\text{M3}} n_{den} t_{vb}^{ij} + \beta_{seat} p_{seat} t_{vb}^{ij} \quad (8.4)$$

$$\text{Bus - M4: } U_b^{ij} = \alpha_b^{\text{M4}} + \beta_a^{\text{M4}} t_{ab}^i + \beta_h^{\text{M4}} h_b + \beta_{vb}^{\text{M4}} t_{vb}^{ij} + \beta_c^{\text{M4}} \tau_b + \beta_{den2} n_{den}^2 t_{vb}^{ij} + \beta_{seat2} p_{seat}^2 t_{vb}^{ij} \quad (8.5)$$

In (8.2) to (8.5),  $t_{ab}^i$  is the access time at zone  $i$ ,  $h_b$  is the headway between two consecutive buses,  $t_{vb}^{ij}$  is the in-vehicle time between zones  $i$  and  $j$ ,  $\tau_b$  is the bus fare,  $n_{den}$  is the density of standees per square metre,  $p_{seat}$  is the proportion of seats been used,  $\alpha_b$  is an alternative specific constant (which will be calibrated to predict an observed modal split) and  $\beta_k$  are the parameters associated with the different attributes. For each model (M1 to M4), the utility of the alternative modes (car and walk) have the same specification:

$$\text{Car: } U_a^{ij} = \beta_{va}^{M \cdot} t_{va}^{ij} + \beta_{ca}^{M \cdot} (c_r^{ij} + \tau_a) / o_r \quad (8.6)$$

$$\text{Walk: } U_e^{ij} = \alpha_e + \beta_{ve}^{M \cdot} t_{ve}^{ij} \quad (8.7)$$

where  $c_r^{ij}$  is the car running cost to travel between zones  $i$  and  $j$ ,  $\tau_a$  is the road charge (decision variable) and  $o_r$  is the average car occupancy rate (therefore  $U_a^{ij}$  is the average utility of car users). Assuming a multinomial logit model for the estimation of demand, the number of trips by mode  $m$  in OD pair  $(i, j)$  is given by:

$$y_m^{ij} = Y^{ij} \frac{e^{U_m^{ij}}}{\sum_n e^{y_n^{ij}}} \quad \forall i, j \quad (8.8)$$

where  $Y^{ij}$  is the total demand between zones  $i$  and  $j$ . The estimation of parameters for models M1 to M4 was discussed in Section 6.3 (Table 6.1) and will be summarised in Section 8.3.1 in this chapter, including the estimation of the travel time parameter for walking.

In this framework, the consumer surplus  $B$  is given by the logsum formula:

$$B = \sum_{ij} \frac{y_m^{ij}}{I_u} \ln \sum_m e^{U_m^{ij}} + B_0 \quad (8.9)$$

where  $I_u$  is the marginal utility of income<sup>49</sup>, equal to minus the cost parameter  $\beta_c^M$  estimated with the choice models<sup>50</sup>, and  $B_0$  is a constant that has no effect on the solution of the problem, and therefore can be set to zero.

### 8.2.3 Travel time, congestion and bus stop delay

We assume that buses and cars share the right-of-way, which is subject to congestion. Furthermore, buses have to stop at bus stops to load and unload passengers. Bus stops are also subject to congestion in the form of queuing delays when the bus frequency is high and/or the dwell time is long. Taking direction 1 for illustration, we model travel time between zone  $i$  and zone  $i+1$  by car ( $t_{va1}^i$ ) and bus ( $t_{vb1}^i$ ) as a function of traffic flow and bus frequency by using the well-known Bureau of Public Roads (BPR) formula<sup>51</sup>:

$$t_{va1}^i(f_{a1}^i, f_b) = t_{a0}^i \left[ 1 + \alpha_0 \left( \frac{f_{a1}^i + \varphi(s_b) f_b}{K_r} \right)^{\alpha_1} \right] \quad (8.10)$$

$$t_{vb1}^i(f_{a1}^i, f_b) = t_{b0}^i \left[ 1 + \alpha_0 \left( \frac{f_{a1}^i + \varphi(s_b) f_b}{K_r} \right)^{\alpha_1} \right] + t_{s1}^i \quad (8.11)$$

where  $t_{a0}^i$ ,  $t_{b0}^i$ ,  $\alpha_0$  and  $\alpha_1$  are parameters ( $t_{a0}^i$  and  $t_{b0}^i$  are the free-flow travel times),  $\varphi \geq 1$  is the passenger car equivalency factor of a bus, which depends on the bus length  $s_b$ , and  $K_r$  is the capacity of the road<sup>52</sup>. The travel time by bus includes the delay due to

---

<sup>49</sup> The marginal utility of income is assumed constant, i.e., we ignore income effects on demand (Jara-Díaz and Videla, 1989; Jara-Díaz, 2007).

<sup>50</sup> Note that  $\beta_c^M$  is estimated with the choice of motorised modes only because walking is for free.

<sup>51</sup> For simplicity, we disregard the formulation that specifically accounts for delays due to traffic lights used in Chapter 7, because the mixed-traffic scenario analysed in this chapter has a saturation degree that produces overflow delays on top of non-random delays (see equation 7.4), which greatly complicates the expression for the overall delay due to traffic signals; however, the road capacity  $K_r$  of the implemented BPR function will take into account the effective green time ratio of traffic signals, as described in Section 8.3.1.

<sup>52</sup> A static model like (8.10)-(8.11) cannot accommodate hypercongestion, which refers to the fact that a small throughput or *outflow* is possible not only with a low *inflow* demand, but also with a high inflow on

bus stops,  $t_{s1}^i$ , which consists of the acceleration and deceleration delay  $t_{ac1}^i$ , the average queuing time  $t_{q1}^i$  and the dwell time  $t_{d1}^i$  (ignoring internal delays at bus stops), i.e.,

$$t_{s1}^i = t_{ac1}^i + t_{q1}^i + t_{d1}^i \quad (8.12)$$

The acceleration and deceleration delay and queuing delay are modelled as in Chapter 7. The delay in the process of accelerating and decelerating at bus stops assumes uniform acceleration (at rate  $r_a$  m/s/s) and deceleration (at rate  $r_d$  m/s/s), thus the extra stopping delay on top of the uniform travel time given by the running speed  $v_{b1}^i$ , is expressed as (8.13), and the queuing delay at bus stops is presented in equation (8.14)

$$t_{ac1}^i = \frac{v_{b1}^i}{2} \left( \frac{1}{r_a} + \frac{1}{r_d} \right) \quad (8.13)$$

$$t_{q1}^i = 0.001 \left[ b_0 + b_{l1}s_b + (b_{d1} + b_{d2}Z_2 + b_{d3}Z_3)t_{d1}^i \right] e^{0.001f_b[b_f + b_{l2}s_b + (b_{d4} + b_{d5}Z_2 + b_{d6}Z_3)t_{d1}^i]} \quad (8.14)$$

where  $s_b$  [m] is the bus length,  $t_{d1}^i$  [s/bus] is the dwell time,  $f_b$  [veh/h] is the bus frequency and  $b_0$ ,  $b_{L1}$ ,  $b_{L2}$ ,  $b_{d1}$ ,  $b_{d2}$ ,  $b_{d3}$ ,  $b_{d4}$ ,  $b_{d5}$ ,  $b_{d6}$  and  $b_f$  are estimated parameters and factors 0.001 are introduced for scaling of the parameters (see Appendix A2 for further details).  $Z_2$  and  $Z_3$  are dummy variables defined as follows:

$$Z_2 = \begin{cases} 1 & \text{if bus stop has two berths} \\ 0 & \text{otherwise} \end{cases}$$

---

traffic breakdown, reached when inflow is higher than capacity (Walters, 1961). This outcome is represented by a backward bending shape of the speed-flow curve. However, it is relevant to note that even though hypercongestion may be necessary to model if the focus is on peak periods with severe congestion, hypercongestion is a temporary phenomenon generally caused by a bottleneck, and despite the fact that the instantaneous relationship between travel time and performed flow can be backward bending, this has nothing to do with the supply curve to be used in the economic analysis of road transport, which has to be a function of the quantity demanded or *inflow*, the latter relationship being monotonic (i.e., when more vehicles use a road, travel time is the same or greater). For a more detailed discussion on hypercongestion, see May *et al.* (2000) and Small and Chu (2003).

$$Z_3 = \begin{cases} 1 & \text{if bus stop has three berths} \\ 0 & \text{otherwise} \end{cases}$$

As discussed in Chapter 7, the estimation of the dwell time per stops requires the cases with boarding allowed at all doors (*TnBn*) and at the front door only (*TnB1*) to be addressed separately, since in *TnBn* boarding and alighting is sequential at all doors, whereas in *TnB1* boarding at the front door occurs simultaneously with alighting at the rear doors. These two cases are summarised in expression (8.15)

$$t_d^i = \begin{cases} c_{oc} + p_b \beta_b \lambda^+ + p_a \beta_a \lambda^- & \text{if boarding at all doors (TnBn)} \\ c_{oc} + \max \{ \beta_b \lambda^+, p_a \beta_a \lambda^- \} & \text{if boarding at front door only (TnB1)} \end{cases} \quad (8.15)$$

where  $c_{oc}$  is the time to open and close doors,  $\beta_a$  and  $\beta_b$  are the average alighting and boarding times per passenger,  $\lambda^+$  and  $\lambda^-$  are the number of passengers boarding and alighting a bus at the bus stop and factors  $p_a$  and  $p_b$  are the proportion of passengers boarding and alighting at the busiest door, given in Table 7.3. Equations (8.13), (8.14) and (8.15) conclude the derivation of the delay at bus stops (8.12).

#### **8.2.4 The choice of bus size and internal layout**

Using data from London, Jansson (1980) finds a linear relationship between bus running costs and bus size measured as the number of seats per bus, a relationship that has been used by Jansson and other authors to find the optimal size of buses in urban routes (e.g., Jara-Díaz and Gschwender, 2003) under the implicit assumption that there is a unique relationship between bus size and capacity, measured as number of seats or total number of passengers that can be carried, as also assumed in Chapter 7. However, the number of passengers that a bus can carry is not only given by the bus size, but also by the internal layout of space allocated to seating and standing, as a passenger sitting takes up more space than a passenger standing. A standard value for the area needed for a passenger sitting is 0.5 square metres (TRB, 2003), whereas, depending on crowding conditions, passengers standing may have a density of up to five or six passengers per square metre, and as such the minimum area required by a standee is approximately 0.17-0.20 square metres, i.e., less than half the space required for a person seated.

Therefore, if the number of seats inside a bus can be manipulated, there is no one-to-one relationship between capacity and bus size, and the final capacity of a bus is outcome function of decisions made about both the bus length and internal layout altogether. In this context, bus capacity is not an absolute value, but rather a function of the maximum density of standees that is acceptable to have, given by policy, demand and cultural constraints<sup>53</sup>.

In this chapter, we consider the number of seats as a variable, which triggers a trade-off between comfort and capacity. That is, decreasing the number of seats on a bus increases its capacity at the expense of reducing the comfort of travelling, represented by a higher cost of standing and crowding. Several physical constraints need to be considered when deciding the number of seats, including minimum space for aisles, doors and in front of the bus (next to the driver) that must be clear of seats. Formulae for seating and standing areas and constraints are presented in Appendix A5.

### **8.2.5 Bus operator cost and problem formulation**

Operator cost is divided into four components which are obtained in the same way as in Chapter 7:

$c_2$ : Station infrastructure cost [\$/station-h]

$c_3$ : Personnel costs (crew) and vehicle capital costs [\$/bus-h], and

$c_4$ : Running costs (fuel consumption, lubricants, tyres, maintenance, etc.) [\$/bus-km]

$c_5$ : Implementation cost related to the fare payment technology (e.g., software requirements) [\$/h]

---

<sup>53</sup> In crowded bus and train systems in Asia or Latin America is not unusual to operate at crush capacity, with 6 passengers standing per square metre in peak periods, however, such a high density of standees could not be acceptable in other regions.

The busway infrastructure and land cost  $c_1$  of Chapter 7 is disregarded because, in this chapter, buses are assumed to run in an existing mixed-traffic road. Following equations (7.17) to (7.20), the total operator cost  $C_o$  can be defined as (8.16):

$$C_o(f_b, s_b, S, \Delta) = c_2(s_b, \Delta)S + c_3(s_b, \Delta)f_b T_c(f_b, s_b, S, v_0, \Delta) + 2c_4(s_b)Lf_b + c_5(\Delta) \quad (8.16)$$

Importantly, we are assuming that the number of seats inside a bus (and consequently, the number of passengers) has no effect on the bus capital cost, which is only determined by the bus size and arrangements regarding fare collection readers (i.e., the cost of seats if assumed negligible relative to the cost of the bus). After obtaining an expression for the operator cost (8.16), we can formulate the social welfare maximisation problem as follows:

$$\text{Max} \quad SW = \sum_{ij} \frac{y^{ij}}{I_u} \ln \sum_m e^{U_m^{ij}} + \sum_{ij} y_a^{ij} \tau_a + \sum_{ij} y_b^{ij} \tau_b - C_o \quad (8.17)$$

Subject to

$$\max_i \{y_{b1}^i, y_{b2}^i\} \leq \kappa f_b K_b(s_b, n_{seat}) \quad (8.18a)$$

$$n_{seat}^{\min} \leq n_{seat} \leq n_{seat}^{\max} \quad (8.18b)$$

$$f_b^{\min} \leq f_b \leq f_b^{\max} \quad (8.18c)$$

$$s_b \in \{s_{b1}, \dots, s_{b4}\} \quad (8.18d)$$

$$\Delta \in \{\Delta_1, \dots, \Delta_6\} \quad (8.18e)$$

Inequality (8.18a) is a capacity constraint that ensures that the bus transport capacity ( $\kappa f_b K_b$ ) is large enough to accommodate the maximum bus load;  $\kappa$  is a design factor introduced to allow spare capacity to absorb random variations in demand (for example,  $\kappa = 0.9$ ) and  $K_b$  is the bus capacity, given by the bus length  $s_b$  and number of seats  $n_{seat}$ . Expression (8.18b) states that the number of seats is constrained by minimum and maximum values, which are obtained in Appendix A5. Frequencies are also constrained



by a minimum policy frequency  $f_{min}$  (set to have a minimum level of service, if desired) and the maximum feasible frequency  $f_{max}$  as given in expression (8.18c). Finally, (8.18d) and (8.18e) establish that bus size  $s_b$  and the boarding and alighting policy and fare collection technology  $\Delta$  are taken from available choices.

The constrained optimisation (8.17)-(8.18) is solved using the optimisation toolbox of Matlab. The solution procedure implemented considers bus frequency as a continuous variable while the number of seats, car toll and bus fare are discrete (fare and toll are constrained to be a multiple of 5 cents). In this setting, modal choice depends on travel times, which in turn depend on modal choice; this fixed-point problem is solved by simply iterating between modal choice and travel times until convergence is reached.

### 8.3 Application

#### 8.3.1 Physical setting and input parameters

The social welfare maximisation model is applied with demand and supply data from Military Road in North Sydney (Figure 8.2). The section modelled comprises 3.44 km of road which is divided in 12 zones (therefore the average zone length is 286 metres). The origin-destination matrix for car and bus trips is obtained from a traffic simulation study undertaken in this corridor by the Roads and Traffic Authority (RTA)<sup>54</sup>. In order to add walking trips to the matrix we use Sydney's Household Travel Survey (TDC, 2010) to obtain the city's modal split by trip distance; 66.7 percent of trips shorter than one kilometre are made on foot, a figure that drops to 24.7 percent for trips between 1 and 2 km, and 5.7 percent for trips between 2 and 5 km (considering car, bus and walk only). Then we amplify each cell (bus+car trips) by the respective percentage of walking trips according to the distance between origin and destination. The matrix obtained with this

---

<sup>54</sup> This corridor is chosen because of the availability of origin-destination demand data at the level of small zones. The estimation of taste parameters for utility functions (8.2) to (8.7) is done with data collected in an adjacent area in Sydney (the CBD and the North West); we assume that the estimated parameters are also applicable to the Military Road area.

procedure is presented in Figure 8.3, with a total of 19,234 trips in the morning peak (7.30 to 8.30am), from which 54.3 percent are from east to west, towards the CBD (Direction 2 in Figure 8.1).

