



HAL
open science

Three essays on the economics of congestion in public transport

Guillaume Monchambert

► **To cite this version:**

Guillaume Monchambert. Three essays on the economics of congestion in public transport. Economics and Finance. Université Paris Saclay (COMUE), 2015. English. NNT: 2015SACLN006. tel-01235400

HAL Id: tel-01235400

<https://theses.hal.science/tel-01235400>

Submitted on 30 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS-SACLAY

Préparée à l'École Normale Supérieure de Cachan

ÉCOLE DOCTORALE N° 578

Sciences de l'homme et de la société

Spécialité de doctorat : Sciences Économiques

Three essays on the economics of congestion
in public transport

Guillaume MONCHAMBERT

Thèse présentée et soutenue à Cachan le 26 Octobre 2015 devant le jury composé de :

André DE PALMA	École Normale Supérieure de Cachan	Directeur de thèse
Marc IVALDI	Toulouse School of Economics	Rapporteur
Stef PROOST	Katholieke Universiteit Leuven	Rapporteur
Erik VERHOEF	Vrije Universiteit Amsterdam	Président du jury
Patrick BONNEL	École Nationale des Travaux Publics de l'État	Examineur
Robin LINDSEY	University of British Columbia	Examineur

Acknowledgements

I want to express my gratitude to André de Palma, for everything he has done: I feel very lucky to have had the opportunity to benefit from his insight, encouragement and ideas. I thank him for pushing me to always go a step further.

I also thank Prof. Marc Ivaldi and Prof. Stef Proost for agreeing to be the rapporteurs for this dissertation, and Prof. Patrick Bonnel and Prof. Erik Verhoef for being members of the jury. Prof. Robin Lindsey deserves a very special thank for having accepted to be both a co-author and a member of a jury.

My indebtedness goes to Martin Koning, who provided me with a clean database of a survey he conducted on his own in the Parisian subway a few years ago, and who accepted to co-write a paper. I sincerely thank him for his help, availability and dynamism. I also thank Luke Haywood for his comments on my econometric works.

I would like to thank the Centre d'Économie de La Sorbonne - Cachan for providing me stimulating research environment. I would also like to thank my PhD colleagues and friends from the Centre d'Économie de la Sorbonne - Cachan, and in particular Marine, Lenka, Elissa, Imen, Samuel, Thilo and Pierre for all the fruitful discussions we had together.

I extend my gratitude to my friends from Lyon, Cachan and Paris who were present at each stage of this thesis. I would like to thank my family, especially my parents, my sister Marianne and my brother Aurélien.

Finally, my biggest thank goes to Marine, who shared this adventure with me and is yet ready for the next ones.

Notice

The three chapters of this dissertation are self-contained research articles. This explains why some information can be redundant, and that the terms “paper” or “article” are used. The first chapter of this dissertation is co-authored with André de Palma and was published in *Journal of Urban Economics* (Monchambert and de Palma, 2014). The second chapter is co-written with André de Palma and Robin Lindsey, and is a work in progress. The third chapter is co-written with Luke Haywood and Martin Koning and is also currently in progress. Data used in Chapter III were provided by Martin Koning.

Contents

Acknowledgements	3
Notice	5
Introduction	15
I Public transport reliability and commuter strategy	27
1 Introduction	27
2 Punctuality in public transport	30
2.1 Transport supply	31
2.2 Demand for bus and taxi	33
2.3 Commuters' strategies	36
3 Competition between bus and taxi companies	39
4 Welfare analysis	41
5 Numerical application	46
6 Conclusion	49
Appendices	51
Appendix A Proof of Proposition 1	51
Appendix B Proof of Proposition 2	52
Appendix C Proof of Proposition 3	53
Appendix D Proof: optimal bus and taxi fare	54
Appendix E Proof of Proposition 4	55
Appendix F Proof of Proposition 5	56
Appendix G Proof of lemma 6	58
Appendix H Discussion of Conjecture 2	59

II Economics of crowding in public transport	61
1 Introduction	61
2 The general model with inelastic demand	65
2.1 User equilibrium	68
2.2 Social Optimum	71
3 Linear crowding costs	77
3.1 User equilibrium with a uniform fare	77
3.2 Social Optimum	80
4 Elastic demand	84
4.1 Uniform-fare regimes	85
4.2 Social optimum	86
4.3 Comparison of the regimes	87
5 Optimal transit service	88
5.1 Optimal timetable	88
5.2 General capacity function	89
5.3 A specific capacity function	93
6 A numerical example	97
6.1 Base-case results	98
6.2 Sensitivity analysis	100
7 Conclusion	103
Appendices	107
Appendix A Proof of Proposition 9	107
Appendix B Proof of Proposition 10	108
Appendix C Proof of Proposition 11	109
Appendix D Proof of Proposition 17	112
Appendix D.1 Optimal timetable for user equilibrium	112
Appendix D.2 Optimal timetable for social optimum	114
Appendix E Derivatives of SS^e with respect to m and s	115
Appendix F Proof of Proposition 21	116
Appendix G Parameter values for numerical example	117

Appendix H	Glossary	119
Appendix H.1	Latin characters	119
Appendix H.2	Greek characters	120

III Well-being in public transport: an empirical approach of the crowding

effect		123
1	Introduction	123
2	Data	128
2.1	Survey design	128
2.2	Descriptive statistics	130
3	The crowding effect	133
3.1	Estimation strategy	133
3.2	Descriptive statistics	134
3.3	Estimates results	135
4	Anatomy of the crowding effect	139
4.1	The nuisance factors	140
4.2	Effects of the nuisance factors on the comfort satisfaction	143
4.3	Heterogeneity of users	146
5	Conclusion	148
Appendices		149
Appendix A	Paris subway map	149
Appendix B	Treatment of the selection bias	150
Appendix C	<i>CS</i> estimates	154
Appendix D	Including <i>NFD</i> to explain the comfort satisfaction	155
Appendix E	Users preferences for the nuisance factors	158
Appendix E.1	Estimation strategy	158
Appendix E.2	Regression results	158

Conclusion		161
-------------------	--	------------

Bibliography		178
---------------------	--	------------

List of Figures

1	Average urban public transport share with respect to the average GDP per inhabitant in 1995(Joly et al., 2006, from the Millennium Cities Database, UITP)	20
2	Distribution of users in the Paris subway network as a function of the time of the day, a winter working day in 2013 (OMNIL, 2011).	22
3	The monocentric city	31
4	The arrival probabilities of the bus	33
5	Distribution of taxi trip time in <i>Group A</i> and <i>Group B</i>	34
6	Share of commuter choosing <i>Strategy O</i> and <i>Strategy L</i> as a function of P , the probability of the bus being on time ($\kappa = 8, \tau = 50$)	38
7	Probability of the bus being on time as a function of the punctuality cost c	46
8	Bus patronage as a function of the punctuality cost c	47
9	Relative social gain compared to equilibrium as a function of the punctuality cost c	48
10	Strategy choice of a commuter in <i>Group A</i> depending on the taxi trip time δ .	52
11	Strategy choice of a commuter in <i>Group B</i> depending on the taxi trip time δ .	53
12	Difference between optimal demand for the bus and equilibrium demand for the bus depending on operating taxi cost d for <i>Group B</i>	60
13	Schedule delay $\delta(t_k)$, crowding cost $g(n_k)$ and equilibrium cost c_e for seven trains, $t_5 = t^*$	69
14	User equilibrium (UE), social optimum (SO) and welfare gain (G^{eo}) with two trains	83
15	Ridership and marginal social cost	111

16	A simple journey satisfaction theoretical framework	126
17	Density show-card used during the field survey	129
18	Representative user expected CS* and expected CS as a function of the in-vehicle density and of the income.	138
19	Map of the Parisian underground PT network. Stations in which the survey has been collected are highlighted by white pentagons.	149

List of Tables

1	Parameters values	32
2	Values of main variables when $c = 5$ and $\theta = 0.75$	48
3	Variation table of the difference between optimal demand for the bus and equilibrium demand for the bus depending on the cost of reliability.	58
4	Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: base-case parameter values	98
5	Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: $\eta = -2/3$	101
6	Effects of increasing parameters β and γ (or parameter h) by 10%	102
7	Effects of increasing parameter λ by 10%	103
8	Individual and journey characteristics for the whole sample, the two subsamples and the Enquête Globale Transport (EGT)	132
9	Distribution of CS	135
10	Effects of the density and the income on the CS	136
11	Rank and score statistics for the 8 underlying nuisance factors of crowding .	142
12	Effects of the crowding on the CS	145
13	Main effects of socio-economic variables on Nuisance Factors Dissatisfaction scores	147
14	Probit model for the selection (Output=1 if the user answers the whole survey)	153
15	Effects of the crowding on the CS	154
16	CS , first specification	155
17	CS , second specification	156
18	CS , third specification	157

19	Ordered probit estimates for NFD_{Standing} (1), $NFD_{\text{Wasted Time}}$ (2), and NFD_{Noise} (3)	159
----	---	-----

Introduction

Congestion arises when the enjoyment of a good by one individual diminishes the consumption possibilities of other individuals (Oakland, 1972). This situation is more likely to happen when the number of individuals having the opportunity to consume the good is large, other things being equal. This explains why congestion is an inherent feature of large agglomerations, and why it may be considered as the negative counterpart to the agglomeration effect. Therefore Ciccone (2002) names "net agglomeration effect" the difference between positive externalities due to agglomeration and congestion effects. A concrete and explicit example of congestion in cities is observable on roads during morning and evening commuting peaks. During these periods characterized by huge traffic jam, an additional road driver suffers from longer travel time, but he is also the direct cause of part of the traffic jam which decreases the utility of other drivers. So the consumption of road by one driver diminishes the road consumption possibility of other drivers. According to the results of the survey provided by the Centre for Economics and Business Research and traffic information company Inrix (CEBR - INRIX, 2014), the cost of road congestion in 2013 is estimated at €93.8 billion for the US economy (0.74% of the GDP), €25.2 billion for the Germany economy (0.92% of the GDP), €17 billion for the France (0.82% of the GDP) and €15.5 billion for the UK (0.81% of the GDP). Road congestion is one of the prevalent and majors issues cities face (OECD, 2007), not only because increased travel times are a loss to the society, but also because this rise in travel times intensifies environment and health-harming emissions. Encouraging public transport seems to be an efficient way to reduce road congestion and its externalities, because it is a cleaner mode. Therefore, a number of supply-side and travel demand management policies are commonly

used to discourage driving and encourage the use of public transport. To attract private car users, gain in travel time but also service quality have to be emphasized. For example, the European Commission wrote in the White Paper on transport (European Commission, 2011):