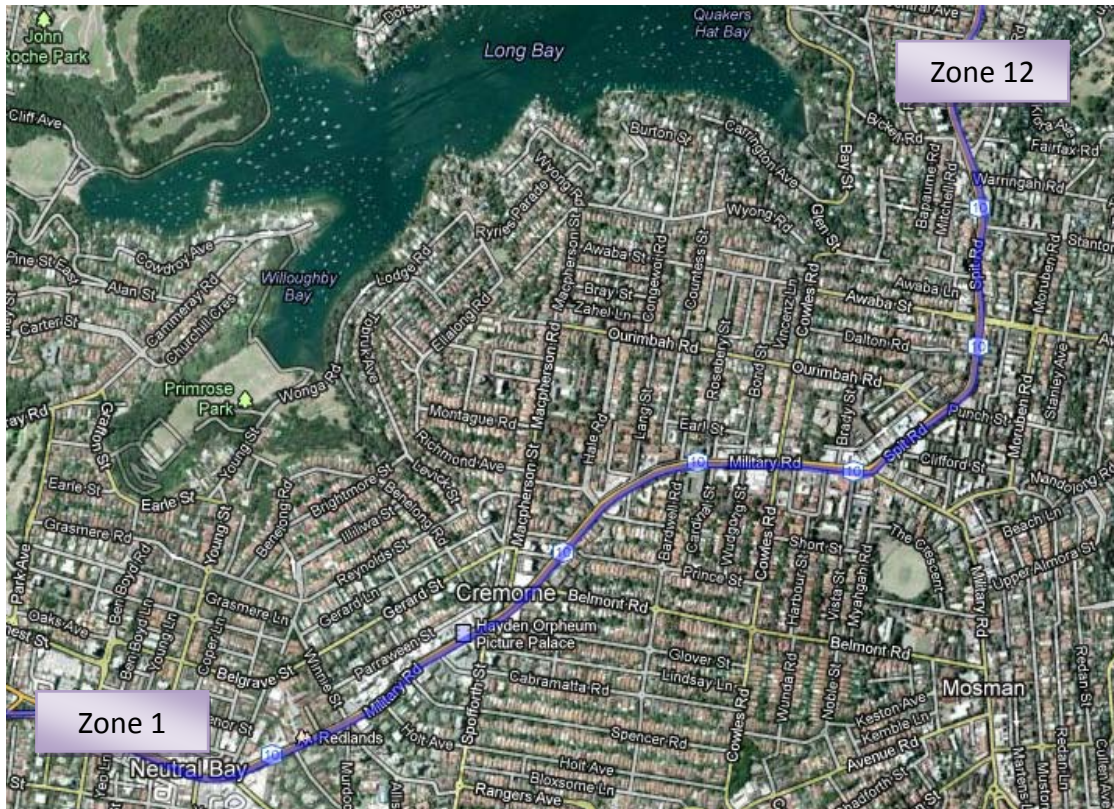


Figure 8.2: Test corridor, Military Road

O/D	1	2	3	4	5	6	7	8	9	10	11	12
1	0	856	1324	54	23	8	74	99	419	71	16	1405
2	165	0	192	15	4	1	20	19	68	14	3	326
3	829	93	0	0	0	0	0	0	0	0	0	0
4	50	12	0	0	0	0	1	3	13	1	0	91
5	146	0	0	0	0	0	0	0	1	0	0	11
6	235	9	3	0	0	0	0	3	9	0	0	17
7	87	13	4	0	0	0	0	12	48	12	0	187
8	18	1	0	0	0	0	0	0	3	9	0	8
9	396	22	5	1	1	3	24	9	0	27	3	763
10	7	0	0	0	0	0	0	0	0	0	12	1511
11	119	11	1	0	0	0	12	0	3	123	0	1027
12	1780	277	54	21	16	27	151	65	207	3763	1685	0

**Figure 8.3: Origin-Destination matrix**

The road has two lanes per direction, BPR functions (8.10) and (8.11) are assumed to represent travel times with commonly used parameter values  $\alpha_0 = 0.15$  and  $\alpha_1 = 4$ , and a capacity  $K_r = 2000 \text{ veh/h}$  obtained by assuming a 60 percent for effective green time ratio at signalised intersections. Speed at free flow is 50 km/h. With these assumptions, the average car speed is 26.3 km/h in direction 1 (outbound) and 21.5 km/h in the direction 2 (inbound), similar to the measured average speed of 22 km/h on this road (RTA, 2011, which only reports average speed in the inbound direction in the morning peak). The bus equivalency factors  $\varphi(s_b)$  are 1.65 for small buses (8 m), 2.19 for standard buses (12 m), 2.60 for rigid long buses (15 m) and 3.00 for articulated buses (18 m), following the linear relationship of Basso and Silva (2010).

Users can choose between travelling by car, bus or to walk; other alternatives like switching time period or changing origin and/or destination are not considered. The car operating cost is 14 cents/km (fuel consumption) and the average car occupancy 1.45 pax/car (TDC, 2010), which we assume remains unchanged after pricing reforms (the

sensitivity of car occupancy to raising tolls is ignored). Walking speed is assumed to be 4 km/h.

The constraints for the minimum and maximum number of seats per bus are explained in Appendix A5. As a minimum, a space free of seats must be left next to the driver, next to doors, for a central corridor and for a wheelchair, which would determine the maximum number of seats  $n_{seat}^{max}$  that can be fit in a bus. On the other hand, the minimum number of seats  $n_{seat}^{min}$  is exogenously fixed such that, regardless of the size of the bus, at least 25 percent of the available area for seating and standing is allocated to seating. It is assumed that a passenger sitting needs  $0.5 \text{ m}^2$  whereas the maximum density of standees is  $4 \text{ pax/m}^2$ , i.e.,  $0.25 \text{ m}^2$  per passenger (see Appendix A5).

Parameters for the utility functions (8.2) to (8.7) are taken from Table 6.1, with the exceptions of the time parameter for walking and the mode specific constants, which are estimated as follows. First, walking as a travel alternative was not considered in the survey of the main stated choice experiment from 2009 in Sydney, described in Section 6.3; therefore a reasonable value for the disutility of travel time while walking has to be supplemented. To this end, a secondary intra-CBD model described in an internal 2009 report by Hensher and Rose is used, in which walking was an alternative to public transport modes and taxi for short CBD trips; in the intra-CBD model, it is found that the time parameter of walking ( $\beta_{ve}$ ) is 1.86 times greater than the in-vehicle time parameter for bus ( $\beta_{vb}$ )<sup>55</sup>. Thus, we assume a constant value of  $\beta_{ve}$  across models, equal to 1.86 times  $\beta_{vb}$  on M1 (because the latter is an average value of  $\beta_{vb}$  for all crowding conditions); therefore,  $\beta_{ve} = 1.86 \cdot -0.019 = -0.035$ .

---

<sup>55</sup> This figure is in the order of the values estimated by Jovicic and Hansen (2003) for Copenhagen (1.36 and 2.32 for the ratio  $\beta_{ve} / \beta_{vb}$  for purposes commuting and education, respectively, considering walking and cycling altogether as a non-motorised mode, and trips up to 30 minutes long).

Second, mode specific constants for demand models M1 to M4 are calibrated to represent the current Sydney modal split of trips shorter than 5 kilometres: 62.5 percent car, 31.6 percent walk, and 5.9 percent bus (TDC, 2010). The current bus frequency of 16 bus/h in the morning peak is used, with a fare of \$2.10 and no car toll. The car specific constant is fixed at zero. With these two considerations for the time parameter for walking and the mode specific constants, the estimated parameters used in this section are presented in Table 8.1 (goodness-of-fit and t-ratios were shown in Table 6.1).

**Table 8.1: Parameter values**

Attribute	M1	M2	M3	M4
Access time $\beta_a$	-0.016	-0.017	-0.017	-0.017
Headway $\beta_h$	-0.009	-0.010	-0.010	-0.010
In-vehicle time bus ( $t_{vb}$ ) $\beta_{vb}$	-0.019	-0.013	-0.004	-0.006
In-vehicle time car ( $t_{va}$ ) $\beta_{va}$	-0.016	-0.018	-0.018	-0.018
Travel time walk $\beta_{ve}$	-0.035	-0.035	-0.035	-0.035
Cost $\beta_c$	-0.062	-0.064	-0.064	-0.064
Modal constant bus $\alpha_b$	-2.080	-2.112	-2.129	-2.134
Modal constant walk $\alpha_e$	-0.092	-0.099	-0.099	-0.100
$t_{vb} \times \text{den stand } \beta_{den}$		-0.004	-0.003	
$t_{vb} \times \text{prop seat } \beta_{seat}$			-0.013	
$t_{vb} \times (\text{den stand})^2 \beta_{den2}$				-0.0005
$t_{vb} \times (\text{prop seat})^2 \beta_{seat2}$				-0.013

Note: Time in minutes, cost in \$ (AUD).

### 8.3.2 Base results

Results with the current OD matrix (Figure 8.3) for demand models M1 to M4 are shown in Table 8.2. First, the solution regarding bus size, frequency, fare, toll and number of seats is similar for M1 and M2, and for M3 and M4. In the case of M1 (no crowding or standing externality internalised) it is optimal to operate with mini buses (8 metre long)

at a frequency of 21.7 veh/h and to charge a fare of 10 cents, whereas in M2 (with standing disutility) the optimal solution has a slightly greater frequency of 23.7 veh/h. The similarity of results is because at this level of bus demand almost all passengers are seating, as shown by the maximum occupancy rate (over number of seats), which is 1.08 for M1 and 0.98 for M2 (tenth row in Table 8.2), therefore, due to the absence of standees, both models have similar optimal outputs. A different result is obtained if we assume that the proportion of bus riders seating also is a source of disutility, either in a linear (M3) or quadratic (M4) form; in these cases the optimal solution comprises bigger (12 m) and more frequent buses (between 25 and 26.1 veh/h), and the optimal fare that escalates to 40 cents. The difference in fare is explained by the fact that in M1 and M2 the marginal cost of carrying an extra passenger is only given by the extra boarding and alighting time, whereas for M3 and M4 the optimal fare also accounts for the discomfort caused by a passenger that reduces the number of free seats on a bus.

Next, regarding the optimal number of seats, in all cases the optimal result is having the maximum number of seats possible (24 for 8 m.-long buses, 39 for 12 m.-long buses), constrained by the minimum area required free of seats<sup>56</sup>. Out of the available area for seating and standing, 80 percent is allocated to seating and 20 percent to standing. The greater frequency and bus size of models M3 and M4 considerably reduces the average occupancy rate (as a function of the number of seats) from over 50 percent in M1 and M2, to 30 percent in M3 and M4 (the supply of seats per hour is almost doubled from 521 in M1 to 1,017 in M3).

---

<sup>56</sup> Note that bus utility in M1 is indifferent to the number of seats inside buses, therefore as long as the capacity constraint is not binding, any number of seats would produce the same level of social welfare. In Table 8.2 the capacity constraint is inactive for buses with the maximum number of seats  $n_{seat}^{max}$ , therefore  $n_{seat}^{max}$  is arbitrarily chosen for M1.

**Table 8.2: Base case results**

<b>Optimal value</b>	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>
Bus length [m]	8	8	12	12
Frequency [veh/h]	21.7	23.7	26.1	25.0
Fare [\$]	0.1	0.1	0.4	0.4
Toll [\$]	2.0	2.0	2.0	2.0
Number of seats	24	24	39	39
Bus capacity [pax/bus]	36	36	58	58
Seating area/total bus area	0.58	0.58	0.63	0.63
Seating area/ (seating plus standing area)	0.80	0.80	0.80	0.80
Average occupancy rate (over number of seats)	0.57	0.52	0.30	0.31
Max. occupancy rate (over number of seats)	1.08	0.98	0.56	0.58
Max. occupancy rate (over total capacity)	0.62	0.56	0.32	0.33
Seat capacity bus route (seats/h)	521	569	1,017	975
Total capacity bus route (pax/h)	782	854	1,512	1,450
Fare collection technology	Off-board	Mag. strip	Mag. strip	Mag. strip
Boarding regime	All doors	All doors	All doors	All doors
Social welfare [\$]	129,544	122,984	122,897	122,801
Consumer surplus [\$]	114,290	107,721	107,454	107,319
Bus operator profit [\$]	-671	-645	-467	-436
Toll revenue [\$]	15,925	15,908	15,909	15,917
Subsidy/bus operator cost	0.83	0.83	0.46	0.44
Fleet size [buses]	11	12	13	13
Modal split bus	7.1%	7.0%	7.0%	7.1%
Modal split car	60.0%	59.9%	60.0%	60.0%
Modal split walk	32.9%	33.1%	33.0%	32.9%

The outputs regarding number of doors, bus boarding policy and fare collection technique are described as follows. First, in all cases it is optimal to have the maximum number of doors given by the bus size, i.e., 2 doors for 8 metre buses and 3 doors for 12 metre buses, as the more doors are in place the shorter are the boarding and alighting times<sup>57</sup>. Second, sequential boarding and alighting at all doors (*TnBn* system) is more efficient than operating with boarding at the front door only (*TnB1*). Third, the optimal fare collection technology is off-board with M1 and on-board with a magnetic strip with M2 to M4.

The consideration of a *crowded seating* disutility has a strong effect on the financial state of the public transport provider and the subsidy required to run the system: in M1 and M2 with an optimal fare of 10 cents it is required a subsidy that needs to cover 83 percent of the operator cost, whereas if the optimal fare of M3 is charged (40 cents) the required subsidy is halved<sup>58</sup>. In all cases the toll revenue is more than enough to cover the bus operator deficit (ignoring toll collection costs). Finally, we observe that the predicted modal splits are almost identical under the four models, and that compared to the observed modal split (62.5 percent car, 31.6 percent walk and 5.9 percent), more people decide to walk (33 percent) and ride a bus (7 percent), reducing the car modal split to 60 percent.

In the next subsections, we analyse how the bus service and pricing levels (fare and toll) should be adapted when faced with an increase in transport demand. The trips by origin and destination of Figure 8.3 are uniformly scaled in five steps, up to a total demand of 28,850 trips/h (50 percent higher than the current number of trips). The main results

---

<sup>57</sup> This result ignores that the time to open and close doors may increase with the number of doors, because drivers may spend more time to check that all doors are clear of passengers if more doors are provided in a bus.

<sup>58</sup> The current operation of Sydney buses has a minimum fare of \$2.10 for a single ticket, which in our model would produce profits, however the current system has to be subsidised. This divergence is explained by a number of elements, including the likely existence of a large amount of fixed costs that is not considered in this application, and that we are only modelling the morning peak period in which demand is the highest.



regarding modal split, bus service design, pricing, crowding and congestion are discussed. Model M4 is not shown because its results are similar to those of M3.

### **8.3.3 Optimal bus frequency: The trade-off between congestion and crowding**

The evolution of the optimal bus frequency is presented in Figure 8.4. It is evident that regardless of the demand model considered, frequency does not vary monotonically with demand, in particular optimal frequency can decrease as demand grows, although the reasons for this result are not the same across the models. Focusing on M2 (standing disutility) first, we observe that frequency is increasing up to 32 veh/h for 1,900 pax/h, but drops to 22 pax/h for 2,100 pax/h; this is because up at 1,900 pax/h the optimal bus is mini (8 metres) whereas at 2,100 pax/h it becomes optimal to operate with standard 12 metre buses with a higher capacity. Similarly for M3, a discrete increase in bus size (from 12 to 15 metres) also explains the drop in frequency from 26 to 24 veh/h with 1,550 pax/h. However, if bus size remains unaltered, frequency is always an increasing function of demand if we assume that crowding and standing disutilities matter (case of M2 up to 1,900 pax/h and M3 beyond 1,550 pax/h) which is in line with all total cost minimisation models that optimise bus frequency either assuming a fixed bus size, or that bus size can be freely adjusted to meet demand once frequency has been optimised (e.g., Mohring, 1972; Jansson, 1980; Jara-Díaz and Gschwender, 2003).

What happens with the optimal frequency in the model that is insensitive to crowding as a source of increasing the valuation of time savings (M1) is even more noteworthy. In this case the optimal bus size does not change across the whole demand range (mini buses) and in spite of that, frequency slightly decreases from 21.7 to 20.8 veh/h as demand increases from 1,370 to 2,250 pax/h<sup>59</sup>. This is because of congestion on the road: as total demand grows so does the number of people that use the congestible road facility (the actual speed drop for cars and buses is shown in Figure 8.5), and given that bus frequency

---

<sup>59</sup> This frequency reduction is not necessarily in opposition of traditional bus optimisation models that predict bus frequency to increase with demand, such as Mohring (1972). In the “square root rule”, frequency decreases with the cycle time, which in this case is increasing with demand because of road congestion.

adds to traffic congestion, the model tries to reduce the number of buses on the street at the expense of increasing crowding levels inside buses, which in M1 is welfare improving because crowding comes at no comfort loss.