The quality, accessibility and reliability of transport services will gain increasing importance in the coming years, inter alia due to the ageing of the population and the need to promote public transport. Attractive frequencies, comfort, easy access, reliability of services and intermodal integration are the main characteristics of service quality. [p. 12]

Yet these efforts are often hampered by transit capacity constraints. Indeed, for a given supply of public transport, an increase in patronage may deteriorate the service quality: less seats are available and some users have to stand for all or part of their journey; some vehicles are so crowded that users on platforms must wait for the next one; more users enter and leave vehicles at each station, increasing dwelling times, travel times and variability of travel times... Consequently, public transport becomes less attractive. Whereas the congestion on roads is known to all and has been the subject of numerous works by engineers, economists, sociologists and others, fewer have been said or written on congestion in public transport. However, optimal modal split and policies efficiency should be studied in an unified theoretical framework. When analyzing the modal choice, the disutilities (or travel costs) of each mode are compared, and the individual is assumed to choose the less costly. Therefore it does not seem consistent to introduce congestion for the private cars without doing the same for public transport modes. It results in biased results and overestimates the appeal of public transport with respect to private cars. This dissertation aims at filling part of this gap and at proposing tools and insights to analyze congestion in public transport.

There is no general consensus on what congestion is in public transport. A vast literature studies and describes the congestion on road networks, but fewer works investigate the congestion in public transport and propose a satisfactory definition of it. In a report written for the European Conference of Ministers of Transport (ECMT), Schallaböck and

Petersen define congestion in public transport as “a situation in which transport participants cannot move in a usual or desirable manner” (ECMT, 1999, p.13). This definition is interesting because it focuses on the disutility supported by the users. However, the fact that congestion is due to an excess of demand is not enhanced here. According to Koning (2011, p.24), a transport infrastructure is “congested when it is not able to meet the demand without lowering the service quality”. Following this, in this dissertation, I consider the congestion in public transport as the condition occurring when the ratio demand / supply of public transport is high enough to decrease the convenience of users. The convenience is related to “absence of effort” in using public transport facilities (OECD, 2014). In the Oxford Dictionary,¹ public transport² is defined as “buses, trains, and other forms of transport that are available to the public, charge set fares, and run on fixed routes”. The definition of congestion proposed above is valid for all public transport modes. This definition does not presume the channels through which congestion decreases the utility of traveling. In this dissertation, I focus on two features of congestion in public transport: the crowding and the unreliability.

In public transport, one of the main manifestation of the congestion takes the form of crowding. Crowding arises when the number of users is too high with regards to the available space. It may occur not only while riding buses and trains, but also when boarding and alighting from them, while waiting on platforms or at stops, and while accessing stations by escalator, elevator, or on foot. A high density of users decreases the amount of space available for each user and creates promiscuity between users. This promiscuity has physical and psychological consequences on the journey experience. One of the main effects is that when traveling in crowded conditions, users are not able to perform regular activities: difficulties to read newspaper because of the lack of space, inability to focus because of noise... This makes the utility of the time spent in crowded vehicles lower than when vehicles are empty. Many other reasons explain why users are averse to a high density when they travel: anxiety, stress, tiredness, threat to safety... Clearly, crowding

¹<http://www.oxforddictionaries.com/definition/english/public-transport?q=public+transport>.

²In North American english, “public transportation” is sometimes used instead of “public transport”. In this manuscript, both expression are used interchangeably.

is a disutility to users. Tirachini et al. (2013) presents the most detailed literature review about crowding in public transport. They report that crowding is shown to increase in-vehicle time (Lin and Wilson, 1992) and waiting time (Oldfield and Bly, 1988), and to deteriorate travel time reliability (Bates et al., 2001). From psychological studies, they relay that crowding causes stress and feeling of exhaustion (Mohd Mahudin et al., 2012). Concerning the users perception of crowding, many papers show that the in-vehicle time cost beared by users increases with the number of users (Li and Hensher, 2011; Wardman and Whelan, 2011; Haywood and Koning, 2015). Crowding is also an important feature when forecasting the route choice of users (Raveau et al., 2011) and when determining the optimal pricing, subsidy and supply (Parry and Small, 2009). I investigate this disutility in Chapters II and III of this dissertation.

Another feature of congestion is the decrease in reliability. Indeed, when more users take a public transport vehicle, they need more time to board and alight the vehicle, because the boarding/loading speed is physically constrained by the width of the gates. This delays the vehicle but also the following ones which have to respect a certain frequency. Consequently, on the network, travel times and delays increase. Congestion may also take the form of the disability to board the vehicle for users, resulting in increased waiting times on the platforms. This is the perspective used by Kraus and Yoshida (2002). Moreover, overground public transport travel times are intimately related to road congestion. For example, buses suffer from road congestion all along the route on shared with private cars network. Road congestion may also affect vehicles circulating on dedicated lane or rail because of the jamming at intersections of roads dedicated to private cars. In Chapter I, I analyze the unreliability in public transport.

Some facts on public transport

The use of public transport is heterogeneous across the world. It depends on the wealth of individual, on geographical characteristics of cities... Figure 1 displays the average public transport share as a function of the GDP per inhabitant for some categories of cities in 1995. These figures have been computed by Joly et al. (2006) by using the Millennium

Cities Database made by the UITP.³ This is one of the very rare sources which supply some figures on public transport patronage for several zones in the world. Figure 1 allows to characterize the use of public transport with respect to the other modes. Public transport is prevailing in cities from developing countries and from Asia, where between 35 and 60% of trips are made by public transport. It is less used in Europe cities (around 25%), and almost residual in North America and Oceania cities (around 6%).

Despite these differences in modal split, public transport networks are well developed in Europe, especially comparing to North America. The example of the metro diffusion is striking. 45 European cities have metro networks that carry in average 31 millions passengers per day, whereas only 15 north-American metro networks carry 11 millions passengers per daily (UITP, 2014c, p.2).⁴ This difference may be partly explained by higher densities of inhabitants in Europe than in North America. Indeed, automobiles seem most economical at low inhabitants densities, bus transit at medium densities, and rail transit at very high densities (Small, 2008). The number of annual UE local public transport - bus, tram and metro - journeys has increased from 45.6 billions in 2000 to 49.5 billions in 2012 (UITP, 2014b, p.2). In average, an UE inhabitant uses the local public transport for 132 journeys per year. In comparison, commercial aviation accounts for 800 million users journeys per year and long distance rails for 1 billion per year (UITP, 2014a). Among these 49.5 billion local public transport journeys, 64% (31.8 billion) are made in bus or trolley, 19% (9.4 billion) by metro and 17% (9.3 billion) by tramway. In Europe, the bus is from far the most used public transport mode. This trend is likely to continue due to the efficiency of new bus rapid transit on dedicated lanes.

According to the definition of the congestion in public transport, the patronage for public transport can not be analyzed without some figures on the supply of public transport. Such statistics are not available for large zones. Therefore the following focuses on Paris region. Moreover, in Chapter III, I use a survey in which users have been interviewed in the Paris subway, and in Chapter II, I propose a calibration of a model on a segment of RER A line, one of the busiest line in the Parisian public transport network. For the Paris

³See <http://www.uitp.org/public-transport-sustainable-mobility>.

⁴The leader continent in patronage for metro is Asia: 50 cities provide metro for a 71 million passengers daily ridership. 16 Latin-American metro systems carry 15 millions passengers per day.

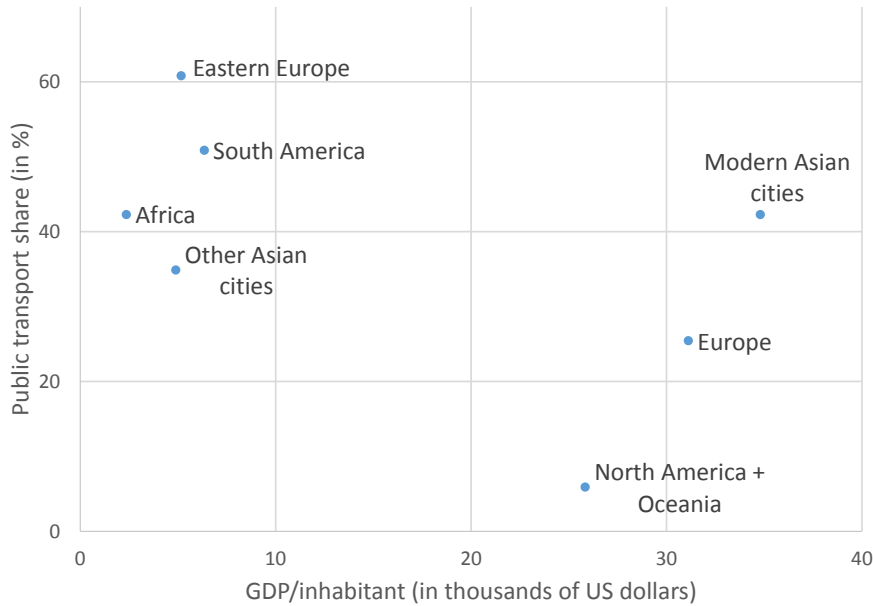


Figure 1: Average urban public transport share with respect to the average GDP per inhabitant in 1995 (Joly et al., 2006, from the Millennium Cities Database, UITP)