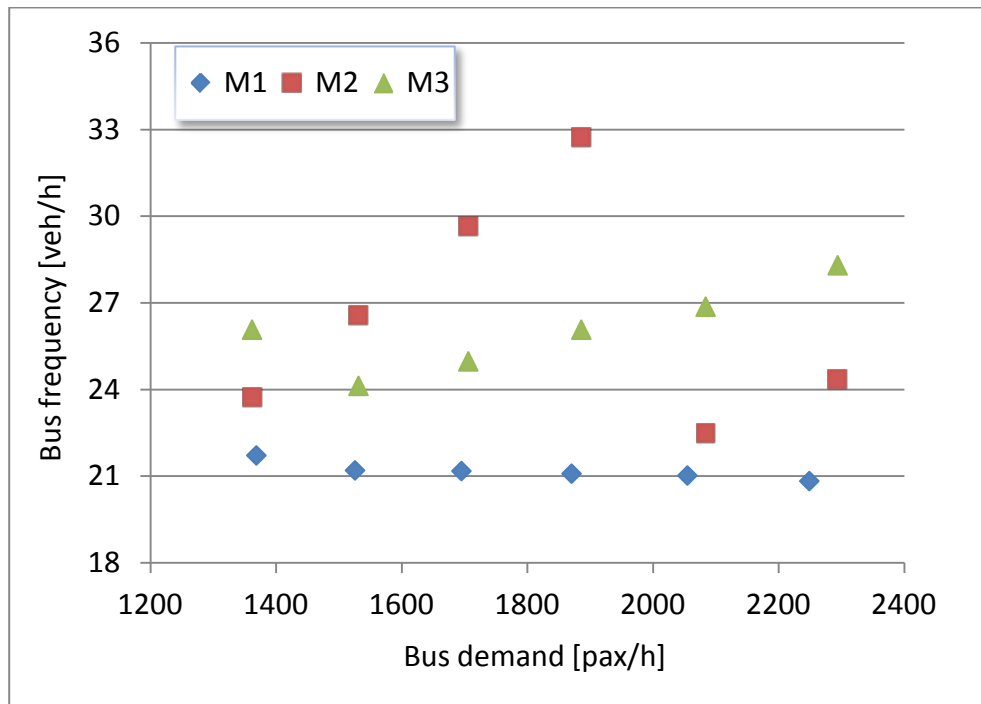


Figure 8.4: Optimal frequency

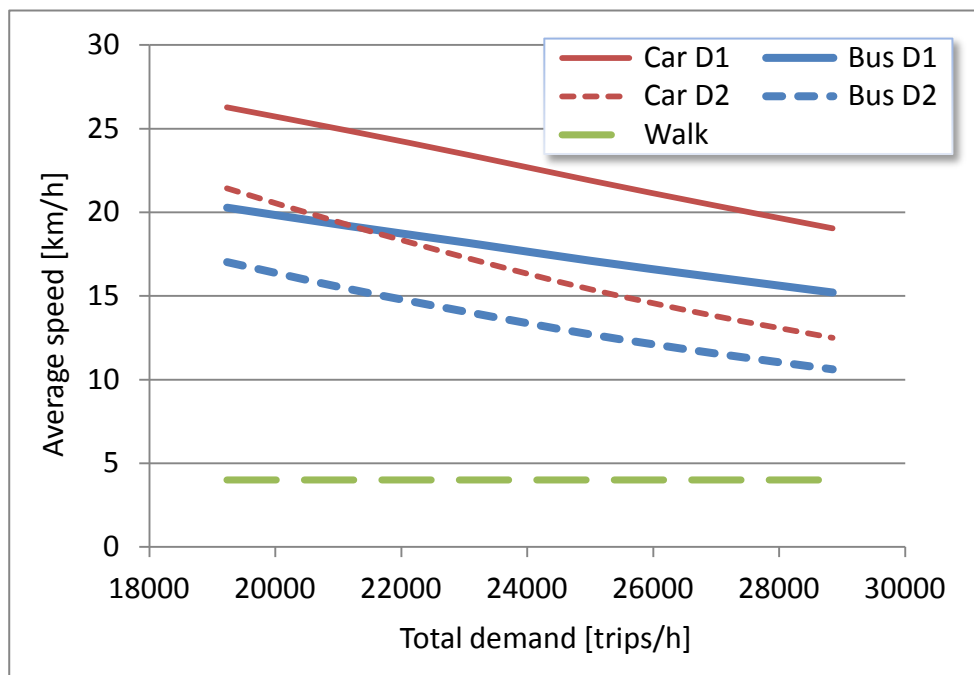


Figure 8.5: Average speed M1

The examination of optimal frequencies does not provide a full picture of the transport supply being provided by the bus operator because different optimal bus sizes are chosen in Figure 8.4. The total seat supply (frequency times number of seats per bus) and seat plus stand supply (frequency times bus capacity) are shown in Figures 8.6 and 8.7. It is clear that the optimal capacity that a planner would choose is quite sensitive to the characterisation of the crowding and standing disutilities. M1 is insensitive to the number of seats chosen as long as the bus capacity constraint (8.18a) is not binding; therefore we have shown the maximum number of seats per bus such that (8.18a) is not active, which passes from 24 seats per bus when demand is 1,370 pax/h, to 11 seats per bus when demand is 2,250 pax/h, as reflected in Figure 8.6 with a total seat supply decreasing for M1. In other words, when confronted with an increase in demand, part of the (optimal) increase in supply is provided simply by reducing the number of seats in order to increase the number of passengers that can be accommodated in a bus, at no crowding cost in M1<sup>60</sup>. A completely different outcome is obtained if crowding and standing matter, in which case the number of seats is kept at the maximum possible given constraint (8.18b) and total seat capacity is increasing for the whole demand range on M2 and M3, as shown in Figure 8.6.

The fact that planners or bus operators would choose to reduce the number of seats per bus if crowding and standing disutilities are not explicitly accounted for (M1) does not mean that the total transport capacity (seat plus stand) is decreasing; as Figure 8.7 shows that with M1, total capacity is actually increasing, due to the increase in bus capacity coupled with a slightly decreasing (almost flat) bus frequency (Figure 8.4).

---

<sup>60</sup> Note that a model in which the number of seats cannot be adjusted would force the frequency and/or bus size to increase if the capacity constraint is binding, which comes at a cost for the operator.

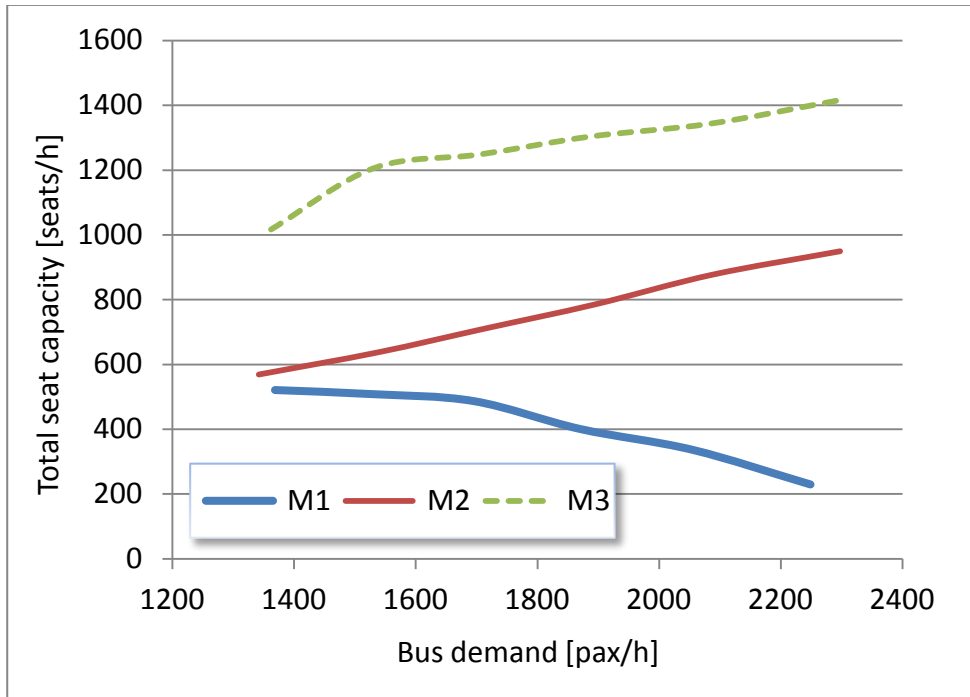


Figure 8.6: Optimal seat supply

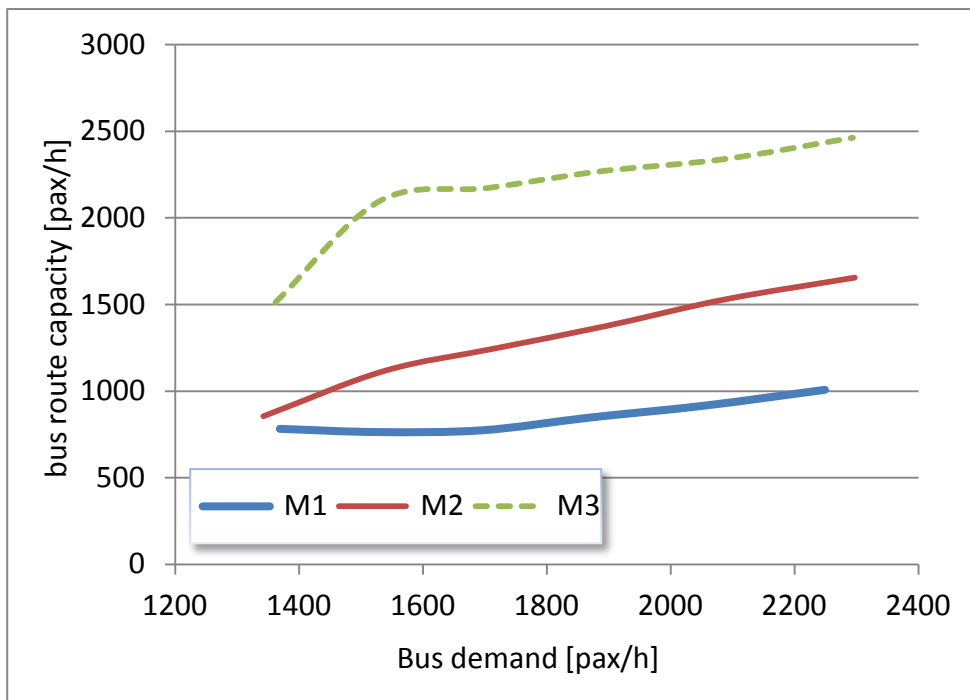


Figure 8.7: Optimal seating+standing supply

In order to show that M1's optimal frequency drop (Figure 8.4) is actually due to the congestion interaction with cars, an alternative scenario in which there is an exclusive bus lane is modelled, and therefore, bus frequency does not influence car travel time, which

is only given by traffic flow (cars remain in two lanes). As depicted in Figure 8.8, when buses do not affect cars optimal frequency increases with demand across the whole demand range.

Finally, it is worth mentioning that in all scenarios, bus frequencies are low enough not to cause any queuing delay at bus stops (equation 8.14), which are assumed to have two berths each. As shown in Chapter 7 (Figure 7.2), for prepaid (cashless) fare collection systems, queuing delays are observable for a bus demand over 3,000 pax/h, a threshold that is not reached in this example.

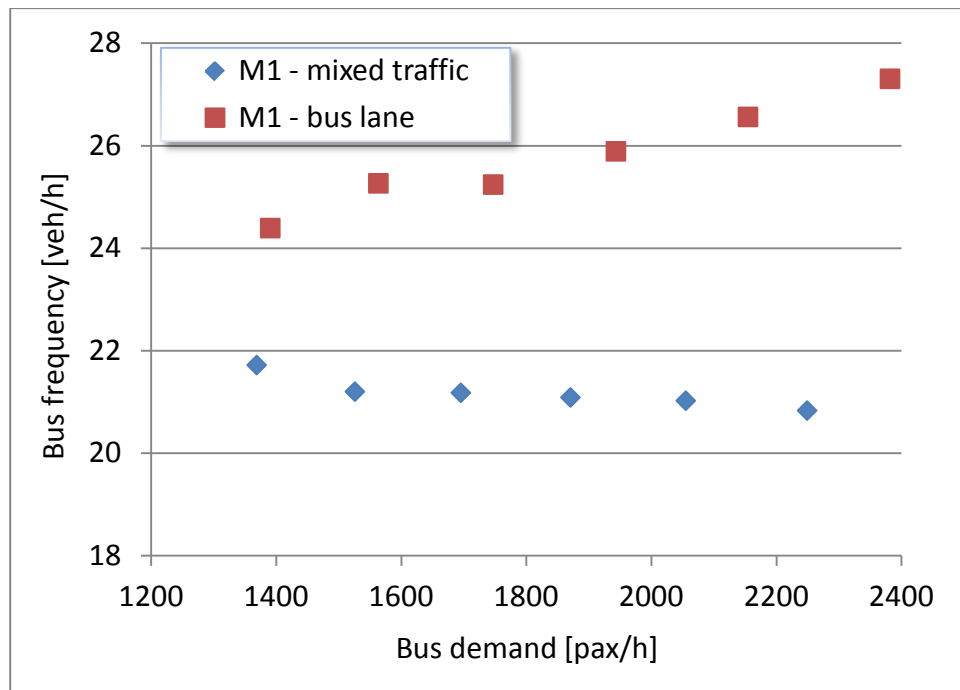
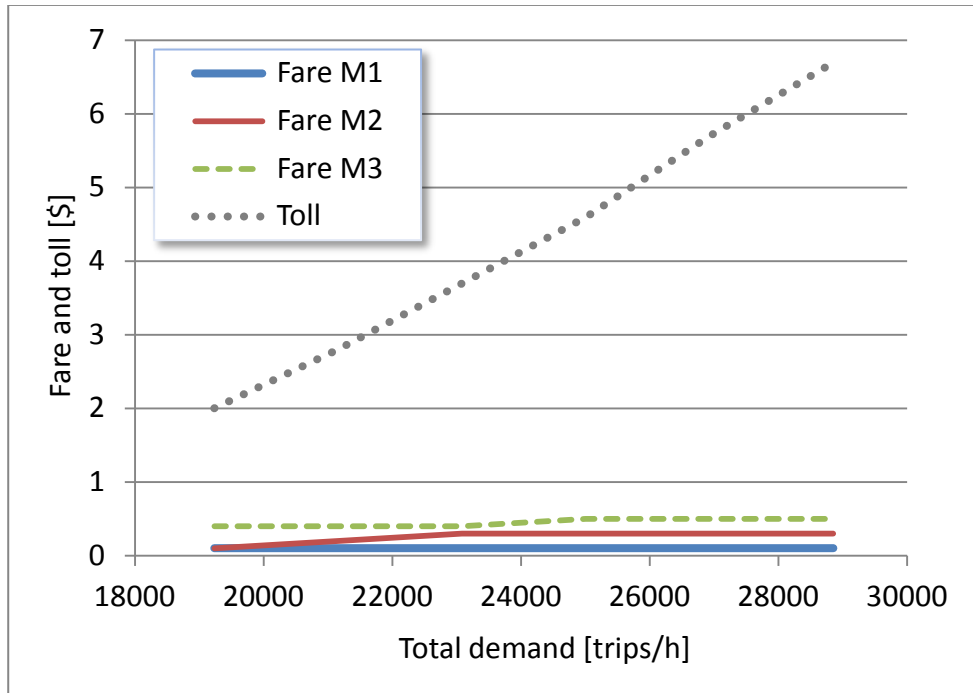


Figure 8.8: Optimal bus frequency on shared and dedicated right-of-way

### 8.3.4 Optimal pricing and modal split

The change in optimal toll and bus fare is analysed in this section. The optimal toll is largely insensitive to the specification of crowding in the bus utility functions, therefore only one value is presented in Figure 8.9, which shows the increase in the optimal toll as total demand (and consequently road congestion) grows. On the other hand, the optimal bus fare slightly increases for M2 and M3, up to 50 cents per ticket, whereas in M1 the fare is maintained at 10 cents.



**Figure 8.9: Optimal toll and bus fare**

Next, modal splits are analysed. The resulting modal split with optimised bus design and pricing structure is almost identical under all demand models (Table 8.1), thus only M2 is shown for illustration in Figure 8.10. The worsening of road congestion<sup>61</sup> as total demand grows encourages walking. This results in a car modal share dropping from 60 to 54 percent, the (assumed uncongestable) alternative of walking increases from 33 to 38 percent, and the bus choice grows from 7 to 8 percent of all trips (due to increased frequency and price difference between toll and fare). Therefore, if transport demand grows in the future and road capacity is held constant, the model predicts walking to become more relevant as a travel alternative, which in this example is supported by the fact that trips are relatively short (the corridor is 3.4 km long). In fact, Figure 8.11 displays the modal split per trip length for trips starting in Zone 1; it is clear that there is a loss of competitiveness of walking as an alternative to motorised modes as trip length increases.

<sup>61</sup> Shown in Figure 8.6 for M1; the result for M2 is similar.