Notes. The urban public transport share is the ratio of the number of trips made with public transport in a city on the total number of trips made in a city. The GDP/inhabitant is also computed at the city level. The Africa average is computed on 8 cities (Abidjan, Cairo, Cape Town, Casablanca, Dakar, Harare, Johannesburg and Tunis), the North America + Oceania average on 20 cities (Atlanta, Brisbane, Calgary, Chicago, Denver, Houston, Los Angeles, Melbourne, Montreal, New York, Ottawa, Perth, Phoenix, San Diego, San Francisco, Sydney, Toronto, Vancouver, Washington and Wellington), the South America on 10 cities (Bogota, Brasilia, Buenos Aires, Caracas, Curitiba, Mexico City, Rio de Janeiro, Salvador, Santiago and Sao Paulo), the modern Asian cities average on 5 cities (Hong Kong, Osaka, Sapporo, Singapore and Tokyo) the others Asian cities of 16 cities (Bangkok, Beijing, Chennai, Guangzhou, Ho Chi Minh City, Jakarta, Kuala Lumpur, Manila, Mumbai, New Delhi, Riyadh, Seoul, Shanghai, Taipei, Tel Aviv and Teheran), the Europe average on 35 cities (Amsterdam, Athens, Barcelona, Berlin, Bern, Bologna, Brussels, Copenhagen, Dusseldorf, Frankfurt, Geneva, Glasgow, Graz, Hamburg, Helsinki, Lille, Lisbon, London, Lyon, Madrid, Manchester, Marseille, Milan, Munich, Nantes, Newcastle, Oslo, Paris, Rome, Ruhr, Stockholm, Stuttgart, Turin, Vienna and Zurich) and the Eastern Europe average on 6 cities (Budapest, Istanbul, Krakow, Moscow, Prague and Warsaw).

region, statistics are available for the annual total of users x kilometers per mode and the annual total of vehicles x kilometers operated by the transit authority. These statistics allow to compute the average number of users per vehicle. In 2013, there were, in average, around 240 users per suburban train (for a capacity going from 200 to 940 seats, depending on the rolling stock), 154 users per metro (seating capacity from 144 to 320), 91 users per tramway (seating capacity from 48 to 80), 20 users per bus in Paris and 14 users per bus outside Paris.⁵ These figures show that the occupancy rates of seats on the network are close to 100% and sometimes higher, meaning that some users have to stand during travel.

However, it is still not sufficient to describe the congestion. Indeed, congestion is dynamic in the sense that the number of users with respect to the supply varies along the day. On Figure 2 is displayed the distribution of users in the Paris metro network as a function of the time of the day for a winter working day in 2013. Two remarks are of interest. First, a morning peak and an evening peak are clearly noticeable, and the evening peak is more spread out in time than the morning peak. Second, the number of users using the facility may vary very quickly, as it is the case between 8am and 9am or between 9am and 10am. This second remark is of first interest. Indeed, in a microeconomic perspective, it means that an individual can decide to travel earlier or later in order to avoid excessive congestion. The departure time decision is the key lever individuals may adapt in order to adapt themselves to congestion, as Knockaert et al. (2012) showed in an experiment conducted on a congested motorway corridor in the Netherlands. In Chapters I and II, I deeply investigate the departure time decisions of users.

The Parisian public transport system is also subject to a lack of reliability. This unreliability may be due to excess of demand or to exogenous incidents. In 2013, 16.1% of RER A users and 16.9% of RER B users arrived at their destination with a late delay higher or equal to 5 minutes (STIF, 2014). With respect to the annual patronage for these two lines (around 310 million RER A users and 220 million RER B users) and to an average value of time of €15 per hour (Wardman et al., 2012), the social cost of unreliability on these two lines amounts at least to €100 million in 2013. This figure highlights the importance of reliability in public transport, an aspect I focus on in Chapter I.

⁵Computations of the author from data supplied by OMNIL (2011).

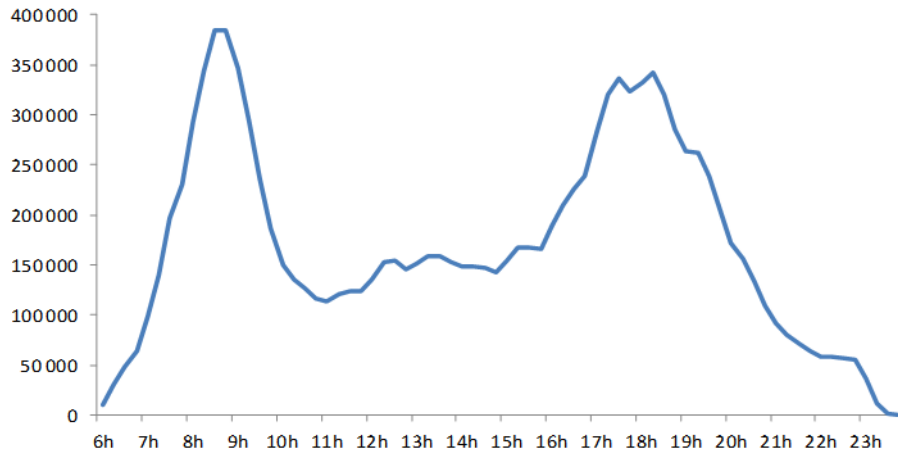


Figure 2: Distribution of users in the Paris subway network as a function of the time of the day, a winter working day in 2013 (OMNIL, 2011).

Roadmap of the dissertation

This dissertation is made of three distinct essays on the congestion in public transport facilities. The two first essays investigate how users get used to lack of punctuality and crowding in public transport. The third essay presents an empirical analysis of the crowding effect.

Chapter I Public transport reliability and commuter strategy

In the first chapter, I focus on the two-way implication between punctuality level of public transport and (potential) customer behavior. The punctuality of public transport is a key element of the service quality. The user cost elements which play an important role in demand analysis are affected by the punctuality level (Bowman and Turnquist, 1981). As a consequence, users and potential users choose both the mode of transport and the departure time depending on public transport punctuality level. Mohring (1972) has shown that scheduled urban public transport is characterized by increasing returns to scale since the frequency increases with demand. Therefore demand is influential in the service quality and the bus company may adapt its punctuality to the level of potential demand. One of the goals of this chapter is to observe if the bus company adapts its service quality to a change in the price of the alternative mode. Another aim is to explore the gap between

the bus punctuality at equilibrium and at optimum. The shift from a mode to another to achieve the optimal modal split is to be highlighted.

I develop a duopoly which embodies a modal competition between public transport and another mode which we call taxi. The attention is focused on the monetary impacts of punctuality. Two different types of variables are observed in the model: the public transport punctuality level which is selected by the bus company and the prices set by bus and taxi companies. Both have a substantial influence on demand for public transport (Paulley et al., 2006). Unreliability has a strong negative impact because it implies excessive wait time and uncertainty (Wardman, 2004; Paulley et al., 2006).

Considering commuting trips, preferences can be analyzed with the dynamic scheduling model. In this model, individual's preferences reflect agents tradeoff between travel time, early schedule delay and late schedule delay. Commuters may choose different strategies to minimize their trip cost. This theory has been first introduced by Vickrey (1969) and then renewed by Arnott et al. (1990). Such analysis usually are specific to road analysis (Fosgerau and Karlström, 2010) therefore I introduce a wait time to extend this model to public transport. In the model, a commuter has the choice between catching the bus and using the taxi service. However he may miss the bus and then he has to use the taxi service. Indeed I assume the headway is so long that all users who miss the bus prefer to use the taxi service. Commuters are differentiated by their preferred arrival time and by their location which is measured as the time to travel to their destination when using the alternative mode. Two different preferred arrival time are considered. The location is uniformly distributed among commuters.

The analysis for the model proceeds in three steps. The first step is to find out for arbitrary public transport and alternative mode prices and punctuality level which commuters will use which mode of transport. The second step is to determine which price and punctuality levels are set by companies at equilibrium given the strategies of commuters identified in step one. The third step is to assess the prices and the punctuality level that minimize the total social cost and to compare these results with the ones found in step two. I find that the public transport reliability set by the public transport firm at the competitive equilibrium increases with the alternative mode fare, via a demand effect. This

is reminiscent of the Mohring Effect. The study of the optimal service quality shows that often, public transport reliability and thereby patronage are lower at equilibrium compared to first-best social optimum.

Chapter II Economics of crowding in public transport

To analyze the welfare effects of public transit crowding, and policies to alleviate it, in a conceptually consistent way, it is necessary to use a structural model that incorporates trip scheduling decisions, an empirically plausible crowding cost function, and alternative pricing (i.e., fare) regimes. Several papers in the transport economics literature have laid much of the groundwork for such a model. Vickrey's (1969) bottleneck congestion model is the seminal work on scheduling of automobile trips. Arnott et al. (1990) extended it to time-varying tolling schemes and capacity investment decisions. Tabuchi (1993) added public transit by considering a setting in which travelers can choose between driving and taking a rail service with scale economies and no crowding. In this chapter, I use this modeling framework to analyze usage of a rail transit line, and assess the potential benefits from internalizing crowding externalities by setting differential train fares. I also present results on optimal train capacity and the number of trains put into service.

To study the behavioral implications and costs of crowding, I develop a structural model in which public transport users face a choice between traveling in a crowded train and arriving when they want, and traveling earlier or later to avoid crowding but arriving at an inconvenient time. Trains run on a fixed timetable between two stations. Riders know the timetable and choose which train to take.

Therefore each commuter chooses when to arrive at the train station by trading off schedule delay costs and crowding costs. Equilibrium obtains when no commuter can decrease his journey cost by changing his departure time, taking all other commuters' departure times as fixed. Thus, as in the Vickrey (1969) model, the equilibrium is a pure-strategy Nash equilibrium with departure times as the decision variables. The social optimum is reached when the marginal social cost of a trip is the same in any train during the peak hour.

I show how the optimum can be decentralized using train-specific fares, and characterize

the welfare gains from optimal pricing. I then allow total demand to be price elastic and compare total usage, private costs, and welfare in the user equilibrium, the social optimum, and a third regime in which an optimal uniform fare is imposed. Finally, I derive the optimal timetable, number of trains, and train capacity for the three fare regimes.

Some of the results parallel those obtained with road traffic congestion models. Passenger loads are often distributed more evenly across trains in the social optimum than the user equilibrium. The social optimum can be decentralized by charging higher fares on more popular trains to internalize the crowding cost externality on each train. Imposing differentiated fares makes users worse off - at least before accounting for how the revenues are used. Other results are less obvious. The welfare gains from tolling are independent of total ridership. Expanding the number of trains can also be more valuable in the social optimum than the user equilibrium even though total system costs are lower in the social optimum.

Chapter III Well-being in public transport: an empirical approach of the crowding effect

Economists often do not distinguish density and crowding and consider these two terms as equivalent. However, according to psychologists, an important distinction has to be done (Stokols, 1972; Baum and Paulus, 1987). In this chapter, the experience of crowding is assumed to be made of several dimensions, the nuisance factors. These nuisance factors of crowding are defined as *the aspects of a journey that are deteriorated by a high density*.⁶ The perception of these dimensions may be different for each user: it is influenced by the travel characteristics, the individual preferences, and of course by the objective in-vehicle density. The concept of comfort satisfaction is larger than the concept of crowding. Indeed, a journey may be experienced as very dis-comfortable despite the train is empty or almost empty, because of heat, smell, accelerating, braking... The expectations or level of requirement may also vary across users. Therefore, in addition to the crowding, the travel and individual characteristics are worth considering when addressing the in-vehicle

⁶In this study, I consider eight dimensions of crowding: *Overcloseness, Standing, Noise, Smell, Time loss, Waste of time, Fall and Robbery*.

comfort.

The objectives of this chapter are twofold. On the one hand, I shed some more light on the in-vehicle comfort satisfaction during public transport journey, by characterizing the effects of crowding, individual preferences and travel characteristics on the user satisfaction. On the other hand, I examine which nuisance factors of crowding construct the crowding experience, and how the dimensions of crowding are influenced by the individual preferences.

The analysis takes the answers to a travel satisfaction question as a measure for user subjective well-being, and the responses to questions about dissatisfaction related to various aspects of comfort when the in-vehicle density is very high as measures for nuisance factors. The data originates from a survey collected on 1,000 Paris subway users. Due to the nature of data, ordered logit models are used during the analysis.

I find a clear crowding effect: on average, an extra-user per square meter decreases by one the expected 0 to 10 scale individual well-being. I do not find any empirical evidence of this effect being intensified by the travel time. However, the crowding effect increases with the income of users. I find three causes of crowding disutility: a higher probability to stand for all or part of the journey, a poorer use of the time during the journey, and noisier travel conditions. These features of discomfort matter more for women and wealthy individuals.

Chapter I

Public transport reliability and commuter strategy

1 Introduction

Despite increasing pollution and congestion in cities, cars remain the most popular mode of transport, because they are usually more convenient than public transport and they keep a strong attractive power due to symbolic and affective motives (Steg, 2005). In the U.S., the predominance of cars is also strengthened, despite the congestion observed on the American highways (The Economist, 2011). Therefore, improving alternative modes of transport and making them attractive is essential in an urban context. Although it has been pointed out that the share of commuters switching from cars to public transport may not be very large (Hensher, 1998), increasing the service quality is still an important determinant of public transport demand (Beirao and Cabral, 2007). Travel time is often presented as the main determinant of trip characteristics. Much less focus has been devoted to trip reliability. However, some studies (see eg, Beirao and Cabral, 2007) have shown that users will shift to cars if public transport is not reliable enough. Several studies strongly suggest that reliability (understood as punctuality) of public transport is crucial to leverage the demand (Bates et al., 2001; Hensher et al., 2003; Paulley et al., 2006; Coulombel and de Palma, 2014). In a qualitative review, Redman et al. (2013) claim that reliability is the most important quality attribute of public transport according to users. Ongoing research

also tries to show that reliability of public transport may have an impact of the price of land.

The reliability issue does not only affect developing countries, but also developed countries. The example of United States is striking: only 77% of the short-haul trains are punctual, whereas 90% of European trains are on time (The Economist, 2011). Moreover, the reliability of long-distance trains is even worse in the US. This is due to decades of under-investment which have led to infrastructure degradation.¹ For a discussion of the relevance of investment in rail transit system, we refer the reader to Winston and Maheshri (2007).

Although there is a long tradition in studying road reliability, a sensitive lack of research is observed in public transport field (Bates et al., 2001). Studies highlight a valuation of road reliability (Bates et al., 2001; Fosgerau and Karlström, 2010), others underline the importance of public transport comfort (de Palma et al., 2013) or punctuality (Jensen, 1999), but few works deal with reliability in an analytical way.

This paper focuses on the two-way implication between punctuality level of public transport and (potential) customer behavior. Indeed, on the one hand the punctuality of public transport is a key element of the service quality. The user cost elements, which play an important role in demand analysis, are affected by the punctuality level (Bowman and Turnquist, 1981). The cost of punctuality differs among commuters. It largely depends on the preferred arrival time of commuters. As a consequence, users and potential users choose both the mode of transport and the departure time as a function of punctuality level in public transport. On the other hand, Mohring (1972) has shown that scheduled urban public transport is characterized by increasing returns to scale since the frequency increases with demand. Demand is influential in the service quality offered and the bus company may adapt its punctuality to the level of potential demand. Thus we show that some users may decide to arrive late at the bus stop when punctuality is too low. As a consequence, the bus company itself may become less strict as regards the punctuality. In a nutshell, this means that user behavior (punctuality of users) is influenced by the

¹To address this situation, Mr Obama plans to spend \$556 billion for transport over 6 years, according to his 2012 budget.

punctuality of public transport. This generates a vicious circle.

In this paper, we study three situations: (i) the reaction of the bus company when it faces a higher price of the alternative mode, (ii) the gap between the bus punctuality at equilibrium and at optimum and (iii) the equilibrium versus optimal modal split when punctuality matters.

We consider a duopoly which symbolizes a modal competition between public transport and another mode, which we call taxi. The attention is focused on the monetary impacts of punctuality. We simplify aspects related to engineering. A duopoly is used because determinants of demand for public transport are related to the demand for private transport (Balcombe et al., 2004). Two different types of variables are observed in the model: the public transport punctuality level, which is selected by the bus company and the prices set by the bus and taxi companies. Both have a substantial influence on demand for public transport (Paulley et al., 2006). Unreliability has a strong negative impact because it implies excessive waiting time and uncertainty (Wardman, 2004; Paulley et al., 2006).

Considering commuting trips, preferences can be analyzed with the dynamic scheduling model. In this model, individual's preferences reflect agents tradeoff between travel time, early schedule and late schedule delays. Commuters may choose different strategies to minimize their trip cost. This theory has been first introduced by Vickrey (1969) and then renewed by Arnott et al. (1990). Such analysis is usually specific to road analysis (Fosgerau and Karlström, 2010); here we introduce a waiting time to extend this model to public transport. The French State-owned railroad (SNCF) suggests to reschedule work arrival and departure times in order to reduce congestion (Steinmann, 2013). For the idea of endogenous schedules and private or public bus company, we refer the reader to Fosgerau and Small (2013).

Commuters are differentiated by their preferred arrival time at workplace and by their residential location which is measured as the time to travel to their destination when using the alternative mode. Two different preferred arrival time are considered and the location is uniformly distributed among commuters.

The analysis for the model proceeds in three steps. The first step consists in finding out the modal choice of commuters depending on prices and punctuality for the public transport

and the alternative mode. The second step determines which price and punctuality levels are set by companies at equilibrium given the behaviors of commuters identified in step one. The third step is to assess the prices and the punctuality level that minimize the total social cost and to compare these results with the ones derived in step two.

The paper is organized as follows. Section 2 describes the model and the commuter's strategies. Section 3 considers equilibrium and its properties. The gain due the transition from equilibrium to optimum is analyzed in Section 4. A numerical application is provided in Section 5 to illustrate our results. The final section concludes and proposes suggestions for further research.

2 Punctuality in public transport

Our model is based on the monocentric city framework defined by Alonso (1964), Mills (1967) and Muth (1969). All jobs are located in the center of the city, referred to as the central business district (CBD). Consequently, all commuters have to reach the CBD every morning. We focus our analysis on a unique radius of the city, assuming that this radius is representative of the set of radius of the city. We consider an unique road which coincides with this radius. It goes straight from the border of the city to the CBD. The radius is measured in time units and is Δ hours long. An unique bus line and a taxi company serve the CBD by using this road and bus stops are uniformly distributed along the radius of the city. We do not take into account congestion on the road. Thus both mode have the same speed and we refer to a bus stop located at δ hours from the CBD as "bus stop δ ". For example, the bus stop Δ is located at the border of the city. Similarly, all commuters live along the radius and we refer to commuters who need δ hours to reach the CBD, whether they use the bus service or the taxi service, as "commuters δ ". For each $\delta \in [0; \Delta]$, all commuters δ live at the same place (see Figure 3).

For analytical tractability, we consider a single bus. However this model can be easily adapted to other modes of public transport that run on a schedule. The bus is scheduled to arrive at the CBD at a given time, but it may be late. The lateness probability is not random: the bus company selects its quality service level and applies it in the same manner

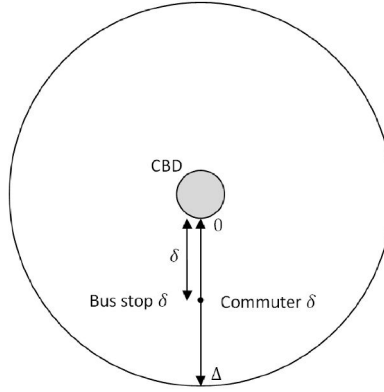


Figure 3: The monocentric city

along the radius. Thus when the bus company chooses to be late, it is late along the whole journey and its lateness is constant over time. Commuters are aware of the punctuality level and adapt their behavior accordingly. In particular they might arrive at the bus stop after the scheduled time even if there is a risk to miss the bus by doing so. This can occur rationally because there is a waiting cost for users. Commuters optimize their tradeoffs between waiting time cost, schedule delay cost and a cost corresponding to the use of an alternative mode, which is the taxi in our model. A commuter may either select *ex ante* the taxi or use the taxi if he misses the bus.