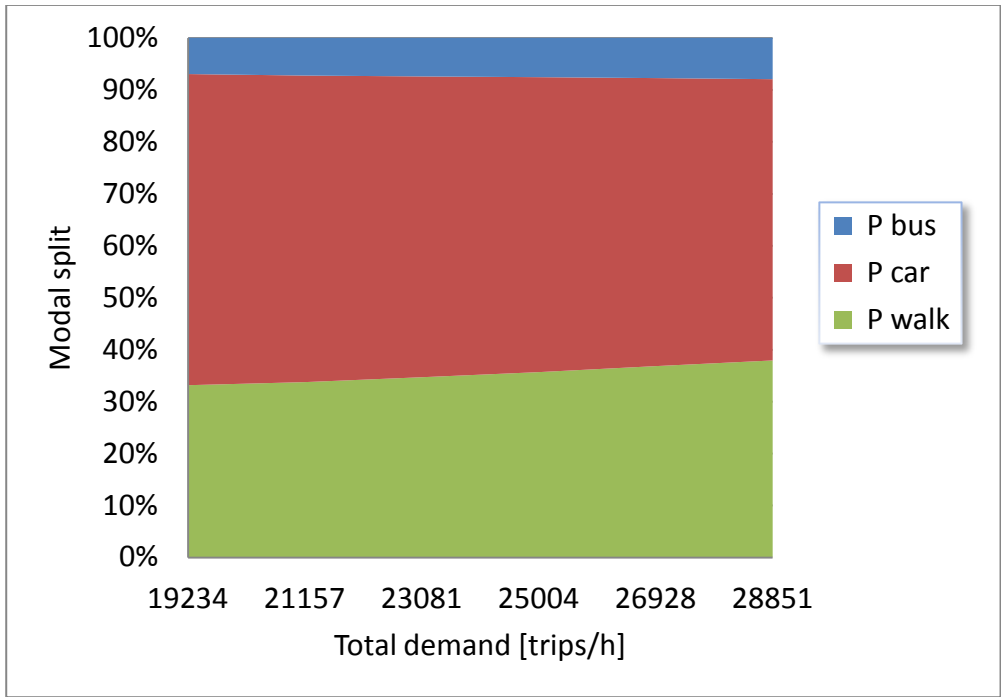


Figure 8.10: Modal split M2

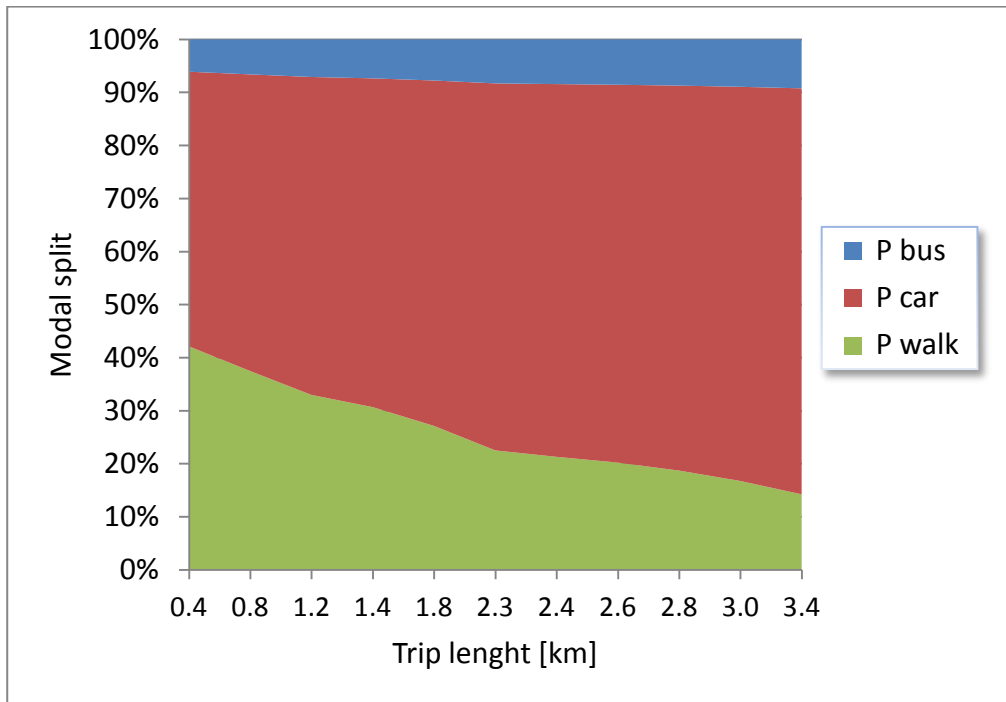


Figure 8.11: Modal split per trip length, M2 base case (total demand=19,234)

### 8.3.5 The case with increased bus-induced congestion

In the previous scenario it was assumed that passenger car equivalency factor for buses  $\varphi(s_b)$  is solely given by bus size, from 1.65 for mini buses (8 metres) to 3 for articulated buses (18 metres). However, as discussed in Chapter 5 on bus congestion, some authors such as Parry and Small (2009) assume that, in mixed traffic, buses should be given a greater weight in the congestion functions (8.10) and (8.11), given that their stops to load and unload passengers have an effect on the capacity of lanes and impose delays on other modes including cars (Koshy and Arasan, 2005; Zhao *et al.*, 2007). We find that when doubling the passenger car equivalency factor (to between 3.3 and 6) optimal bus frequency is reduced, and that the impact is stronger if no crowding externalities are explicitly modelled (M1, Figure 8.12) than when the crowding disutility is accounted for (M2, Figure 8.13). In Figure 8.12, the increase in frequency for a bus demand beyond 1,800 pax/h is because the minimum number of seats  $n_{seat}^{min}$  has been reached, the capacity constraint (8.18a) is binding and therefore the operator has no option but increasing the bus frequency to meet demand.

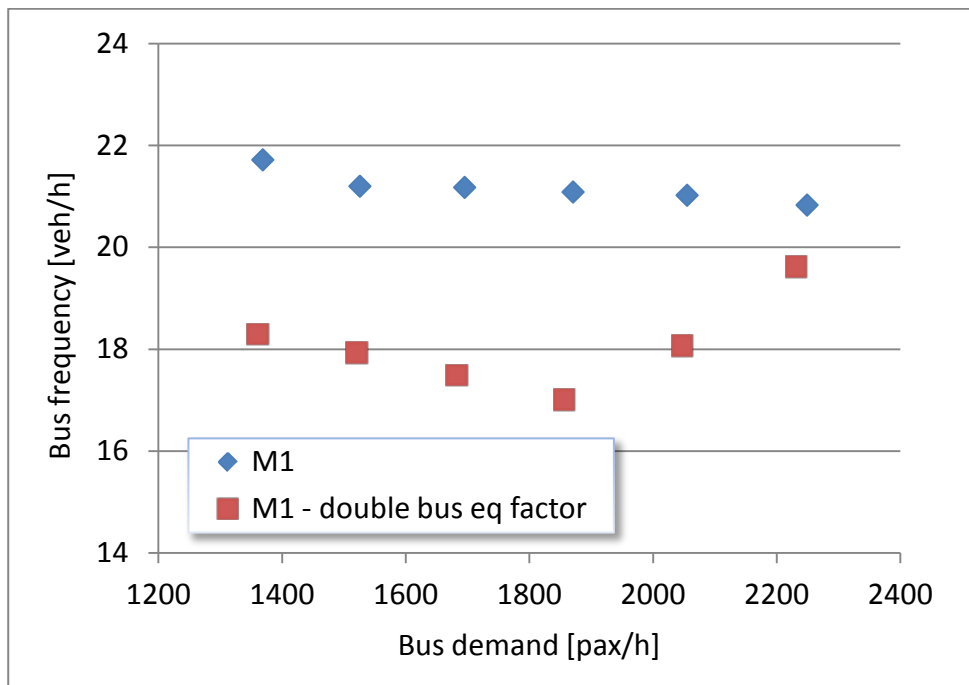


Figure 8.12: Optimal bus frequency M1, double equivalence factor for buses



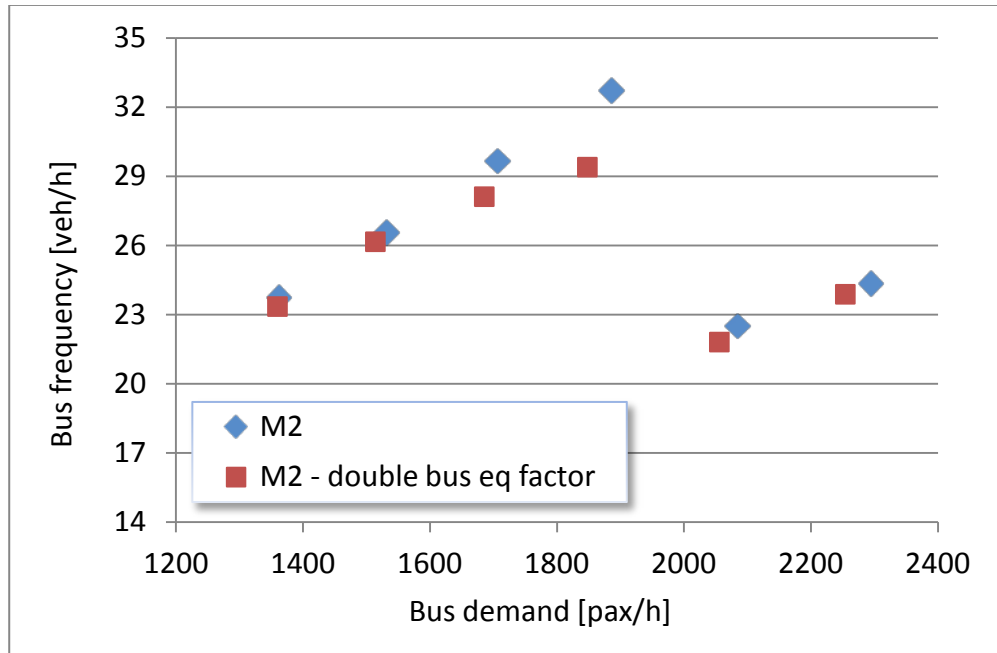
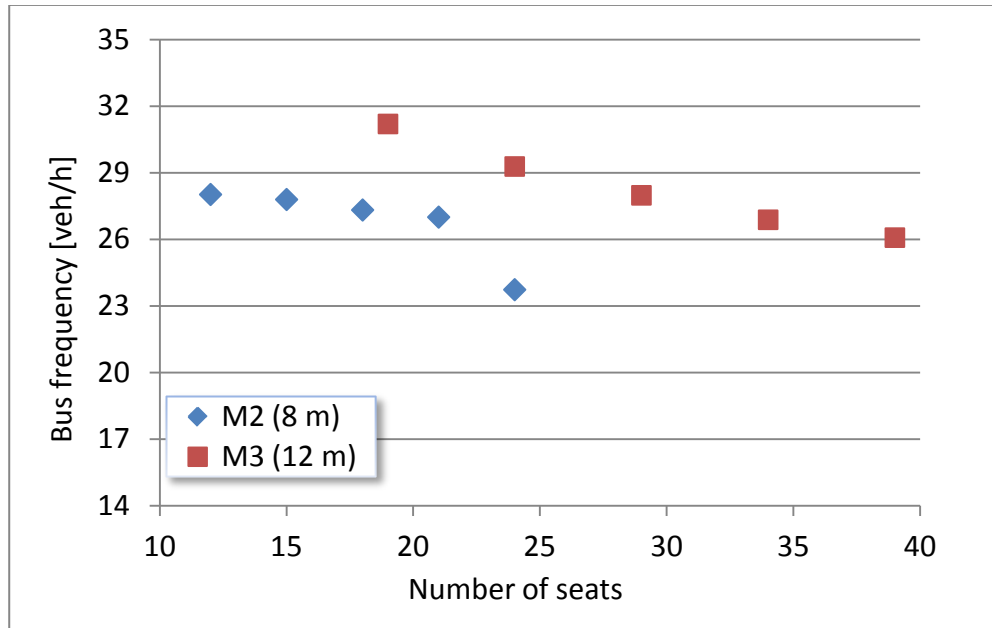


Figure 8.13: Optimal bus frequency M2, double equivalency factor for buses

### 8.3.6 The relationship between the number of seats and optimal frequency

In this section, we study the sensitivity of the optimal bus frequency to alternative bus layouts regarding number of seats. As previously discussed, in all scenarios in which the crowding externality is considered (M2, M3 and M4), the optimal bus design comprises having as many seats as possible, given an optimal bus size, in order to reduce the crowding effects of seating and reduce the number of standees. In this context, we study what happens if the number of seats is exogenously chosen to be lower than the maximum (and therefore the bus capacity is increased); Figure 8.14 shows that for both M2 (mini buses, 8 m.) and M3 (standard buses, 12 m.) frequency should be increased as a response to the users' discomfort of having fewer seats. In other words, the number of seats inside a bus does have an effect on the optimal design of a public transport system if the planner acknowledges that users dislike crowding.



**Figure 8.14: Optimal bus frequency for suboptimal numbers of seat**

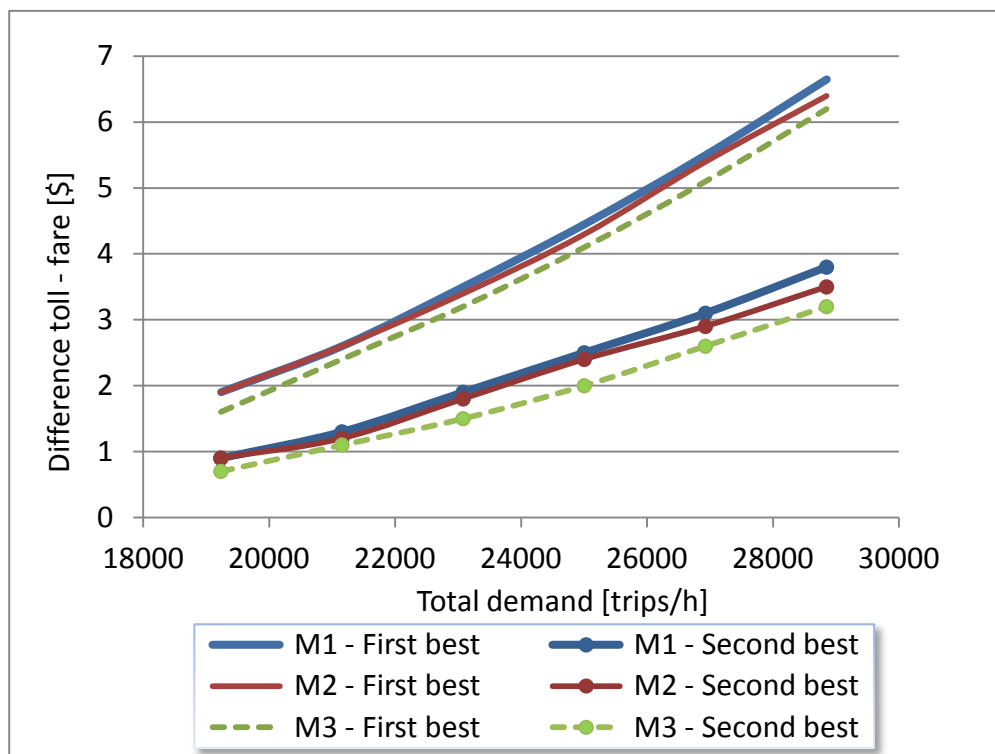
### 8.3.7 The second best scenario

The preceding analysis was undertaken by assuming that a congestion toll on cars is in place, as shown in Figure 8.9. In this section, the second best case in which there is no car toll is investigated. The principles behind first best and second best pricing were extensively discussed in Chapter 3. In this section we limit the analysis to a graphical comparison of relevant optimisation outputs between the first best and second best scenarios.

The second best bus fare is negative across the demand range tested and for all utility specifications (M1 to M4), i.e., the optimal decrease in bus fare to face a zero toll policy is larger than the optimal first best bus fare (between 10 and 40 cents)<sup>62</sup>. Figure 8.15 shows the difference between optimal toll and fare in the first best and second best scenarios for demand models M1, M2 and M3 (therefore, in the second best scenarios the curves are equal to the absolute value of the negative bus fare). The difference between toll and fare is lower in the second best scenario, as also found by Ahn (2009) with a numerical application of a model similar to the one developed in Chapter 3. In Ahn (2009), the

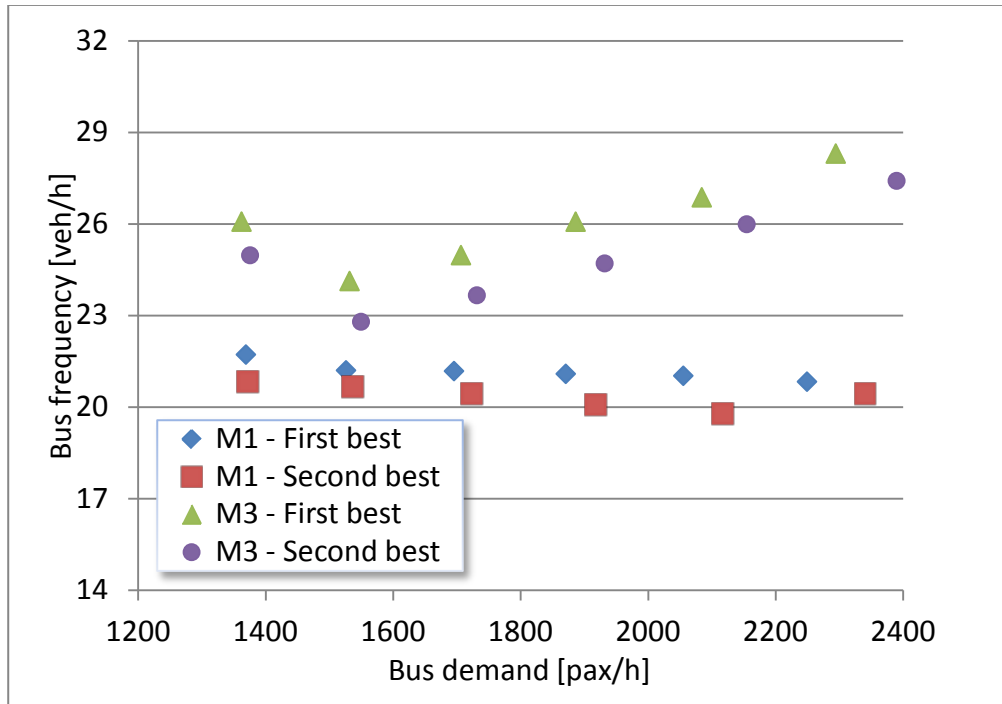
<sup>62</sup> A negative second best bus fare is also obtained by Ahn (2009).

second best bus fare does not decrease sufficiently to maintain the difference between fare and toll in the first best scenario because such a low bus fare would produce a greater than socially optimal amount of total trips; whereas in our framework the amount of total trips is fixed but the amount of motorised trips is not, and hence a low (negative in this case) bus fare attracts not only car users but also walkers to public transport. This explains that the second fare bus fare is not so low as to maintain the first best toll-minus-fare difference.



**Figure 8.15: Optimal toll minus bus fare, first best and second best scenarios**

Finally, Optimal bus frequency is lower in the second best scenario as shown in Figure 8.16 (model M2 is eliminated for easiness of exposition), because the underpricing of car traffic generates a greater than optimal amount of car trips, and therefore increased congestion for both cars and buses.



**Figure 8.16: Optimal bus frequency, first best and second best scenarios**

#### 8.4. Conclusions

In this chapter we have introduced a social welfare maximisation model with disaggregated origin destination demand and multiple travel alternatives, with the aim of optimising the design of urban bus routes including pricing decisions for both bus and car. The influence of bus crowding is highlighted as we analyse its impact on both the design of the bus service and the congestion level on the road. The consideration of crowding externalities as increasing the discomfort of public transport users pushes towards having bigger and more frequent buses (Jara-Díaz and Gschwender, 2003), which in turn may worsen both bus and traffic congestion on shared roads. The number of seats in buses is introduced as a decision variable for the first time in a microeconomic public transport model; the number of seats is the result of the trade-off between passengers' comfort (that drives the number of seats up) and vehicle capacity (which might be increased by removing seats). The model is applied to the Military Road corridor in North Sydney and results are discussed in several scenarios with different demand levels and modelling assumptions.

A number of results stand out from our numerical application. The consideration of crowding externalities (at both seating and standing) imposes a higher optimal bus fare, and consequently, a reduction of the optimal bus subsidy. Optimal bus frequency results from a trade-off between the level of *congestion inside buses*, i.e., passengers' crowding, and the level of *congestion outside buses*, i.e., the effect of frequency on slowing down both buses and cars in mixed-traffic. In particular, optimal bus frequency is quite sensitive to the assumptions regarding crowding costs, the impact of buses on traffic congestion and the overall congestion level. We show that if crowding matters, bus frequency should increase (for a given bus size) with demand even under heavy congestion, however that might not be the case if the crowding externality is not accounted for, in which case an increase of total demand might be met by a decrease of both frequency and number of seats per bus, at the expense of crowding passengers inside buses and making more passengers stand while travelling.

Regarding the relevance of non-motorised modes in urban mobility, in a corridor of 3.4 km, an increase in total transport demand worsens traffic congestion which increases the choice of walking relative to its motorised alternatives (with optimised bus service, fare and toll). This suggests that at least for short trips, improving the travel conditions of non-motorised modes is a wise strategy to tackle worsening congestion problems in cities.

Finally, the existence of a crowding externality implies that buses should have as many seats as possible, up to a minimum area that must be left free of seats. If for any other reason planners decide to have buses with fewer seats than optimal (e.g., to increase bus capacity), bus frequency (and the number of buses itself) should be increased to compensate for discomfort imposed on public transport users. Future research should test the optimality of providing the maximum number of seats in a high demand scenario, in which an active capacity constraint may push the number of seats down.

## Chapter 9

### Conclusions

#### 9.1 Summary

The main topics of this thesis are the design of urban bus systems and the determination of pricing levels in multimodal settings that include cars, public transport and non-motorised transport (walking or cycling) as travel alternatives. Analytical and empirical models for the circulation of buses on a route are developed. Variables analysed comprise bus frequency and size, the number of bus stops, fare collection technique and bus boarding policy, investment in dedicated bus infrastructure, number of bus seats, fare and road price for car users. We analyse the influence of elements such as bus congestion, bus crowding and fare collection techniques and boarding policies in the optimal design of a bus system. A number of engineering aspects are embedded into the microeconomic analysis of urban transport, such as the determination of bus boarding times with alternative fare payment technologies and the estimation of bus queuing delay at bus stops as a source of congestion; the relationship between queuing delays and fare collection and bus boarding times is also explored. Finally, the effect of non-motorised travel alternatives on the pricing decisions of cars and public transport is studied, with the objective of maximising social welfare.

Several results of both scientific and practical relevance have been obtained. These are summarised next.

## **9.2 Contributions**

### **9.2.1 Result on research questions**

Seven research questions were put forward in Chapter 1. The analysis undertaken in this thesis suggests the following main conclusions.

**Question 1:** *What is the effect of including non-motorised transport alternatives in the optimal pricing of motorised modes (public transport and car)?*

A non-motorised alternative (walking and cycling) only has an influence on the second best public transport fare (when there is no marginal cost pricing for cars), and the final result depends on the substitution between motorised and non-motorised modes. We obtain analytically the conditions that lead to the underestimation or overestimation of the optimal public transport fare when non-motorised modes are ignored in the calculation of optimal public transport fares. If the modal substitution between public transport and non-motorised modes is strong relative to the substitution between car and public transport, and between car and non-motorised modes, it is more likely that the optimal public transport fare is underestimated if non-motorised alternatives are ignored.

**Question 2:** *What is the impact of alternative fare collection systems and bus boarding policies on bus travel times and associated costs (e.g., fleet size, operating cost, environmental cost)?*

The effect of alternative fare collection systems and bus boarding policies on bus performance and key operator and service outputs was quantified with an empirical bus travel time model in Chapter 4, and with a theoretical total cost minimisation model in Chapter 7. Sizeable savings on fleet size, operator cost, environmental cost and travel time for users are accruable when speeding up the boarding and alighting process of

passengers, by upgrading the fare collection system from a slow (e.g. on-board cash payment) to a quicker one (e.g., on-board or off-board fare payment verification by means of a contactless card), and/or allowing boarding at all available bus doors.

**Question 3:** *What is the optimal fare collection system and bus boarding rule, given demand and operator cost parameters?*

The optimal fare collection system and bus boarding rule depend on the demand level; as bus patronage increases it is efficient to invest money in speeding up boarding and alighting times. In general, it was found that once on-board cash payment has been ruled out, allowing boarding at all doors is more important as a tool to reduce both users' and operator's costs than technological improvements on fare collection. This analysis is based on the travel times associated with each fare collection technology and the cost of implementing such systems.

**Question 4:** *What is the impact of disregarding the effects of crowding on people's preferences on the design of the optimal road pricing and public transport service and fare levels?*

The consideration of crowding externalities (in respect of both seating and standing) imposes a higher optimal bus fare, and consequently, a reduction of the optimal bus subsidy. Bus crowding has no practical effect on the determination of congestion tolls. Optimal bus frequency results from a trade-off between the level of *congestion inside buses*, i.e., passengers' crowding, and the level of *congestion outside buses*, i.e., the effect of frequency on slowing down both buses and cars in mixed-traffic. We show that if crowding matters, bus frequency increases (for a given bus size) as a function of demand, even under heavy congestion, however that might not be the case if the crowding externality is not accounted for, in which case an increase of total demand might be met by a decrease in both frequency and number of seats per bus, at the expense of crowding passengers inside buses, and making more passengers stand while travelling.



**Question 5:** *Considering both crowding and standing disutilities, how many seats should buses provide?*

The existence of a crowding externality implies that buses should have as many seats as possible, up to a minimum area that must be left free of seats. If for any other reason planners decide to have buses with fewer seats than optimal (e.g., to increase bus capacity), bus frequency (and the number of buses itself) should be increased to compensate for discomfort imposed on public transport users.

**Question 6:** *What is the effect on the design of bus systems of misrepresenting bus congestion (or not considering it at all) for scenarios with high bus demand (and which are consequently, subject to bus congestion).*

Disregarding bus congestion at bus stops in the design of the service would yield greater frequencies than optimal when congestion is noticeable, i.e. for high demand (over 10,000 pax/h in the example of Chapter 7). In congested mixed-traffic operation, optimal bus frequency depends on the marginal effect of an extra bus on overall congestion (for both buses and cars); if detentions in bus stops are captured through an increased value of the car equivalency factor of a bus, then optimal bus frequency decreases.

**Question 7:** *If bus speed can be increased by investment in infrastructure in dedicated bus corridors, what is the optimal level of investment in busways (which in turn determines the running speed of buses)?*

For the off-board payment system and the on-board payment alternatives before congestion builds, the optimal investment in infrastructure grows with the logarithm of demand. This result depends on the existence of a positive and linear relationship between investment in infrastructure and desired running speed. If additional bus stops are added to a high demand area with the objective of reducing the number of passengers per stop, and consequently the duration of queuing delays, then bus running speed should be decreased.

### **9.2.2 Methodological contributions**

The second-best public transport fare with explicit account of underpriced cars and an uncongestible non-motorised mode as travel alternatives is analytically derived. We obtain the conditions that lead to the underestimation or overestimation of the optimal public transport fare when non-motorised transport is ignored in the determination of optimal prices for motorised transport.

The decision by the operator of a fare collection system and bus boarding policy is embedded into microeconomic models for the optimisation of bus systems. Faster bus boarding techniques (e.g., upgrading from of-board cash payment to off-board contactless card payment validation) present the trade-off of reducing ride time and increasing capital cost.

Bus congestion in the form of queuing delays behind bus stops is estimated using simulation. The delay function depends on the bus frequency, bus size, number of berths and dwell time (which is given by the number of passengers boarding and alighting, the number of doors per bus and the fare collection technology). Therefore, we conclude that models that use flow measures (including frequency only or frequency plus traffic flow) as the only explanatory variables for bus congestion are incomplete.

The use of bus running speed as an optimisation variable, linked to the investment in infrastructure, is also a novelty in the literature on urban bus transport optimisation, which has traditionally focused on determining the optimal value for bus frequency, size, distance between stops, density of routes and fare.

The existence of a crowding externality and standing disutility for passengers inside buses is used to determine the optimal internal layout of buses in respect of spacing allocated to seating and standing. Explicit constraints are considered for the determination of seating and standing areas, including areas free of seats next to the bus driver, doors, for a central aisle and for a wheelchair.

### **9.2.3 Further results of practical relevance**

An empirical model on bus running times with and without the influence of traffic congestion is used to compare two policies aimed at reducing bus travel times: providing bus lanes versus upgrading the fare collection system. It was found that the bus demand level in terms of passengers per bus-kilometre is crucial in determining the superiority of one policy or the other. A demand threshold is identified beyond which speeding up the boarding and alighting process is more effective in increasing bus operating speed than segregating cars from buses. The demand threshold depends on the congestion level associated with mixed-traffic operation.

Using empirical data on the actual number of bus stops from an on-call bus service (i.e., in which buses are allowed to skip bus stops if no one desires to get off or on) in Sydney, it was possible to estimate a relationship between the actual number of stops, the scheduled (designated) number of stops, and the total number of passengers riding a bus. This function is useful to estimate the number of stops per bus ride given the number of passengers boarding and alighting, and consequently total ride time in on-call services.

The bus congestion function estimated with the bus stop simulator IRENE is helpful to analyse the influence of several factors on the occurrence of queuing delays at bus stops, such as the number of passengers boarding and alighting, the fare collection technology, the number of doors to board and alight, the bus frequency and size and the number of berths.

### **9.3 Caveats of the Research**

The empirical and theoretical analysis developed in this thesis adds to the literature in several respects; however there are a number of caveats or limitations that have to be highlighted in order to understand the scope of the contributions and to provide a research path for ongoing inquiry.

In terms of geographical or spatial scope, the empirical analysis in this thesis has been limited to a single corridor, either for the optimisation of bus services only (i.e., a single route in Chapter 7) or for the multimodal social welfare maximisation model that includes cars and non-motorised modes (Chapter 8). Transport networks are not treated and would be relevant, for example, for the adoption of a new fare collection system (in which the analysis of a network may suggest the convenience of applying off-board fare collection in only a subset of bus stops or bus routes with high demand).

The empirical estimation of bus travel times in Chapter 4 considers several time periods; however, the total cost minimisation and social welfare maximisation models are developed in a single period framework, which is assumed to be the morning peak in the numerical applications. In this respect, we assume that travellers can choose a mode but cannot switch period of travel, which is relevant if a percentage of the peak transport demand can be actually spread to off-peak periods when faced with, for example, the application of a time-of-day congestion toll. In Chapter 8, changes in origin and/or destination, and the option of not travelling at all were not considered either.

The focus of the thesis is on adding to the knowledge base in the field of public transport economics, and although private transport is taken into account, the policy tools considered for car users are limited to the application of a fixed road user price. Other policies such as optimal parking fees or dynamic tolls are disregarded. The latter have an influence on aggregate social welfare and resulting modal split; however it is considered unlikely that a more refined modelling of the car alternative would significantly affect the main results regarding the optimisation of bus systems, in terms of crowding effects on service design and pricing, bus congestion, and boarding and alighting policy.

On the subject of bus operations, the models assume that buses maintain a regular headway, i.e., the issue of bus bunching is ignored. The optimisation of bus timetables is not addressed either, which might be important when translating the potential time benefits of, for example, upgrading the fare collection system from a slow to a quicker technique. The physical and fare integration between buses and other public transport modes such as rail are also ignored. For example, a contactless smartcard fare collection

system would provide further benefits if is implemented in a multimodal public transport network, in which transfers are charged at a different rate (or are for free); in such a setting it is more cumbersome to have paper- or cash-based payment systems. Smartcards can also be used to pay for both transport and non-transport services.

Implementation issues have not been embedded in our analysis. These include the existence of fare evasion and the fact that advanced smartcard fare collection systems may be difficult to implement due to technological or contractual obstacles (e.g., delays on approval by various authorities), as reported by Hidalgo and Carrigan (2010).

The economic analysis considers the transport sector only, abstracting from distortions in the rest of the economy and general equilibrium issues (reviewed in Section 2.3.4). Distributional concerns were also ignored, which emphasise the need for a general equilibrium approach to fully gauge the distributional and welfare effects of transport pricing reforms.

#### **9.4 Areas for Further Research**

The discussion on the caveats imposed on the research focus of this thesis suggests several lines of future research inquiry that are natural extensions of the contribution of this thesis. Some of these extensions are summarised in this section.

The examination of a public transport network with multiple routes and modes (e.g., bus, light rail and heavy rail) is a natural extension of this work. Simple networks such as those proposed by Chang and Schonfeld (1991) or Jara-Díaz and Gschwender (2003b) could be used as a first step to determine the sensitivity of the new results of this thesis on the design of a single bus route, to elements particular to public transport networks like the existence of two- or three-legged trips with transfers, or to the possibility of passengers choosing between two or more public transport alternatives to complete a trip. Looking beyond this immediate extension, the ultimate goal would be the optimisation of a real-world public transport network, including decisions on network design, number and size of bus stops (given by the number of berths), fare collection system, bus size and

frequency, number of doors and number of seats per bus, and cost of tickets for different pricing structures (with and without fare integration across modes). The optimal choice of rail or bus based services for a particular route can also be considered.

The effect of crowding on bus service design was only partially incorporated as an influence on increasing the valuation of in-vehicle time savings. As reviewed in Chapter 6, a more comprehensive account of the crowding externality needs to also consider the effect of crowding on the valuation of waiting times (if bus or train stations get crowded) and on increasing waiting and boarding and alighting times. All these considerations are expected to increase the optimal bus frequency to reduce crowding levels in vehicles and at stations.

Regarding the relationship between non-motorised modes and the optimal pricing of cars and public transport, an application of the model of Chapter 3 could include safety issues, as empirical findings suggest a *safety in numbers* effect for non-motorised travel; that is, the probability of getting involved in an accident by a person walking and cycling decreases the more people are walking and cycling (Jacobsen, 2003; Robinson, 2005).

Travel time variability and modal reliability are known to have a significant role in influencing the quality of service and hence demand (Hensher and Prioni, 2002; Hensher *et al.*, 2003); however travel time variability and modal reliability have been disregarded in this study. The relationship between service reliability and public transport supply and pricing decisions should be incorporated in future research efforts.

Finally, another area worthy of further research is building in preference heterogeneity in modal choice, which can be embedded in a social welfare maximisation framework through, for example, the specification of a mixed multinomial logit model (MMNL). It is not known to what extent outputs on the optimisation of public transport systems are sensitive to the choice model assumed to explain travel behaviour.

## 9.5 Concluding Remarks

This thesis presents, on the one hand, a number of extensions and new elements in the analysis of urban bus operations, and on the other hand, a number of contributions designed to integrate engineering and economic aspects, as a way of improving our understanding of economic and operational measures that can be implemented to increase social welfare in the passenger transport sector. Contributions of both scientific and practical significance have been presented and substantiated with analytical and empirical methods.

First, we have shown how sensitive the optimal design of urban bus services is to the consideration of elements such as bus congestion and passenger crowding. Second, we have set out and implemented methods for the study of other interventions in the public transport system, including decisions on bus running speed and investment in road infrastructure for buses, and the choice of a fare collection system and bus boarding and alighting rules. Finally, the role of walking and cycling in the setting of optimal prices for motorised transport was revealed. The generally agreed proposition that public and non-motorised transport are fundamental for the development of sustainable transport policies and sustainable cities suggests that it is vital to assess the relevance of these findings. This thesis is a step forward towards a more comprehensive treatment of sustainable transport alternatives in formal policy analysis.

## References

Abkowitz, M. D. and Engelstein, I., 1983. Factors affecting running time on transit routes. *Transportation Research Part A* 17(2), 107-113.

Abkowitz, M. and Tozzi, J., 1987. Research contributions to managing transit service reliability. *Journal of Advanced Transportation* 21(1), 47-65.

Ahn, K., 2009. Road pricing and bus service policies. *Journal of Transport Economics and Policy* 43(1), 25-53.

Akçelik, R., 1981. Traffic signals-capacity and timing analysis. Research Report ARR 123. Australian Road Research Board. Melbourne. Australia.

Akçelik, R. and Roupail, N. M., 1993. Estimation of delays at traffic signals for variable demand conditions. *Transportation Research Part B* 27(2), 109-131.

Akçelik, R. and Besley, M., 2002. Queue discharge flow and speed models for signalized intersections. 15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia.

Allport, R. J., 1981. The costing of bus, light rail transit and metro public transport systems. *Traffic Engineering and Control* 22, 633-639.

Arnott, R., de Palma, A. and Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *The American Economic Review* 83(1), 161-179.

Arnott, R. and Yan, A., 2000. The two-mode problem: second-best pricing and capacity. *Review of Urban & Regional Development Studies* 12(3), 170-199.

ATC, 2006. National Guidelines for Transport System Management in Australia – Volume 4 Urban Transport. Australian Transport Council.

Balcombe, R., Mackett, R., Paulley, N., J., P., Shires, J., Titheridge, H., Wardman, M. and White, P., 2004. The demand for public transport: a practical guide. Transportation Research Laboratory Report TRL 593, London, UK.

Basso, L. J. and Jara-Díaz, S. R., 2010. The case for subsidisation of urban public transport and the Mohring effect. *Journal of Transport Economics and Policy* 44, 365-372.