Table 1 presents important notations and their numerical values that will be used in Section 2 and Section 5. We first characterize the network and then the commuter behavior. Finally we present the modal split.

2.1 Transport supply

Bus stops are uniformly distributed between 0 and Δ . The bus is scheduled to arrive at its destination, the CBD, at time T . As there is no road congestion, it is also scheduled to serve the bus stop δ at time $T - \delta$ and also leaves at time $T - \delta$ i.e there is no transfer cost.² The bus company may choose that the bus is late and arrives at CBD time $T + x$. In this case, the bus stops at every bus stop δ at time $T + x - \delta$. The bus arrives at the CBD at time T with probability P and at time $T + x$ with probability $1 - P$ (Figure 4). Whatever the bus lateness, the total bus trip time is constant and equal to Δ . The potential lateness

²The loading time is assumed to be set to zero without loss of generality.

Table 1: Parameters values

Parameter	Comment	Suggested value
T	Scheduled arrival time	-
tt	Bus travel time	25/60 (hour)
x	Lateness	10/60 (hour)
$P \in [\frac{1}{2}; 1]$	Probability of the bus being late	-
$t_p \in \{T; T + x\}$	Arrival time at the bus stop of the bus	-
$\delta \in [0; \Delta]$	Taxi trip time	(hour)
Δ	Maximal taxi trip time	35/60 (hour)
$t^* \in \{T; T + x\}$	Preferred arrival time of users	-
$t_a \in \{T; T + x\}$	Arrival time at the bus stop of the user	-
$\theta \in [\frac{1}{2}; 1]$	Share of population in <i>Group A</i>	-
α_{bus}	In-bus time cost	15 (\$/hour)
α_{taxi}	In-taxi time cost	4 (\$/hour)
η	Waiting time cost	20 (\$/hour)
β	Early delay cost	10 (\$/hour)
γ	Late delay cost	30 (\$/hour)
κ	Bus fare	(\$)
τ	Taxi fare	(\$/hour)
c	Cost of punctuality (bus)	(\$)
d	Operating cost per unit of time (taxi)	40 (\$ /hour)

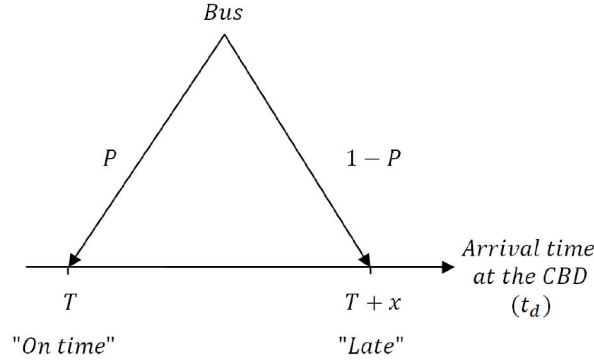


Figure 4: The arrival probabilities of the bus

is also constant and equal to x .

The probability of the bus being on time is endogenous: the bus company sets its level. It does not depend on traffic conditions, number of passengers or loading time. The worst quality of service occurs when the bus has the same probability of being on time and late. We assume that a regulator imposes this constraint to assure a consistent timetable.³ The “punctuality level” corresponds to the probability of the bus being on time.

Assumption 1. *The probability P of the bus being on time satisfies the following inequality:*

$$\frac{1}{2} \leq P \leq 1.$$

We assume that there is no capacity constraint in the bus. The bus fare, priced by the bus company, is κ for each passenger.

Commuters have access to an alternative mode of transport. In our model we consider this option as a taxi service, but it can also be walking or personal car use. The taxi company sets a fare τ which corresponds to the price charged per minute of travel.

2.2 Demand for bus and taxi

We consider two firms located in the CBD. Firm A employs a part of θ in commuters population and firm B a part of $(1 - \theta)$. The share of commuters working for firm A is bigger than the one working for firm B ($\theta \geq 1/2$). The workday in the first firm starts

³Minimal value of P is $1/2$, otherwise we would face another schedule than the expected one.

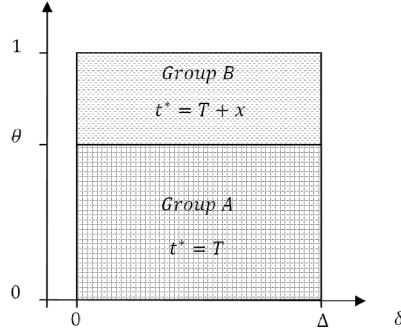


Figure 5: Distribution of taxi trip time in *Group A* and *Group B*

at time T whereas it starts at time $T + y$ in the other firm.⁴ This reflects the fact that even though a majority of commuters wishes to arrive at work place at the same time, all commuters have not the same preferred arrival time.

For tractability, we assume that the gap between the beginnings of workday equals to the lateness of the bus ($x = y$). Therefore, each type has a different preferred arrival time denoted $t^* \in \{T; T + x\}$. The first type of commuters (referred to as *Group A*) would rather arrive at time T , and the second one (referred to as *Group B*) at time $T + x$ (see Figure 5).

Commuters locations are uniformly distributed among each group in the same manner (Figure 5) and the distribution is assumed to have a support $[0; \Delta]$ so that $F(0) = 0$ and $F(\Delta) = 1$.

They are assumed to incur a schedule delay cost if traveling at time $t \neq t^*$. There is no transfer cost: commuters do not incur a cost by reaching the bus stop because bus stops are uniformly distributed along the radius where they live.

A commuter has the choice between catching the bus and using the taxi service. However he may miss the bus and then he has to use the taxi service. Indeed we assume the headway is so long that all users who miss the bus prefer to use the taxi service. If he tries to catch the bus, the commuter δ uses the bus stop δ because it minimizes its transfer cost. A commuter δ choosing to catch the bus bears the following schedule delay cost function that is assumed to depend on its arrival time at the bus stop, denoted $t_a \in \{T - \delta; T - \delta + x\}$,

⁴This gap between working start times is conceivable if there is no Marshallian externality between these two firms (see Henderson, 1997).

the arrival time of the bus, denoted $t_p \in \{T - \delta; T - \delta + x\}$, its most preferred trip time, denoted $t^* \in \{T; T + x\}$ as well as on the arrival time at destination of the bus, denoted $t_d \in \{T; T + x\}$:

$$CC_{bus} = \begin{cases} \kappa + \delta\alpha_{bus} + \eta(t_p - t_a) + \beta[t^* - t_d]^+ + \gamma[t_d - t^*]^+ & \text{if } (t_a \leq t_p), \\ \delta(\alpha_{taxi} + \tau) + \gamma[t_d - t^*]^+ & \text{if } (t_a > t_p), \end{cases}$$

with $[x]^+ = x$ if $x \geq 0$ and 0 if $x < 0$, κ the bus fare, α_{bus} the in-bus time cost, η the waiting time cost, β the early delay cost, γ the late delay cost, α_{taxi} the in-taxi time cost, τ the taxi fare and δ the trip time of commuter δ .

If a commuter chooses from the start to use the taxi service, he incurs the following cost:

$$CC_{taxi} = \delta(\alpha_{taxi} + \tau),$$

with α_{taxi} the taxi travel time value, τ the taxi fare and δ the taxi trip time.

By considering that the value of time in bus $\delta\alpha_{bus}$ is incurred by every commuter whatever is its choice, we can normalize the cost functions to:

$$CC_{bus} = \begin{cases} \kappa + \eta(t_p - t_a) + \beta[t^* - t_d]^+ + \gamma[t_d - t^*]^+ & \text{if } (t_a \leq t_p), \\ \delta(\check{\alpha} + \tau) + \gamma[t_d - t^*]^+ & \text{if } (t_a > t_p), \end{cases} \quad (1)$$

$$CC_{taxi} = \delta(\check{\alpha} + \tau), \quad (2)$$

with $\check{\alpha} = \alpha_{taxi} - \alpha_{bus}$.

Assumption 2. *The cost of waiting one minute for a bus, η , is lower than the cost of being one minute late, γ , and higher than the cost of being one minute early, β :*

$$\gamma \geq \eta \geq \beta.$$

This assumption is consistent with literature valuations (Wardman, 2004).

2.3 Commuters' strategies

Commuters dispose of three different strategies to minimize the cost of a trip. A strategy is defined by an arrival time at the bus stop. Arriving at the bus stop at time T corresponds to *Strategy O* (On-time at the bus stop), arriving at time $T + x$ to *Strategy L* (Late at the bus stop) and *Strategy T* (Taxi) embodies the decision to use the taxi and to not arrive at the bus stop. If a commuter chooses *Strategy O*, he waits until the bus arrives.

As a convention, we assume that a commuter who is indifferent between two strategies has a preference for maximizing its chance to get the bus. The commuter chooses:

Strategy O (arrive at time T) if $EC(O) \leq EC(T)$ and $EC(O) \leq EC(L)$;

Strategy L (arrive at time $T + x$) if $EC(L) < EC(O)$ and $EC(L) \leq EC(T)$;

Strategy T (choose the taxi) if $EC(T) < EC(O)$ and $EC(T) < EC(L)$;

where $EC(i)$ represents the expected cost of strategy i .

Proposition 1. *Under A.1 and A.2, the commuter δ in Group A selects:*

$$\text{Strategy O (time } T) \quad \text{if} \quad \delta \geq \delta_{T,O}^A,$$

$$\text{Strategy T (taxi)} \quad \text{if} \quad \delta < \delta_{T,O}^A,$$

where $\delta_{T,O}^A \equiv [\kappa + (1 - P)(\eta + \gamma)x] / (\check{\alpha} + \tau)$.

Proof. See Appendix A. □

For a commuter wishing to arrive at time T , *Strategy L* is never selected. Indeed a commuter chooses *Strategy L* instead of *Strategy T* if he prefers a late bus trip over a taxi trip. However such a commuter prefers an on time bus trip over taxi trip and consequently, he will choose *Strategy O*.