Basso, L. J. and Silva, H. E., 2010. A microeconomic analysis of congestion management policies. 5th Kuhmo Nectar Conference in Transport Economics Valencia, Spain, July 8-9.

Bates, J., Polak, J., Jones, P. and Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research Part E* 37(2-3), 191-229.

Berglas, E., Fresko, D. and Pines, D., 1984. Right of way and congestion toll. *Journal of Transport Economics and Policy* 18(2), 165-187

Bertini, R. L. and El-Geneidy, A. M., 2004. Modeling transit trip time using archived bus dispatch system data. *Journal of Transportation Engineering* 130(1), 56-67.

Bertrand, T. J., 1977. 'Second best' congestion taxes in transportation systems. *Econometrica* 45(7), 1703-1715.

Bhat, C. R. and Sardesai, R., 2006. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B* 40(9), 709-730.

BITRE, 2007. Estimating urban traffic and congestion cost trends for Australian cities. Working Paper 71, Bureau of Transport and Regional Economics, Department of Transport and Regional Services, Canberra, Australia.

Bly, P. H., Webster, F. V. and Pounds, S., 1980. Effects of subsidies on urban public transport. *Transportation* 9(4), 311-331.

Browning, E. K., 1976. The marginal cost of public funds. *The Journal of Political Economy* 84(2), 283-298.

Calthrop, E., De Borger, B. and Proost, S., 2010. Cost-benefit analysis of transport investments in distorted economies. *Transportation Research Part B* 44(7), 850-869.

Ceder, A., Golany, B. and Tal, O., 2001. Creating bus timetables with maximal synchronization. *Transportation Research Part A* 35(10), 913-928.

Cepeda, M., Cominetti, R. and Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation Research Part B* 40(6), 437-459.

Cervero, R., 1984. Cost and performance impacts of transit subsidy programs. *Transportation Research Part A* 18(5-6), 407-413.

Chang, S. K. and Schonfeld, P. M., 1991. Multiple period optimization of bus transit systems. *Transportation Research Part B* 25(6), 453-478.

Chien, S. and Schonfeld, P., 1998. Joint optimization of a rail transit line and its feeder bus system. *Journal of Advanced Transportation* 32(3), 253-284.

Chien, S. I. J. Y. and Tsai, C. F. M., 2007. Optimization of fare structure and service frequency for maximum profitability of transit systems. *Transportation Planning and Technology* 30(5), 477 - 500.

Cominetti, R. and Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. *Transportation Science* 35(3), 250-267.

Cortés, C. E., Jara-Díaz, S. and Tirachini, A., 2011. Integrating short turning and deadheading in the optimization of transit services. *Transportation Research Part A* 45(5), 419-434

Daganzo, C. F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B* 43(10), 913-921.

Danielis, R. and Marcucci, E., 2002. Bottleneck road congestion pricing with a competing railroad service. *Transportation Research Part E* 38(5), 379-388.

Davidson, B., Vovsha, P., Abedini, M., Chu, C. and Garland, R., 2011. Impact of capacity, crowding, and vehicle arrival adherence on public transport ridership: Los Angeles and Sydney experience and forecasting approach. *Proceedings of the Australasian Transport Research Forum, Adelaide, Australia, 28-30 September 2011.*

De Borger, B., Mayeres, I., Proost, S. and Wouters, S., 1996. Optimal pricing of urban passenger transport: a simulation exercise for Belgium. *Journal of Transport Economics and Policy* 30(1), 31-54.

De Borger, B. and Wouters, S., 1998. Transport externalities and optimal pricing and supply decisions in urban transportation: a simulation analysis for Belgium. *Regional Science and Urban Economics* 28(2), 163-197.

De Borger, B. and Swysen, D., 1999. Public transport subsidies versus road pricing: an empirical analysis for interregional transport in Belgium. Working Paper 1999006, Faculty of Applied Economics, University of Antwerp.

Delgado, F., Muñoz, J., Giesen, R. and Cipriano, A., 2009. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. *Transportation Research Record* 2090, 59-67.

Delle Site, P. and Filippi, F., 1998. Service optimization for bus corridors with short-turn strategies and variable vehicle size. *Transportation Research Part A* 32(1), 19-38.

dell'Olio, L., Moura, J. L. and Ibeas, A., 2006. Bi-level mathematical programming model for Locating bus stops and optimizing frequencies. *Transportation Research Record* 1971, 23-31.

Department of Transport, 1989. Central London Rail Study. A joint study by the Department of Transport, British Rail Network South East, London Regional Transport and London Underground Ltd.

Dodgson, J. S. and Topham, N., 1987. Benefit-cost rules for urban transit subsidies: an integration of allocational, distributional and public finance issues. *Journal of Transport Economics and Policy* 21(1), 57-71.

Douglas, N. and Karpouzis, G., 2005. Estimating the cost to passengers of station crowding. 28th Australasian Transport Research Forum (ATRF), Sydney, September 2005.

Dueker, K. J., Kimpel, T. J. and Strathman, J. G., 2004. Determinants of bus dwell time. *Journal of Public Transportation* 7(1), 21-39.

Dupuit, J., 1844. De la Mesure de l'Utilité des Travaux Publiques. In *Annales des Ponts et Chaussées* Vol. 8. Translated and reprinted in: K. Arrow & T. Scitovski (Eds.), 1969. *AEA Readings in Welfare Economics*, AEA, 255-283.

El-Geneidy, A., J. Horning and Krizek, K., 2008. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. 87th Annual Meeting of the Transportation Research Board, Washington, D.C.

Else, P. K., 1985. Optimal pricing and subsidies for scheduled transport services. *Journal of Transport Economics and Policy* 19(3), 263-279.

Frankena, M. W., 1983. The efficiency of public transport objectives and subsidy formulas. *Journal of Transport Economics and Policy* 17(1), 67-76.

Fernández, J. E., de Cea, J. and de Grange, L., 2005. Production costs, congestion, scope and scale economies in urban bus transportation corridors. *Transportation Research Part A* 39(5), 383-403.

Fernández, R., 2010. Modelling public transport stops by microscopic simulation. *Transportation Research Part C* 18(6), 856-868.

Fernández, R., Valenzuela, E. and Gálvez, T., 2000. Incorporación de la capacidad y rendimiento de paraderos en el programa TRANSYT ("Incorporation of capacity and performance of bus stops in the program TRANSYT", in Spanish). Proceedings of the XI Panamerican Conference of Transport and Traffic Engineering, Gramado, Brasil.

Fernández, R. and Planzer, R., 2002. On the capacity of bus transit systems. *Transport Reviews* 22(3), 267–293.

Fernández, R. and Tyler, N., 2005. Effect of passenger-bus-traffic interactions on bus stop operations. *Transportation Planning and Technology* 28(4), 273-292.

Fernández, R., del Campo, M. and Swett, C., 2009. Data collection and calibration of passenger service time models for the Transantiago system. European Transport Conference, The Netherlands, 6-8 October.

Fernández, R., Zegers, P., Weber, G. and Tyler, N., 2010. Influence of platform height, door width, and fare collection on bus dwell time: laboratory evidence for Santiago de Chile. *Transportation Research Record* 2143, 59-66.

Fickling, R., Gunn, H., Kirby, H., Bradley, M. and Heywood, C., 2008. The productive use of rail travel time and value of travel time saving for travellers in the course of work. European Transport Conference.

Fleurent, C., Lessard, R. and Seguin, L., 2004. Transit timetable synchronization: evaluation and optimization. Proceedings of the Ninth International Conference on Computer-Aided Scheduling of Public Transport (CASPT), San Diego, California.

Florian, M., He, S. and Constantin, I., 2005. An EMME/2 macro for transit equilibrium assignment which satisfies capacity of transit services. EMME/2 Conference, Shanghai.

Frey, H. C., Roupail, N. M., Zhai, H., Farias, T. L. and Gonçalves, G. A., 2007. Comparing real-world fuel consumption for diesel- and hydrogen-fueled transit buses and implication for emissions. *Transportation Research Part D* 12(4), 281-291.

FTA, 2008. Advanced Network Planning for Bus Rapid Transit: The “Quickway” Model as a Modal Alternative to “Light Rail Lite”. Report No. FL-26-7104-4, Federal Transit Administration, U.S. Department of Transportation.

FTA, 2009. Characteristics of Bus Rapid Transit for Decision-Making. Project No. FTA-FL-26-7109.2009.1, Federal Transit Administration, U.S. Department of Transportation.

Furth, P. and Rahbee, A., 2000. Optimal bus stop spacing through dynamic programming and geographic modeling. *Transportation Research Record: Journal of the Transportation Research Board* 1731, 15-22.

Furth, P. G. and SanClemente, J. L., 2006. Near side, far side, uphill, downhill: impact of bus stop location on bus delay. *Transportation Research Record* 1971, 66-73.

Gatta, V. and Marcucci, E., 2007. Quality and public transport service contracts. *European Transport \ Trasporti Europei* 36, 92-106.

Gibson, J., Baeza, I. and Willumsen, L. G., 1989. Bus-stops, congestion and congested bus-stops. *Traffic Engineering and Control* 30(6), 291-302.

Gibson, J., Fernández, R. and Albert, A., 1997. Operación de paraderos formales en Santiago ("Operation of formal bus stops in Santiago", in Spanish). *Proceedings VIII Chilean Conference of Transport Engineering*

Glaister, S., 1974. Generalised consumer surplus and public transport pricing. *The Economic Journal* 84(336), 849-867.

Glaister, S. and Lewis, D., 1978. An integrated fares policy for transport in London. *Journal of Public Economics* 9(3), 341-355.

Hamdouch, Y., Ho, H. W., Sumalee, A. and Wang, G., 2011. Schedule-based transit assignment model with vehicle capacity and seat availability. *Transportation Research Part B* 45(10), 1805-1830.

Hensher, D. A., 2002. A systematic assessment of the environmental impacts of transport policy: An end use perspective. *Environmental and Resource Economics* 22(1-2), 185-217.

Hensher, D. A., 2003. Influences on distance-based unit costs for non-urban bus operations in New South Wales. *Road & Transport Research* 12(1), 16-27.

Hensher, D., 2010. Knowing your costs. Notes for the Certificate of Transport Management, Institute of Transport and Logistics Studies, The University of Sydney.

Hensher, D. A. and Brewer, A. M., 2000. Transport Economics and Management, Oxford University Press, Oxford.

Hensher, D. and Prioni, P., 2002. A service quality index for area-wide contract performance assessment. *Journal of Transport Economics and Policy* 36(1), 93-113.

Hensher, D. A. and Stanley, J., 2003. Performance-based quality contracts in bus service provision. *Transportation Research Part A* 37(6), 519-538.

Hensher, D. A., Stopher, P. R. and Bullock, P., 2003. Service quality—developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A* 37(6), 499-517.

Hensher, D. A. and Houghton, E., 2004. Performance based quality contracts for the bus sector: Delivering social and commercial value for money. *Transportation Research Part B* 38(2), 123-146.

Hensher, D. A. and Wallis, I., 2005. Competitive tendering as a contracting mechanism for subsidising transportation: The bus experience. *Journal of Transport Economics and Policy* 39(3), 295-322.

Hensher, D. A., Rose, J. M. and Collins, A., 2011. Identifying commuter preferences for existing modes and a proposed Metro in Sydney, Australia with special reference to crowding. *Public Transport* 3(2), 109-147.

Hensher, D. A., Truong, T. P., Mulley, C. and Ellison, R., 2012. Assessing the wider economy impacts of transport infrastructure investment with an illustrative application to the North-West Rail Link project in Sydney, Australia. Forthcoming in the *Journal of Transport Geography*.

Hidalgo, D. and Carrigan, A., 2010. Modernizing Public Transportation. Report Embarq, World Resources Institute.

Hossain, M. and Kennedy, S., 2008. Estimating energy savings from bus improvement options in urban corridors. *Journal of Public Transportation* 11(3): 19-40.

Huang, H.-J., 2002. Pricing and logit-based mode choice models of a transit and highway system with elastic demand. *European Journal of Operational Research* 140(3), 562-570.

Ibeas, Á., dell'Olio, L., Alonso, B. and Sainz, O., 2010. Optimizing bus stop spacing in urban areas. *Transportation Research Part E* 46(3), 446-458.

Jacobsen, P. L., 2003. Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention* 9, 205-209.

Jaiswal, S., Bunker, J. and Ferreira, L., 2007. Operating characteristics and performance of a busway transit station. 30th Australasian Transport Research Forum, Melbourne, 25-27 September.

Jaiswal, S., Bunker, J. and Ferreira, L., 2010. Influence of platform walking on BRT station bus dwell time estimation: Australian analysis. *Journal of Transportation Engineering* 136(12), 1173-1179.

Jansson, J. O., 1979. Marginal cost pricing of scheduled transport services. *Journal of Transport Economics and Policy* 13(3), 268-294.

Jansson, J. O., 1980. A simple bus line model for optimization of service frequency and bus size. *Journal of Transport Economics and Policy* 14(1), 53-80.

Jansson, K., 1993. Optimal public transport price and service frequency. *Journal of Transport Economics and Policy* 27(1), 33-50.

Jansson, K., 2010. Public transport policy with and without road pricing 5th Kuhmo-Nectar Conference on Transport Economics, Valencia, Spain, July 8-9.

Jara-Díaz, S. R., 1990. Consumer's surplus and the value of travel time savings. *Transportation Research Part B* 24(1), 73-77.

Jara-Díaz, S. R., 2007. *Transport Economic Theory*, Elsevier. Linacre House, Jordan Hill, Oxford OX2 8DP, UK.

Jara-Díaz, S. R. and Videla, J., 1989. Detection of income effect in mode choice: Theory and application. *Transportation Research Part B* 23(6), 393-400.

Jara-Díaz, S. R. and Gschwender, A., 2003a. Towards a general microeconomic model for the operation of public transport. *Transport Reviews* 23(4), 453 - 469.

Jara-Díaz, S. R. and Gschwender, A., 2003b. From the single line model to the spatial structure of transit services: corridors or direct? . *Journal of Transport Economics and Policy* 37(2), 261-277.

Jara-Díaz, S. R. and Gschwender, A., 2005. Making pricing work in public transport provision. In: Button, K.J., Hensher, D.A. (eds.) *Handbook of Transport Strategy, Policy and Institutions*, Elsevier Ltd, Amsterdam: 447-459.

Jara-Díaz, S. and Tirachini, A., 2012. Urban bus transport: open all doors for boarding. Forthcoming in the *Journal of Transport Economics and Policy*.

Jovicic, G. and Hansen, C. O., 2003. A passenger travel demand model for Copenhagen. *Transportation Research Part A* 37(4), 333-349.

Kerin, P. D., 1992. Efficient bus fares. *Transport Reviews* 12(1), 33 - 47.

Kikuchi, S., 1985. Relationship between the number of stops and headway for a fixed-route transit system. *Transportation Research Part A* 19(1), 65-71.

Kleven, H. J. and Kreiner, C. T., 2006. The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics* 90(10-11), 1955-1973.

Knight, F. H., 1924. Some fallacies in the interpretation of social cost. *The Quarterly Journal of Economics* 38(4), 582-606.

Kocur, G. and Hendrickson, C., 1982. Design of local bus service with demand equilibration. *Transportation Science* 16(2), 149-170.

Koshy, R. Z. and Arasan, V. T., 2005. Influence of bus stops on flow characteristics of mixed traffic. *Journal of Transportation Engineering* 131(8), 640-643.

Kraus, M., 1989. The welfare gains from pricing road congestion using automatic vehicle identification and on-vehicle meters. *Journal of Urban Economics* 25, 261-281.

Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. *Journal of Urban Economics* 29(2), 249-259.

Kraus, M. and Yoshida, Y., 2002. The commuter's time-of-use decision and optimal pricing and service in urban mass transit. *Journal of Urban Economics* 51(1), 170-195.

Kraus, M., 2003. A new look at the two-mode problem. *Journal of Urban Economics* 54(3), 511-530.



Kuah, G. K. and Perl, J., 1988. Optimization of feeder bus routes and bus-stop spacing. *Journal of Transportation Engineering* 114(3), 341-354.

Kurauchi, F., Bell, M. and Schmöcker, J.-D., 2003. Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms* 2(4), 309-327.

Lam, W. H. K., Cheung, C.-Y. and Lam, C. F., 1999. A study of crowding effects at the Hong Kong light rail transit stations. *Transportation Research Part A: Policy and Practice* 33(5), 401-415.

Leiva, C., Muñoz, J. C., Giesen, R. and Larrain, H., 2010. Design of limited-stop services for an urban bus corridor with capacity constraints. *Transportation Research Part B* 44(10), 1186-1201.

Levinson, H. S., 1983. Analyzing transit travel time performance. *Transportation Research Record* 915, 1-6.

Lévy-Lambert, H., 1968. Tarification des services a qualité variable- application aux péages de circulation. *Econometrica* 36(3/4), 564-574.

Leurent, F. and Liu, K., 2009. On seat congestion, passenger comfort and route choice in urban transit: a network equilibrium assignment model with application to Paris. 88th Annual Transportation Research Board Meeting, Washington, DC, January 2009.

Li, Z., Hensher, D. A. and Rose, J. M., 2010. Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E* 46(3), 384-403.

Li, Z. and Hensher, D. A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. *Transport Policy* 18(6), 880-887.

Lin, T. and Wilson, N. H. M., 1992. Dwell time relationships for light rail systems. *Transportation Research Record* 1361, 287-295.

Lindsey, R., 2006. Do economists reach a conclusion on road pricing? The intellectual history of an idea. *Econ Journal Watch* 3(2), 292-379.