Proposition 2. *Under A.1 and A.2, the commuter δ in Group B selects:*

$$\text{Strategy O (time } T) \quad \text{if} \quad \delta \geq \delta_{L,O}^B,$$

$$\text{Strategy L (time } T + x) \quad \text{if} \quad \delta \in \left[\delta_{T,L}^B; \delta_{L,O}^B \right],$$

$$\text{Strategy T (taxi)} \quad \text{if} \quad \delta < \delta_{T,L}^B,$$

where $\delta_{L,O}^B \equiv [\kappa + (\frac{1-P}{P}\eta + \beta)x] / (\check{\alpha} + \tau)$ and $\delta_{T,L}^B \equiv \kappa / (\check{\alpha} + \tau)$.

Proof. See Appendix B. □

Strategy L is selected by some commuters from *Group B* unlike commuters from *Group A*. It can be explained by the fact that in this case, *Strategy L* corresponds to a possibility of the bus arriving on time. A commuter who prefers an on-time bus trip over a taxi trip, yet prefers a taxi trip more than an early-arrival bus trip, chooses *Strategy L*.

The share of commuters choosing *Strategy T* is independent of the probability of the bus being on time. Indeed P has no influence in the arbitrage between *Strategy L* and *Strategy T*. For commuters in *Group B*, choosing *Strategy L* is equivalent to choosing *Strategy T* except that they take the bus when it is late. Consequently, *Strategy L* is preferred to *Strategy T* as long as the cost of taking the bus when it is late is lower than the cost of taking a taxi. Then this arbitrage is independent of the probability of the bus being on time.

When the punctuality decreases, the share of commuters arriving late at the bus stop increases. The cut in the service quality makes the cost of *Strategy O* higher (because of A.2) and the cost of *Strategy L* smaller (except for commuters living so close to the CBD that a taxi trip is still cheaper than a bus trip, but we do not take account of these commuters because they still prefer *Strategy T*). Then among commuters who chose *Strategy O* before the service quality fall, those living the closest to the CBD were the most indifferent between both strategies and switched from *Strategy O* to *Strategy L*. The bus company may also itself become less strict, and generate a vicious circle.

When the taxi fare, τ , increases, more commuters choose to arrive at the bus stop at T and less commuters choose *Strategy L* and *Strategy T*. This is due to the fact that on the one hand some commuters have a bigger interest to minimize the probability of taking the taxi by shifting from *Strategy L* to *Strategy O* and from *Strategy T* to *Strategy L*. On the other hand, the shift from *Strategy L* to *Strategy O* is larger than the one from *Strategy T* to *Strategy L*.

Figure 6 illustrates these results. Other things being equal, the share of commuters arriving at T (and by doing so they are sure to catch the bus) among *Group A* increases from around 40% when $P = 1/2$ to almost 55% when $P = 1$. The share of commuters

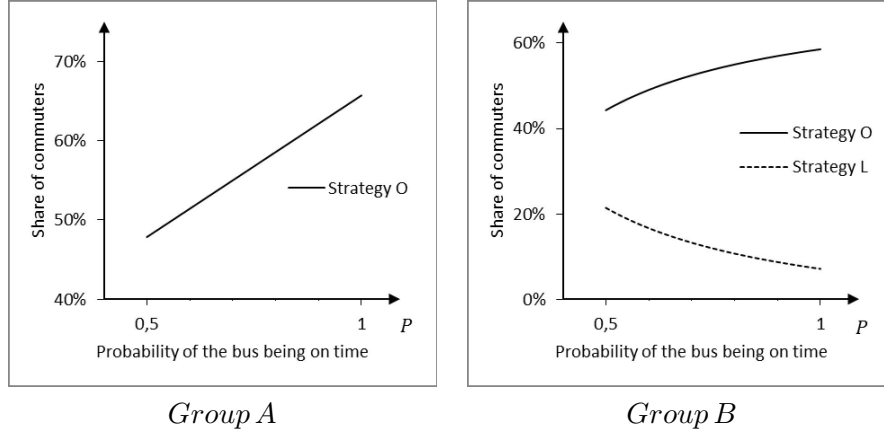


Figure 6: Share of commuter choosing *Strategy O* and *Strategy L* as a function of P , the probability of the bus being on time ($\kappa = 8$, $\tau = 50$)

in *Group B* choosing to arrive late at the bus stop (*Strategy L*) depends inversely on the probability of the bus being on time. If the bus arrives later, some users switch from *Strategy O* to *Strategy L* which leaves the bus company no incentive to restore the service quality.

Assumption 3. *The maximum cost of the taxi use, priced at the operating cost, is higher than the cost of the bus use, when priced at zero and when the bus arrives on time with probability 1/2:*

$$\Delta(\tilde{\alpha} + d) \geq \frac{1}{2}(\eta + \gamma)x.$$

Once the commuters strategy are defined, shares of commuters who are at the bus stop at time T or $T + x$ are known. Demands are described by

$$D_{bus} = \theta \left(1 - \frac{\delta_{T,O}^A}{\Delta} \right) + (1 - \theta) \left[1 - \frac{\delta_{L,O}^B}{\Delta} + (1 - P) \frac{\delta_{L,O}^B - \delta_{T,L}^B}{\Delta} \right], \quad (3a)$$

$$D_{taxi} = \theta \frac{\delta_{T,O}^A}{\Delta} + (1 - \theta) \left[P \frac{\delta_{L,O}^B - \delta_{T,L}^B}{\Delta} + \frac{\delta_{T,L}^B}{\Delta} \right]. \quad (3b)$$

Thus the bus (and taxi) patronage depends on the probability of the bus being on time. *Group A* is more sensitive to the service quality than *Group B* (see also Figure 6). This is due to the fact that commuters from *Group A* incur late arrival costs while commuters from *Group B* incur early arrival costs and, as seen in A.2, the penalty for lateness is much higher than the penalty for arriving early at the destination.

3 Competition between bus and taxi companies

In this section, we explore equilibrium pricing and punctuality level in a duopoly competition. We assume that following condition holds:

$$\Delta (\check{\alpha} + \tau) \geq \kappa + \frac{1}{2} (\eta + \gamma) x. \quad (4)$$

This condition assures that the price selected by the bus company is low enough to preserve a demand for bus trips. Thus the demand functions formulation (equations (3a) and (3b)) is still correct. We will check if it holds once the equilibrium values of τ and κ are solved.

Both companies incur a cost. The cost incurred by the bus company depends on the punctuality level and is assumed to be quadratic. It is a sunk cost in the sense of being unrecoverable (Sutton, 1991). The cost of the taxi company linearly depends on the total travel time and can be viewed as an operating cost:

$$Cost_{bus} = \frac{c}{2} P^2, \quad (5a)$$

$$Cost_{taxi} = d \times TTT, \quad (5b)$$

with c the punctuality cost, d a cost per hour traveled and TTT the total travel time of the taxi company.

The bus company chooses the bus fare κ and the punctuality level P , so as to maximize its expected profit. From equations (3a) and (5a), the bus company profit can be written as

$$\Pi_{bus} = \kappa D_{bus} - \frac{c}{2} P^2.$$

There exists a unique solution⁵ satisfying the first-order conditions $\partial \Pi_{bus} / \partial \kappa = 0$ and

⁵Second-order conditions are satisfied as $\partial^2 \Pi_{bus} / \partial \kappa^2 = -2 / \check{\Delta}$ and $\partial^2 \Pi_{bus} / \partial P^2 = -c / \check{\Delta}$. The Hessian matrix of second partial derivatives is also negative definite, and the solution is a global maximum.

$\partial \Pi_{bus} / \partial P = 0$, given by

$$\kappa^e = \frac{1}{2} (\check{\Delta}^e - \Gamma^e x), \quad (6a)$$

$$P^e = \begin{cases} \frac{1}{2} & \text{if } c > c_2^e, \\ \frac{\kappa^e \check{\eta} x}{c \check{\Delta}^e} & \text{if } c \in [c_1^e; c_2^e], \\ 1 & \text{if } c < c_1^e, \end{cases} \quad (6b)$$

where $\check{\Delta}^e = \Delta (\check{\alpha} + \tau^e)$, $\Gamma^e = (1 - P^e) \eta + (1 - \theta) P^e \beta + \theta (1 - P^e) \gamma$, $\check{\eta} = \eta - (1 - \theta) \beta + \theta \gamma$, $c_1^e \equiv \kappa^e \check{\eta} x / \check{\Delta}^e$ and where $c_2^e \equiv 2c_1^e$.

The price of a minute traveled in a taxi, τ , is set by the taxi company to maximize its profit. From equations (3b) and (5b), taxi profit is given by

$$\begin{aligned} \Pi_{taxi} = (\tau - d) & \left[\theta \int_0^{\delta_{T,O}^A} \delta f(\delta) d\delta \right. \\ & \left. + (1 - \theta) \left(\int_0^{\delta_{T,L}^B} \delta f(\delta) d\delta + P \int_{\delta_{T,L}^B}^{\delta_{L,O}^B} \delta f(\delta) d\delta \right) \right]. \end{aligned}$$

The level of price satisfying the first-order condition⁶ $\partial \Pi_{taxi} / \partial \tau = 0$ is

$$\tau^e = \check{\alpha} + 2d. \quad (7)$$

Condition (4) requires $\check{\Delta}^e \geq \kappa^e + \frac{1}{2} (\eta + \gamma) x$ and yet

$\Delta (\check{\alpha} + d) \geq \{P\eta - (1 - \theta) P\beta + [1 - \theta (1 - P)] \gamma\} x / 2$. It holds according to A.3.

Note that the probability of the bus being on time in (6b) is continuous.

The core component of the bus fare corresponds to the average taxi trip cost cut by the average schedule and waiting time cost incurred by commuters. The bus company takes account of its service quality to remain attractive regarding the alternative mode. As expected, the punctuality decreases when the punctuality cost c increases. Since the punctuality level decreases with the maximal taxi trip time, Δ , a high scatter of commuter's locations makes the service quality regress (see equation (6b)). In addition, the longer of

⁶Second-order condition requires that $4\check{\alpha} - 2\tau^e + 6d \geq 0$ or $\tau^e \leq 2\check{\alpha} + 3d$.

the radius where commuters live, the higher is the mark-up for the bus company. The taxi fare is independent of the bus company choices. It only depends on the values of taxi and bus travel time and operating cost. *Ceteris paribus* when d increases, both bus and taxi fares become higher.

There is a unique simultaneous Nash equilibrium which is given by equations (6a), (6b) and (7).

Proposition 3. *At equilibrium, P^e , the probability of the bus being on time and κ^e , the bus fare, increase with τ , the taxi fare.*

Proof. See Appendix C. □

Consider an initial rise in taxi fare, τ^e , for example due to an increase in the taxi operating cost or in the petrol price. This increase leads to a standard modal shift from taxi service to bus service, other things being equal (see Propositions 1 and 2). Consequently, the cost of the bus punctuality per user decreases. The bus company therefore will have an incentive to increase the punctuality level when τ rises. By doing so, the bus company attracts additional commuters. In this model, an increase in bus patronage improves the service quality of the bus. This can be viewed as an extension of the Mohring Effect (Mohring, 1972) according to which the service quality measured as the frequency increases when the demand for public transport rises.

The increase in the bus fare is explained by two aspects: on the one hand the service quality has been improved, and on the other hand, the rises in the taxi fare increase the average taxi trip cost and therefore the bus fare. There is no strategic complementarity because the taxi company does not react to a change in bus fare (see Vives, 1990).

4 Welfare analysis

Welfare is the sum of the aggregate commuter surplus and the companies profits. Since a cost function is used instead of a surplus function to study the commuter strategies, the social welfare function is defined as the opposite of the social cost function SC which is the difference between aggregate commuter costs and firm profits. From equations of

commuter cost (1) and (2), of demand (3a) and (3b), and of companies cost (5a) and (5b), the social cost function can be written as

$$SC = \frac{\Delta \alpha_{bus}}{2} + \theta CC_{\theta=1} + (1 - \theta) CC_{\theta=0} - \Pi_{bus} - \Pi_{taxi},$$

where

$$\begin{aligned} CC_{\theta=1} &= (\check{\alpha} + \tau) \int_0^{\delta_{T,O}^A} \delta f(\delta) d\delta \\ &\quad + \{\kappa + [(1 - P)(\eta + \gamma)]x\} \int_{\delta_{T,O}^A}^{\Delta} f(\delta) d\delta, \end{aligned}$$

and where

$$\begin{aligned} CC_{\theta=0} &= (\check{\alpha} + \tau) \int_0^{\delta_{T,L}^B} \delta f(\delta) d\delta \\ &\quad + \int_{\delta_{T,L}^B}^{\delta_{L,O}^B} [(1 - P)\kappa + P(\check{\alpha} + \tau)\delta] f(\delta) d\delta \\ &\quad + \{\kappa + [(1 - P)\eta + P\beta]x\} \int_{\delta_{L,O}^B}^{\Delta} f(\delta) d\delta. \end{aligned}$$

The social planner chooses the punctuality level P , the bus fare κ and taxi fare τ so as to minimize social cost. The first-order conditions for the socially optimal bus and taxi prices are given by

$$\kappa^o = 0, \tag{8a}$$

$$\tau^o = d. \tag{8b}$$

As expected, optimal bus and taxi fares equal to the marginal costs incurred by bus and taxi companies. Indeed, as there is no variable cost for the bus, the optimal bus fare is null.

The expression of the optimal punctuality level P^o is not explicit in the general case because the equation to solve is a cube root i.e it has three solutions with only one real.

$$P^o = \arg \min_{P \in [\frac{1}{2}; 1]} SC. \quad (9)$$

However in the extreme case where $\theta = 1$, there exists a unique solution⁷ satisfying $\partial SC_{\theta=1} / \partial P = 0$. By using (8a) and (8b), we obtain, for *Group A*

$$P_{\theta=1}^o = \begin{cases} \frac{1}{2} & \text{if } c > c_{2;\theta=1}^o, \\ \frac{[\check{\Delta}^o - (\eta + \gamma)x](\eta + \gamma)x}{c\check{\Delta}^o - [(\eta + \gamma)x]^2} & \text{if } c \in [c_{1;\theta=1}^o; c_{2;\theta=1}^o], \\ 1 & \text{if } c < c_{1;\theta=1}^o, \end{cases} \quad (9a)$$

where $\check{\Delta}^o = \Delta(\check{\alpha} + \tau^o)$, $c_{1;\theta=1}^o \equiv (\eta + \gamma)x$, $c_{2;\theta=1}^o \equiv [2\check{\Delta}^o - (\eta + \gamma)x](\eta + \gamma)x / \check{\Delta}^o$ and where $c_{1;\theta=1}^o < c_{2;\theta=1}^o$. Note that the probability of the bus being on time when $\theta = 1$ is continuous.

We generalize the above result to the other extreme case where $\theta = 0$ in the following conjecture.

Conjecture 1. *For Group B ($\theta = 0$), the punctuality level of the bus $P_{\theta=0}^o$ weakly decreases when the cost of reliability c increases. There are two critical values of c , $c_{1;\theta=0}^o$ and $c_{2;\theta=0}^o$ with $c_{1;\theta=0}^o \leq c_{2;\theta=0}^o$ such that:*

$$P_{\theta=0}^o = \begin{cases} \frac{1}{2} & \text{if } c > c_{2;\theta=0}^o, \\ 1 & \text{if } c < c_{1;\theta=0}^o. \end{cases} \quad (9b)$$

with $c_{1;\theta=0}^o < c_{2;\theta=0}^o$.

Equations (8a), (8b) and (9) provide the values at optimum in the general case. Equations (9b) and (9a) point out the optimal punctuality level in extreme cases.

The optimal probability of the bus being on time has the same properties we describe in Section 3: it decreases when the punctuality cost c or the travel time of the commuter

⁷Second-order condition is verified as $c\check{\Delta}^o \geq [(\eta + \gamma)x]^2$.

living the farthest Δ increases. The important observation is that the optimal probability of the bus being on time does not necessarily equal 1. It may be lower than 1 and even equal $1/2$ under some conditions. Critical values $c_{1;\theta=0}^o$ and $c_{2;\theta=0}^o$ are expected because $P_{\theta=0}^o \in [1/2; 1]$. The above conjecture is illustrated in Figure 7.

From now on, as the expression of P^o is not explicit and $P^o = \theta P_{\theta=1}^o + (1 - \theta) P_{\theta=0}^o$, properties of the optimal probability of the bus being on time will be addressed separately according to the structure of the population. The two extreme cases $\theta = 1$ and $\theta = 0$ are highlighted, even if $\theta \geq 1/2$.

Proposition 4. *For Group A ($\theta = 1$), the punctuality level of the bus is higher at optimum than at equilibrium.*

Proof. See Appendix E. □

Commuters in *Group A* want to arrive at T therefore the later is the bus, the more commuters incur a cost. The bus company wishes to maximize the probability of the bus being on time at equilibrium, as the social planner does at optimum, while taking into account the punctuality cost per user incurred by the bus company. The difference between equilibrium and optimum bus punctuality is mainly explained by a price-effect. Indeed, the gap between the bus fare relative to the taxi fare is much higher at equilibrium than at optimum. Thus other things being equal, the bus company attracts less customers at equilibrium than at optimum. Consequently, the bus company has to reduce the bus punctuality at equilibrium more than the social planner does at optimum to keep the punctuality cost per user small enough. This result is summarized in Proposition 4.

As there is no explicit expression for P^o and $P_{\theta=0}^o$, a discussion with a figure is provided in Section 5.

Proposition 5. *For Group A ($\theta = 1$), if the taxi operating cost d is higher than d_1^c , the bus patronage is higher at optimum than at equilibrium.*

When $d \leq d_1^c$, the bus patronage is higher at optimum than at equilibrium if and only if the cost of punctuality for the bus company is small enough ($c \leq c_1^c$).⁸

⁸The critical value of the taxi operating cost d is $d_1^c = \frac{3(\eta+\gamma)x}{8\Delta} - \check{\alpha}$. The critical value of the punctuality cost c_1^c is defined as the unique solution of $D_{\theta=1}^o = D_{\theta=1}^e$.

Proof. See Appendix F. □

As the expression of $P_{\theta=0}^o$ is not explicit, the analysis is more difficult for *Group B*. However we formulate a proposition, as well as a conjecture.

Proposition 6. *For Group B ($\theta = 0$), if the taxi operating cost d is smaller than d_2^c (higher than d_3^c , resp.), the bus patronage is smaller (resp. higher) at optimum than at equilibrium.⁹*

Proof. See Appendix G. □

We conjecture the variations in demand for *Group B* when $d \in [d_2^c; d_3^c]$.

Conjecture 2. *For Group B, when $d \in [d_2^c; d_3^c]$, the bus patronage is higher at optimum than at equilibrium if the punctuality cost for the bus company is small enough ($c > c_2^c$).¹⁰*

This conjecture is discussed in Appendix H. The basic idea in Propositions 5 and 6 and in Conjecture 2 is that when the taxi operating cost is small, the bus company tends to underprice which consequently attracts too many customers. As the taxi operating cost is high, the bus company overprices. This is due to the fact that the bus fare highly depends on the taxi fare (see equation (6a)). We refer the reader to Proost et al. (2002) for a detailed discussion of the optimal pricing in transport.

The equilibrium modal split meets the optimal modal split under two conditions. First the taxi operating cost d has to be included between the two critical values we defined. Then the punctuality cost incurred by the bus company c has to equal a critical value. If the taxi operating cost is higher than the interval defined by critical values, the optimal modal split is reached by a partial commuters shift from taking a taxi to taking a bus. This shift can also be in the opposite direction if the taxi operating cost is smaller than the critical interval. This reflect the fact that the bus company underprovides quality relative to the social optimum when c is small (see De Borger and Van Dender, 2006, for a detailed discussion).