Lu, L., Su, Y., Yao, D., Li, L. and Li, Z., 2010. Optimal design of bus stops that are shared by multiple lines of buses. 13th International IEEE Annual Conference on Intelligent Transportation Systems. Madeira Island, Portugal, September 19-22, 2010.

Maier, H., 2011. CAPTRAS and CONGTRAS: Alternative ways of modelling crowding in RAILPLAN. UK EMME Users' Conference, London, 22 June 2011.

Maloney, M. and Boyle, D., 1999. Components of travel time on the Glendale Beeline bus network. *Transportation Research Record* 1666, 23-27.

Marchand, M., 1968. A note on optimal tolls in an imperfect environment. *Econometrica* 36(3/4), 575-581.

Maunsell, F. and Macdonald, M., 2007. Rail overcrowding, reliability and frequency. Report for Centro, the West Midlands Passenger Transport Executive.

May, A. D., Shepherd, S. P. and Bates, J. J., 2000. Supply curves for urban road networks. *Journal of Transport Economics and Policy* 34(3), 261-290

May, A., Koh, A., Blackledge, D. and Fioretto, M., 2010. Overcoming the barriers to implementing urban road user charging schemes. *European Transport Research Review* 2(1), 53-68.

Mazloumi, E., Rose, G., Currie, G. and Sarvi, M., 2011. An integrated framework to predict bus travel time and its variability using traffic flow data. *Journal of Intelligent Transportation Systems* 15(2), 75-90.

Mazzulla, G. and Eboli, L., 2006. A service quality experimental measure for public transport. *European Transport \ Trasporti Europei* 34, 42-53.

McKnight, C. E., Levinson, H. S., Ozbay, K., Kamga, C. and Paaswell, R. E., 2003. Impact of congestion on bus operations and costs. FHWA Report, October 2003.

Milkovits, M. N., 2008. Modeling the factors affecting bus stop dwell time: use of automatic passenger counting, automatic fare counting, and automatic vehicle location data. *Transportation Research Record* 2072, 125-130.

Mohring, H., 1972. Optimization and scale economies in urban bus transportation. *American Economic Review* 62(4), 591-604.

Mohring, H., 1983. Minibuses in urban transportation. *Journal of Urban Economics* 14(3), 293-317.

Oldfield, R. H. and Bly, P. H., 1988. An analytic investigation of optimal bus size. *Transportation Research Part B* 22(5), 319-337.

Parry, I. W. H. and Bento, A., 2001. Revenue recycling and the welfare effects of road pricing. *The Scandinavian Journal of Economics* 103(4), 645-671.

Parry, I. W. H. and Small, K. A., 2009. Should urban transit subsidies be reduced? *American Economic Review* 99(3), 700-724.

Pedersen, P. A., 2003. On the optimal fare policies in urban transportation. *Transportation Research Part B* 37(5), 423-435.

Pels, E. and Verhoef, E., 2007. Infrastructure pricing and competition between modes in urban transport. *Environment and Planning A* 39(9), 2119-38.

Pickrell, D., 1985. Rising deficits and the uses of transit subsidies in the United States. *Journal of Transport Economics and Policy* 19(3), 281-298.

Pigou, A. C., 1920. *Wealth and Welfare*, MacMillan, London.

Preston, J. M., 2008. Public transport subsidisation. In: Ison, S. & Rye, T. (Eds) *The Implementation and Effectiveness of Transport Demand Management measures: An International Perspective*, Chapter 10, 189-209.

Proost, S. and Van Dender, K., 2004. Marginal social cost pricing for all transport modes and the effects of modal budget constraints. In: *Road Pricing: Theory and Evidence*. G.Santos (Ed.). Elsevier, 159-177.

Proost, S., De Borger, B. and Koskenoja, P., 2007. Public finance aspects of transport charging and investments. in A.De Palma, R.Lindsey, and S.Proost (eds.), *Investment and the Use of Tax and Toll Revenues in the Transport Sector*, *Research in Transportation Economics* 19, Elsevier, 59-80.

Proost, S. and Van Dender, K., 2008. Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. *Transportation Research Part A* 42(9), 1220-1230.

Puong, A., 2000. Dwell time model and analysis for the MBTA Red Line. MIT's Open Course Ware project.

Quinet, E., 2005. Alternative pricing doctrines. *Research in Transportation Economics* 14, 19-47.

Ramsey, F. P., 1927. A contribution to the theory of taxation. *The Economic Journal* 37(145), 47-61.

RAND, 2006. Benefits of Punctuality Improvements for the Paris Suburban Railway Network. Technical report.

Robinson, D. L., 2005. Safety in numbers in Australia: more walkers and bicyclists, safer walking and bicycling. *Health Promotion Journal of Australia* 16(1), 47-51.

Rouwendal, J. and Verhoef, E. T., 2006. Basic economic principles of road pricing: from theory to applications. *Transport Policy* 13(2), 106-114.

RTA, 2011. Key Roads Performance Report. Roads and Traffic Authority, NSW Government, September 2011.

Sáez, D., Cortés, C. E., Milla, F., Núñez, A., Tirachini, A. and Riquelme, M., 2012. Hybrid predictive control strategy for a public transport system with uncertain demand. *Transportmetrica* 8(1), 61-86

Safirova, E., Houde, S., Lipman, D. A., Harrington, W. and Baglino, A., 2006. Congestion pricing, long-term economic and land-use effects. RFF Discussion Paper 06-37.

Safirova, E., Houde, S. and Harrington, W., 2007. Marginal social cost pricing on a transportation network: a comparison of second-best policies. RFF Discussion Paper 07-52.

Saka, A. A., 2001. Model for determining optimum bus-stop spacing in urban areas. *Journal of Transportation Engineering* 127(3): 195-199.

Savage, I. and Small, K. A., 2010. A comment on 'subsidisation of urban public transport and the mohring effect'. *Journal of Transport Economics and Policy* 44, 373-380.

Schmöcker, J.-D., Fonzone, A., Shimamoto, H., Kurauchi, F. and Bell, M. G. H., 2011. Frequency-based transit assignment considering seat capacities. *Transportation Research Part B* 45(2), 392-408.

Senna, L. A. D. S., 1994. The influence of travel time variability on the value of time. *Transportation* 21, 203-228.

Shepherd, S., 2008. The effect of complex models of externalities on estimated optimal tolls. *Transportation* 35(4), 559-577.

Sherman, R., 1971. Congestion interdependence and urban transit fares. *Econometrica* 39(3), 565-576.

Small, K.,A., 1983. Bus priority and congestion pricing on urban expressways. *Research in Transportation Economics* 1, 27-74.

Small, K. A., 2008. Urban transportation policy: A guide and road map. Forthcoming in Robert P. Inman, ed., *The Urban Enigma: City Problems, City Prospects*, Princeton University Press.

Small, K. A. and Chu, X., 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37, 319-352.

Small, K. A. and Verhoef, E. T., 2007. *The Economics of Urban Transportation*, Routledge, Taylor & Francis Group.

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

Strathman, J., Dueker, K., Kimpel, T., Gerhart, R., Turner, K., Taylor, P., Callas, S., Griffin, D. and Hopper, J., 1999. Automated bus dispatching, operations control, and service reliability: baseline analysis. *Transportation Research Record* 1666, 28-36.

Strathman, J., K. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, Callas, S. and Griffin, D., 2000. Service reliability impacts of computer-aided dispatching and automatic vehicle location technology: A Tri-Met case study. *Transportation Quarterly* 54(3), 85-102.

Strathman, J. G., Kimpel, T. J., Dueker, K. J., Gerhart, R. L. and Callas, S., 2002. Evaluation of transit operations: data applications of Tri-Met's automated Bus Dispatching System. *Transportation* 29(3), 321-345.

Sumalee, A., Tan, Z. and Lam, W. H. K., 2009. Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B* 43(8-9), 895-912.

Sun, A. and Hickman, M., 2008. The holding problem at multiple holding stations., *Computer-aided Systems in Public Transport. Lecture Notes in Economics and Mathematical Systems* 600, 339-359.

Tabuchi, T., 1993. Bottleneck congestion and modal split. *Journal of Urban Economics* 34(3), 414-431.

TDC, 2010. 2008/09 Household Travel Survey. Summary Report, Transport Data Centre, Transport NSW.

Tétreault, P. R. and El-Geneidy, A. M., 2010. Estimating bus run times for new limited-stop service using archived AVL and APC data. *Transportation Research Part A* 44(6), 390-402.

Tirachini, A., 2011. Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers. Forthcoming in *Transportmetrica*, First published on: 30 March 2011 (iFirst).

Tirachini, A., 2012. Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services. Forthcoming in *Transportation Research Part C*.

Tirachini, A., Hensher, D. A. and Jara-Díaz, S., 2010a. Restating modal investment priority with an improved model for public transport analysis. *Transportation Research Part E* 46(6), 1148-68.

Tirachini, A., Hensher, D. A. and Jara-Díaz, S. R., 2010b. Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network. *Research in Transportation Economics* 29(1), 231-42.

Tirachini, A., Cortés, C. E. and Jara-Díaz, S. R., 2011. Optimal design and benefits of a short turning strategy for a bus corridor. *Transportation* 38(1), 169-189

Tirachini, A. and Hensher, D. A., 2011. Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transportation Research Part B* 45(5), 828-844.

Tirachini, A. and Hensher, D. A., 2012. Multimodal transport pricing: first best, second best and extensions to non-motorized transport. *Transport Reviews* 32(2), 181-202.

Tisato, P., 1998. Optimal bus subsidy and cross subsidy with a logit choice model. *Journal of Transport Economics and Policy* 32(3), 331-350.

Turvey, R. and Mohring, H., 1975. Optimal bus fares. *Journal of Transport Economics and Policy* 9(3), 280-286.

TRB, 1996. Guidelines for the location and design of bus stops. TCRP Report 19, Transit Cooperative Research Program, Washington D.C.

- TRB, 2000. Highway capacity manual. National Research Council, Washington D.C.
- TRB, 2003. Transit capacity and quality of service manual. TCRP Report 100, Transit Cooperative Research Program, Washington D.C.
- Valencia, A. and Fernández, R., 2007. Representación macroscópica de la operación segregada del transporte público de superficie ("Macroscopic representation of the segregated operation of buses", in Spanish). Proceedings of the 13th Chilean Conference on Transport Engineering, Santiago.
- Van Reeve, P., 2008. Subsidisation of urban public transport and the Mohring effect. *Journal of Transport Economics and Policy* 42, 349-359.
- Vickrey, W. S., 1969. Congestion theory and transport investment. *The American Economic Review* 59(2), 251-260.
- Vuchic, V. R., 2005. *Urban Transit Operations, Planning and Economics*. John Wiley & Sons, 2005.
- Walters, A. A., 1961. The theory and measurement of private and social cost of highway congestion. *Econometrica* 29(4), 676-699.
- Wang, B. and Legaspi, J., 2012. Developing a train crowding economic costing model and estimating passenger crowding cost of Sydney CityRail network. Submitted for presentation at the 2012 Australasian Transport Research Forum, 26-28 September 2012, Perth, Australia.
- Wardman, M. and Whelan, G. A., 2011. Twenty years of rail crowding valuation studies: evidence and lessons from British experience. *Transport Reviews* 31(3), 379-398.
- Watkiss, P., 2002. Fuel taxation inquiry: the air pollution costs of transport in Australia. Report for the Commonwealth of Australia, March 2002.
- Welding, P. I., 1957. The instability of a close-interval service. *OR* 8(3), 133-142.
- Whelan, G. and Crockett, J., 2009. An investigation of the willingness to pay to reduce rail overcrowding. International Conference on Choice Modelling, Harrogate, England, April.
- Wichiensin, M., Bell, M. G. H. and Yang, H., 2007. Impact of congestion charging on the transit market: An inter-modal equilibrium model. *Transportation Research Part A* 41(7), 703-713.

Wright, L. and Hook, W., 2007. Bus Rapid Transit Planning Guide, 3rd edition, Institute for Transportation and Development Policy, New York.

York, I. O., 1993. Factors Affecting Bus-Stop Times, Transport Research Laboratory, Project Report 2, Crowthorne.

Zhao, X.-m., Gao, Z.-y. and Jia, B., 2007. The capacity drop caused by the combined effect of the intersection and the bus stop in a CA model. *Physica A: Statistical Mechanics and its Applications* 385(2), 645-658.

Zorn, L., Sall, E. and Wu, D., 2012. Incorporating crowding into the San Francisco activity-based travel model. Forthcoming in *Transportation*.



# Appendices

## Appendix A1: First Order Conditions, First Best and Second Best Pricing Models (Chapter 3)

### First order conditions for first best (Section 3.3)

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda [f_b K_b - q_b] \quad (\text{A1.1})$$

$$\frac{\partial L}{\partial q_a} = \frac{\partial B}{\partial q_a} - c_a - q_a \frac{\partial c_a}{\partial q_a} - q_b \frac{\partial c_b}{\partial q_a} = 0 \quad (\text{A1.2})$$

$$\frac{\partial L}{\partial q_b} = \frac{\partial B}{\partial q_b} - q_a \frac{\partial c_a}{\partial q_b} - c_b - q_b \frac{\partial c_b}{\partial q_b} - \lambda = 0 \quad (\text{A1.3})$$

$$\frac{\partial L}{\partial q_e} = \frac{\partial B}{\partial q_e} - c_e = 0 \quad (\text{A1.4})$$

$$\frac{\partial L}{\partial f_b} = -q_a \frac{\partial c_a}{\partial f_b} - q_b \frac{\partial c_b}{\partial f_b} + \lambda K_b = 0 \quad (\text{A1.5})$$

$$\frac{\partial L}{\partial K_b} = -q_a \frac{\partial c_a}{\partial K_b} - q_b \frac{\partial c_b}{\partial K_b} + \lambda f_b = 0 \quad (\text{A1.6})$$

$$\lambda [f_b K_b - q_b] = 0 \quad (\text{A1.7})$$

Recalling the equilibrium condition (3.4), (A1.2) and (A1.4) yield results (3.8) and (3.9).

First order conditions for second best (Section 3.4)

$$L = B(q_a, q_b, q_e) - q_a c_a(q_a, q_b, f_b, K_b) - q_b c_b(q_a, q_b, f_b, K_b) - q_e c_e + \lambda [f_b K_b - q_b] + \gamma_a \left( c_a - \frac{\partial B}{\partial q_a} \right) + \gamma_b \left( c_u + \tau_b - \frac{\partial B}{\partial q_b} \right) + \gamma_e \left( c_e - \frac{\partial B}{\partial q_e} \right) \quad (\text{A1.8})$$

$$\frac{\partial L}{\partial q_a} = \frac{\partial B}{\partial q_a} - c_a - q_a \frac{\partial c_a}{\partial q_a} - q_b \frac{\partial c_b}{\partial q_a} + \gamma_a \left( \frac{\partial c_a}{\partial q_a} - \frac{\partial^2 B}{\partial q_a^2} \right) + \gamma_b \left( \frac{\partial c_u}{\partial q_a} - \frac{\partial^2 B}{\partial q_a \partial q_b} \right) - \gamma_e \frac{\partial^2 B}{\partial q_a \partial q_e} = 0 \quad (\text{A1.9})$$

$$\frac{\partial L}{\partial q_b} = \frac{\partial B}{\partial q_b} - q_a \frac{\partial c_a}{\partial q_b} - c_b - q_b \frac{\partial c_b}{\partial q_b} - \lambda + \gamma_a \left( \frac{\partial c_a}{\partial q_b} - \frac{\partial^2 B}{\partial q_b \partial q_a} \right) + \gamma_b \left( \frac{\partial c_u}{\partial q_b} - \frac{\partial^2 B}{\partial q_b^2} \right) - \gamma_e \frac{\partial^2 B}{\partial q_b \partial q_e} = 0 \quad (\text{A1.10})$$

$$\frac{\partial L}{\partial q_e} = \frac{\partial B}{\partial q_e} - c_e - \gamma_a \frac{\partial^2 B}{\partial q_e \partial q_a} - \gamma_b \frac{\partial^2 B}{\partial q_e \partial q_b} - \gamma_e \frac{\partial^2 B}{\partial q_e^2} = 0 \quad (\text{A1.11})$$

$$\frac{\partial L}{\partial f_b} = -q_a \frac{\partial c_a}{\partial f_b} - q_b \frac{\partial c_b}{\partial f_b} + \lambda K_b + \gamma_a \frac{\partial c_a}{\partial f_b} + \gamma_b \frac{\partial c_u}{\partial f_b} = 0 \quad (\text{A1.12})$$

$$\frac{\partial L}{\partial K_b} = -q_a \frac{\partial c_a}{\partial K_b} - q_b \frac{\partial c_b}{\partial K_b} + \lambda f_b + \gamma_a \frac{\partial c_a}{\partial K_b} + \gamma_b \frac{\partial c_u}{\partial K_b} = 0 \quad (\text{A1.13})$$

$$\frac{\partial L}{\partial \gamma_a} = c_a - \frac{\partial B}{\partial q_a} = 0 \quad (\text{A1.14})$$

$$\frac{\partial L}{\partial \gamma_b} = c_u + \tau_b - \frac{\partial B}{\partial q_b} = 0 \quad (\text{A1.15})$$

$$\frac{\partial L}{\partial \gamma_e} = c_e - \frac{\partial B}{\partial q_e} = 0 \quad (\text{A1.16})$$

$$\frac{\partial L}{\partial \tau_b} = \gamma_b = 0 \quad (\text{A1.17})$$

$$\lambda [f_b K_b - q_b] = 0 \quad (\text{A1.18})$$

## Appendix A2: Estimation of the Queuing Delay Function (Chapter 5)

To estimate the queuing delay of buses, we use the bus stop simulator IRENE, which can determine the capacity, queuing delay, dwell time, berth usage and other indicators of the performance of a bus stop as a function of a number of inputs such as the boarding and alighting demand, number of berths, bus size and frequency. For a more detailed description of the program see Fernández and Planzer (2002).

Regarding inputs, the following assumptions are made for the simulations:

- Bus size: Four different bus sizes are considered in accordance with standard commercial vehicle sizes: 8-, 12-, 15- and 18-metre long buses.
- Number of berths: Three configurations are simulated, with one, two and three contiguous berths.
- Berth length: Each berth is assumed to be 1.5 times the bus length, which is the minimum distance necessary for buses to manoeuvre and overtake a preceding bus if necessary (Wright and Hook, 2007).
- Bus saturation flow: This parameter depends on the length of the bus and influences the queuing delay. We assume a basic saturation flow of  $s = 2086$  passenger cars per hour per lane (Akçelik and Besley, 2002) and apply the following equivalency factors depending on the size of the bus (Basso and Silva, 2010): 1.65 (8 m), 2.19 (12 m), 2.60 (15 m) and 3.00 (18 m), yielding estimated saturation flows of 1262, 951, 823 and 694 bus/h for 8, 12, 15 and 18-metre buses, respectively.

A total of 265 simulations were run encompassing all bus sizes and bus stop designs previously described for a range of frequencies from 20 to 220 bus/h and dwell times between 10 and 65 seconds. Buses are assumed to arrive at a constant rate at stops (no bus bunching) and bus stops are isolated from traffic lights.

### Appendix A3: Estimation of Delay at Intersections (Chapter 7)

The mean queuing delay at intersections is modelled as equation (7.4), where  $d_1$  is the non-random delay,  $d_2$  is the overflow delay,  $t_i$  is the acceleration and deceleration delay given by (7.3), and  $h_i$  is the average number of stops per intersection. The non-random delay can be expressed as (A3.1) (Akçelik, 1981; Akçelik and Roupail, 1993).

$$d_1 = \frac{0.5C_T(1-u)^2}{1-ux} \quad (\text{A3.1})$$

$C_T$  is the traffic light cycle time [s],  $u = g/C_T$  is the ratio of effective green time  $g$  [s] to the cycle time  $C_T$ , and  $x_b = f_b/K_r$  is the degree of saturation, given the capacity of the intersection  $K_r = s_f \cdot u$  [veh/h], where  $s_f$  is the saturation flow rate [veh/h] and  $f_b$  is the bus frequency [veh/h]. On the other hand, the overflow delay  $d_2$  is positive only for a degree of saturation greater than 0.67 (Akçelik, 1981). Assuming  $u=0.6$  and  $s_b=694$  bus/h, the critical bus flow that would yield overflow delays is 279 bus/h, which is larger than the capacity of the corridor (given by the capacity of the busiest bus stations, typically between 100 and 200 buses per hour), and then the overflow delay  $d_2$  can be ignored.

The average number of times that vehicles stop per intersection (stops/veh) is given by (A3.2) and accounts for the fact that not all vehicles get to stop at intersections (note that A3.2 is imbedded in the uniform delay A3.1).

$$h_i = \frac{1-u}{1-ux} \quad (\text{A3.2})$$

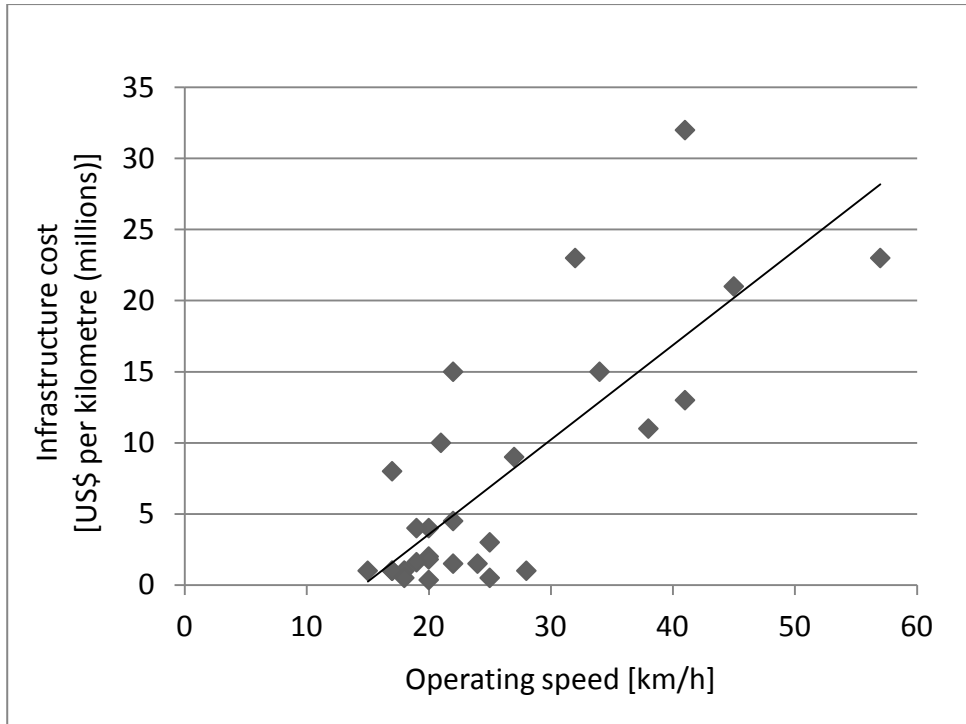
Then, given (A3.1), (A3.2) and the acceleration and deceleration delay (7.3), the total delay at intersections is obtained as (7.5).

## Appendix A4: Estimation of Parameters of the Operator Cost Functions (Chapters 7 and 8)

For the infrastructure and land cost  $c_1$ , there is no empirical data to directly support equation the linear function proposed in equation (7.16); nevertheless, drawing on data on several Bus Rapid Transit systems around the world reported in Wright and Hook (2007), we found a positive correlation between investment in infrastructure per kilometre and the operating (commercial) speed  $V$  (total speed including running time and stops) achieved by the buses<sup>63</sup>, as shown in Figure A4.1 (the straight line represents the linear fit).

---

<sup>63</sup> In principle, it is always possible to increase the bus running speed in a corridor up to a limit, even if it comes at a very high cost, e.g., if space limitations make it impossible to have busways on the ground, tunnels could be provided for a fast circulation of buses. However, if a costly solution is suggested by the optimality analysis carried out in Chapter 7 but is not possible due to financial constraints, there will be an active constraint on the running speed of buses given by the budget that is available for infrastructure investment. In this case, if demand increases and improvements in speed are not possible, the gap in total cost between the optimal (unconstrained) and the actual solution is increased.



**Figure A4.1: Infrastructure cost per kilometre versus operating (commercial) speed<sup>64</sup>**

If the cost of bus stations is excluded from the infrastructure cost (because it will be considered in  $c_2$ ), we postulate that the main impact of investment in infrastructure is on the running speed  $v_0$ . Unfortunately there is no information in the available data sources on the relative investment in land, roads, bus stops, etc., that leads to the values presented in Figure A4.1. It seems reasonable to assume that 70 percent of the infrastructure cost is allocated to road infrastructure (land acquisition plus busway construction), and also that the constant running speed  $v_0$  is 30 percent higher than the commercial speed; hence we can estimate the linear relation (7.16) for the data in Figure A4.1, where  $c_{10} = -235.9$   $\$/h\text{-km}$ ,  $c_{11} = 12.4$   $\$/\text{km}^2$ . A range of non-linear functional forms support the linear approximation (7.16). Calculation assumes an asset life of 50 years, discount rate of 7 percent. To translate annuity calculations into hourly costs, it is necessary to estimate the amount of equivalent hours of operation of a particular period

<sup>64</sup> Based on data from Wright and Hook (2007), including a total of 29 BRT systems in which both data items (infrastructure cost and average commercial speed) are provided: Brisbane, Sydney, Mexico City, León, Quito Central, Quito Ecovía, Guayaquil, Bogotá, Pereira, Curitiba, Sao Paulo, Goiana, Porto Alegre, Miami, Eugene, Los Angeles, Pittsburgh West, Pittsburgh East, Pittsburgh South, Amsterdam, Eindhoven, Rouen, Crawley, Beijing, Kunming, Hangzhou, Seoul, Jakarta and Taipei. Guided busways (i.e., systems with side guide wheels, such as Nagoya and Adelaide) are excluded.

over a year. Using the morning peak, it can be estimated that a year is equivalent to 2947 peak hours of operation for a typical urban bus service in Australia (see Tirachini *et al.*, 2010), value that was used to calculate  $c_{10}$  and  $c_{11}$ . The overall explained variation in infrastructure cost per kilometre attributable to commercial speed is  $R^2=0.625$ .

The station infrastructure cost depends on the amenities provided, quality of shelter and overall design, ranging from \$15,000 for a simple shelter to \$150,000 or more for stations with passenger enclosure, at-level boarding, retail services and detailed passenger information (FTA, 2009). In this paper we assume that the cost increases linearly with bus length: \$50,000 (8 m. bus), \$75,000 (12 m. bus), \$100,000 (15 m. bus) and \$125,000 (18 m. bus), values that are amplified by 25 percent if off-board payment is provided.

There are two vending machines per station, and four fare collection readers in case of off-board payment. Fare collection costs are taken from Wright and Hook (2007), the cost of a fare collection reader is \$750 (coins), \$1,750 (magnetic strip) and \$2,500 (contactless card), while the cost per vending machine is \$10,000 (magnetic strip) and \$15,000 (smart card). The cost of software is \$100,000 for coin payment, \$300,000 for magnetic strip and \$500,000 for contactless card. Bus driving cost is \$29.9 (Hensher, 2010), value that is increased by 21 percent to account for overhead operating costs (e.g., administration, supervision, depot-relating costs), following ATC (2006). The cost of buses is \$160,000 (8 m.), \$370,000 (12 m.), \$520,000 (15 m.) and \$700,000 (18 m.). The estimated parameters in Tables A4.1 and A4.2 are adjusted to 2011 Australian Dollars assuming 20 years of asset life for buses, 15 years for stations and 5 years for software, card readers and vending machines; one year is equivalent to 2947 peak hours of operation.

**Table A4.1: Cost items related to bus size**

<b>Bus size [m]</b>	<b>Bus cost [\$/bus-h]</b>	<b>Driver cost [\$/bus-h]</b>	<b>Station cost [\$/station-h]</b>
8	5.1	37.6	4.4
12	11.9	37.6	6.5
15	16.9	37.6	8.7
18	22.0	37.6	10.9

**Table A4.2: Cost items related to fare collection technology**

<b>Technology</b>	<b>Software cost [\$/h]</b>	<b>Card reader [\$/h]</b>	<b>Vending machine [\$/h]</b>
Coin	12.1	0.1	0.0
Magnetic strip	36.3	0.2	2.4
Contactless card	60.5	0.3	3.6
Off-board	60.5	0.3	3.6

In general, operators have a reserve fleet to deal with unexpected breakdowns and maintenance, which in the model is internalised by applying a safety factor  $\eta > 1$  to the calculation of the fleet size (e.g.,  $\eta = 1.05$  meaning that 5 percent of vehicles are not used and kept at depots). Equivalently, in the model,  $\eta$  is applied to the part of  $c_3$  (cost per bus-hour) that accounts for the rolling stock capital cost.

Finally, for the running cost per vehicle-kilometre  $c_4$  (equation 7.19) the estimated parameters are<sup>65</sup>  $c_{40} = 0.077$  \$/bus-km,  $c_{41} = 0.029$  \$/m-km and  $c_{42} = -0.0013$  \$-h/bus-km<sup>2</sup>. The negative value of  $c_{42}$  means that the fuel consumption is a decreasing function of speed on a distance base (litres/km) as also found by Hossain and Kennedy (2008) for speeds up to 80 km/h, even though the fuel consumption can increase with speed on a time base (litres/h). The overall goodness of fitness of expression (21) is  $R^2 = 0.53$ . A number of non-linear functional forms did not produce a significant improvement over the linear function (21).

---

<sup>65</sup> Using data collected by Hensher (2003). The sample size is 82 observations. It is assumed that the running speed is 30 percent higher than the average (commercial) speed.



## Appendix A5: Bus Internal Layout: Passengers Seating and Standing and Constraints for the Determination of the Number of Seats (Chapter 8)

Let  $A(s_b)$  be the total area available in a bus for seating and standing, which is a function of the bus length  $s_b$ . If  $P_s$  is the proportion of  $A$  allocated to seating, the areas for seating  $A_{seat}$  and standing  $A_{stand}$  can be formulated as:

$$A_{seat}(P_s, s_b) = P_s A(s_b) \quad (A5.1)$$

$$A_{stand}(P_s, s_b) = [1 - P_s] A(s_b) \quad (A5.2)$$

If  $a_{seat}$  is the area required by one bus seat ( $m^2$ ), then the number of seats  $n_{seat}$  per bus is

$$n_{seat} = \frac{A_{seat}}{a_{seat}} \quad (A5.3)$$

For the estimation of in-vehicle time costs, it is necessary to determine the proportion of seats being occupied  $p_{seat}$  and the density of standees  $n_{den}$  (if any) in each segment of a bus trip. Taking direction 1, if  $\lambda^{i+}$  and  $\lambda^{i-}$  are the number of passengers getting on and off a bus at stop  $i$ , the number of passengers  $q^i$  on board a bus between stops  $i$  and  $i+1$  is calculated recursively:

$$q^0 = 0 \quad (A5.4)$$

$$q^i = q^{i-1} + \lambda^{i+} - \lambda^{i-} \quad \forall i \in \{1, P-1\} \quad (A5.5)$$

Separating  $q^i$  among passengers seating  $q_{seat}^i$  and standing  $q_{stand}^i$ , we can obtain  $p_{seat}^i$  and  $n_{den}^i$  as follows:

$$p_{seat}^i = \frac{q_{seat}^i}{n_{seat}} = \frac{\min\{n_{seat}, q^i\}}{n_{seat}} \quad (A5.6)$$

$$n_{den}^i = \frac{q_{stand}^i}{A_{stand}} = \frac{q^i - q_{seat}^i}{A_{stand}} \quad (A5.7)$$

To calculate the area available for seating  $A_{seat}$  and standing  $A_{stand}$ , we need an estimation of the area occupied by seats, standees, doors and other elements. The U.S. Transportation Research Board recommends the following values (TRB, 2003):

**Table A5.1: Area occupied by passengers sitting, standing and other objects**  
(Source: TRB, 2003)

Situation	Projected area [m <sup>2</sup> ]
Standing	0.15-0.20
Standing with briefcase	0.25-0.30
Standing with daypack	0.30-0.35
Standing with suitcase	0.35-0.55
Transverse seating	0.50
Longitudinal seating	0.40
Wheelchair space	0.95
Rear door	0.80

We use Table A5.1 and the following assumptions in order to calculate seating and standing areas, feasible numbers of seats and total bus capacity:

(a1) Buses have transverse seating only, therefore 0.5 m<sup>2</sup> is the value used for the area occupied by passengers sitting.

(a2) The maximum density of standees  $d_{max}$  is around 6.7 pax/m<sup>2</sup>, equivalent to an area of 0.15 m<sup>2</sup> per standee. However, given the Sydney context  $d_{max}$  is set as 4 pax/m<sup>2</sup> in Chapter 8.

(a3) Buses are 2.55 metre wide (regardless of length)

(a4) The front area must be left clear of passengers, for the driver and front door. This area is 1.5 metre long.

(a5) Next to each rear door there has to be a 0.8 m<sup>2</sup> area clear of standees. The number of doors per bus is denoted as  $n_{doors}$ .

(a6) Buses must have a  $0.95 \text{ m}^2$  area reserved for wheelchairs.

Using (a3) to (a6), the total area  $A \text{ (m}^2\text{)}$  available for seating and standing is:

$$A = A_{sit} + A_{stand} = 2.55(s_b - 1.5) - 0.8(n_{doors} - 1) - 0.95 \quad (\text{A5.8})$$

And the capacity of a bus (maximum number of passengers that can be accommodated) is:

$$K(s_b, P_s, n_{doors}) = \left[ \frac{P_s}{a_{seat}} + (1 - P_s)d_{max} \right] A \quad (\text{A5.9})$$

### Constraints

(c1) An aisle is provided in the centre of the bus, with a minimum width of 0.5 metre. This aisle does not necessarily have to cover the full length of the bus as the back row may have a seat in the middle (where the aisle ends). Therefore, assuming that 1.5 metre is left at the front and 0.7 metres is used for a seat at the back, the minimum area that has to be reserved for the aisle is  $A_{stand}^{\min} = 0.5(s_b - 2.2)$ . Then, the number of seats is upper bounded by:

$$n_{seat} \leq n_{seat}^{\max} = \frac{A - A_{stand}^{\min}}{a_{seat}} \quad (\text{A5.10})$$

(c2) A minimum number of seats must be provided, i.e., the proportion  $P_s$  of  $A$  allocated to seating has a lower bound  $P_s^{\min}$ , which is arbitrarily decided (e.g.,  $P_s^{\min} = 0.3$  meaning that at least 30 percent of the available area must be reserved for passengers sitting). Therefore

$$n_{seat} \geq n_{seat}^{\min} = \frac{P_s^{\min} A}{a_{seat}} \quad (\text{A5.11})$$

Combining constraints (A5.10) and (A5.11) the number of seats per bus must comply with inequality (8.18b):  $n_{seat}^{\min} \leq n_{seat} \leq n_{seat}^{\max}$ .