⁹The critical values of the taxi operation cost d are $d_2^c = -(\frac{1}{2}\eta - \frac{7}{2}\beta)x/2\Delta - \alpha_{taxi}$ and $d_3^c = (2\eta + \beta)x/2\Delta - \alpha_{taxi}$, with $d_2^c < d_3^c$.

¹⁰The critical value of reliability c_2^c is the unique solution of $D_{\theta=0}^o = D_{\theta=0}^e$.

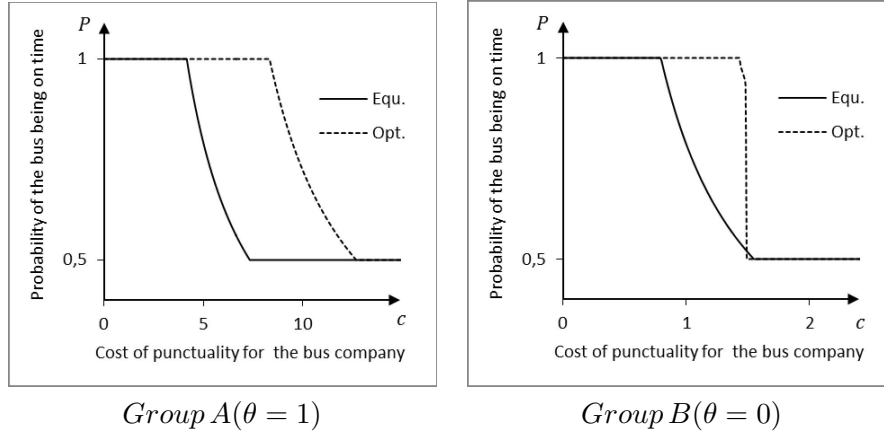


Figure 7: Probability of the bus being on time as a function of the punctuality cost c

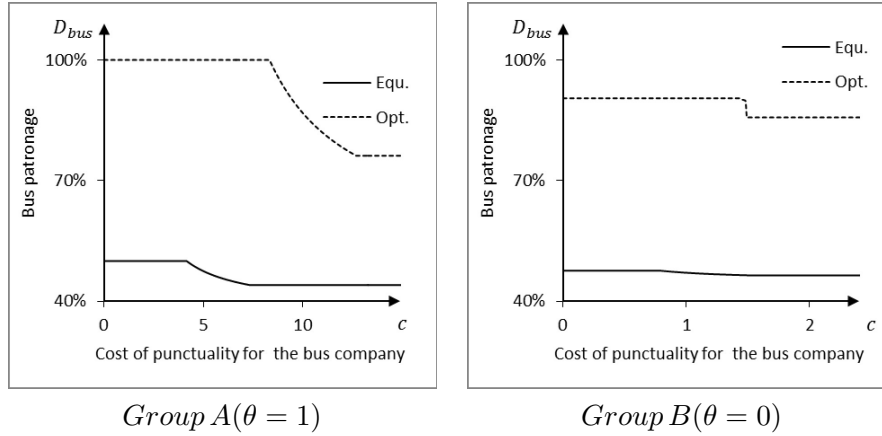
The taxi operating cost corresponds to the traditional costs as fuel or insurance, but it may also be viewed as an extra tax set by the planner to account for the externalities such as pollution or noise.¹¹ In this sense, the operating cost trend should be growing and in the long run, the bus patronage would increase at the expense of the taxi service.

5 Numerical application

We develop an applied case to illustrate previous theoretical findings. Numerical results are obtained with values specified in Table 1. The studied case is related to a 25 minutes bus trip. The bus has a probability P of being on time and a probability $1 - P$ of being 10 minutes late at departure. The commuter living the farthest from their trip destination has a taxi trip time equal to 35 minutes. We consider a uniform distribution of the taxi trip time. The operating taxi cost d is constant and equal to 40 \$/hour. Lastly, cost parameters α_{bus} , α_{taxi} , η , β and γ are equal to 15, 4, 20, 10 and 30 \$/hour, resp. Each variable is drawn depending on the reliability cost for the bus c .

A reminder to the readers, P^e and P^o are respectively the probability of the bus being on time at equilibrium and at optimum. As expected, the probability of the bus being on time decreases when the reliability cost increases (see Figure 7). The more expensive the punctuality is, the less interesting is the reliability for both the bus company and the

¹¹See Proost and Van Dender (2001) for an evaluation of alternative fuel efficiency, environmental and transport policies regarding atmospheric pollution.

Figure 8: Bus patronage as a function of the punctuality cost c

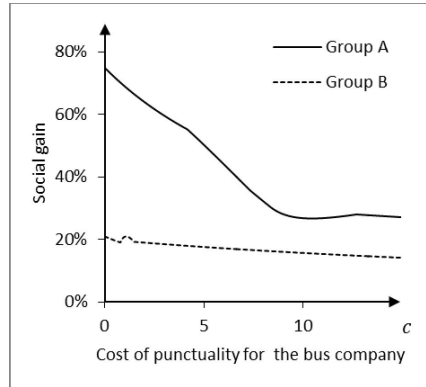
social planner.

As indicated in Proposition 4, the probability of the bus being on time when $\theta = 1$ is higher at optimum than at equilibrium. The opposite extreme case where $\theta = 0$ is more complex as P^o is not continuous. It seems that the probability of the bus being on time is higher at optimum than at equilibrium when c is small and that after a critical value of c this relation is inverted. Probabilities of the bus being on time are higher when $\theta = 1$ than when $\theta = 0$. This is due to the fact that users from *Group A* are more sensitive to unreliability because when the bus is late they incur a late delay cost which is higher than the waiting time cost incurred by commuters from *Group B*. Thus when $\theta = 1$, the bus company needs to maintain a better level of service than when $\theta = 0$ in order to keep their patrons. An important observation is that the optimal punctuality may be very low and even equal to 0.5 which is the worst reliability level. Indeed, since the reliability cost is not too high, the social planner makes the bus company increase the punctuality of the bus to minimize the cost born by users. However, if the punctuality cost for the bus is too high, it is socially better to share cost with users by making or allowing the bus to be late.

Two points are especially interesting in Figure 8. First, the bus patronage is weakly decreasing when the punctuality cost increases. This drop is higher at optimum than at equilibrium. Along with Figure 7 we note that the punctuality has a strong effect on demand. The variations of the bus patronage corresponds to the variations of the bus punctuality. When the bus punctuality is stable, the split between the bus and the taxi

Table 2: Values of main variables when $c = 5$ and $\theta = 0.75$

Variable	Equilibrium	Optimum
Probability of the bus being on time	0.72	0.88
Bus fare	16.9\$	0\$
Taxi fare	70\$/hour	40\$/hour
Bus patronage	47%	96%
Social gain	-	42%

Figure 9: Relative social gain compared to equilibrium as a function of the punctuality cost c

is constant. Secondly, note that the demand for the bus is higher at optimum than at equilibrium in both extreme cases. Regarding Proposition 5, this example illustrates the common case where the bus patronage is higher at optimum than at equilibrium. At equilibrium the bus patronage is sub-optimal. Too much commuters use the taxi service because catching the bus is too expensive and the bus is not reliable enough.

Table 2 provides the values of main variables when $c = 5$ and $\theta = 0.75$. The probability of the bus being on time at equilibrium equals to 0.72 (note that this measure is consistent with observed average lateness in Paris Area (STIF, 2014)). The bus fare may seem high, but it is not surprising as we do not take into account subsidies which are important in the public transport sector (Ponti, 2011). Indeed for example in Paris Area in 2010, monetary public transport revenues equal to 29.7% of total operating cost (STIF, 2013). The optimum is reached by increasing the reliability at its maximal level and decreasing prices. Consequently, the bus patronage becomes much higher and the social gain is about 42%.

The relative social gain is computed as the ratio of the absolute gain, due to the transition from equilibrium to optimum, to the absolute social cost at equilibrium (see Figure 9). Such curves allow to determine when the gain is high enough to justify public intervention: the lower the punctuality cost is, the more useful is public intervention. Indeed when c is high (see equations 6b, 9a and 9b), punctuality at equilibrium and at optimum is the same. The only difference between equilibrium and optimum is the modal split, but the gain due to this difference is gradually offset by the growing punctuality cost. Consequently, for both *Group A* and *Group B*, the social gain tends to 0 when the punctuality cost, c , tends to infinity. The cut in social gain is faster for *Group A* because variation in patronage is more sensitive with respect to the rise in c .

The brief application in this section illustrates that the effectiveness of public intervention varies according to punctuality cost. In the more general and realistic case, a stronger intervention seems useful in relation to the current situation.

6 Conclusion

The modeling of the bus punctuality reported here has provided an improved understanding of the two-way implication between punctuality level of public transport and customer public transport use. Commuters develop adaptive strategies to fit the transport system. Thus a rise in the fare of a mode decreases the patronage for this mode. In particular, an increase in the taxi fare rises the share of commuters arriving on time at the bus stop because they wish to minimize the probability of missing the bus. Moreover when the bus company becomes less strict as regards punctuality, more bus users will prefer to arrive late at the bus stop. Then the bus company is not incited to maintain a high level of reliability. This can generate a vicious circle. We also appreciate the efficiency of the punctuality when it is viewed as an instrument of service quality that can be adapted to fit and regulate the public transport patronage.

The main findings of this paper follow. At equilibrium, the probability of the bus being on time increases with the price of the alternative mode. The service quality reacts well to a rise in the taxi fare. Indeed, a new market share of commuters is assailable with a

reasonable effort in terms of service quality. Compared with the optimum, buses are very often too late at equilibrium. Commuters bear the cost of this extra-lateness, because they have to wait for the bus more often or take the taxi which is expensive. However, it does not mean that the bus should not be late. Indeed if the cost of the punctuality is too high relative to the cost of the alternative mode, a late bus is socially preferable. Finally, we find that the sign and the amplitude of the gap between the equilibrium and optimal modal split first depends on the cost of the alternative mode and secondly on the punctuality cost incurred by the bus company. Nevertheless, in the more general and realistic case the bus patronage seems under-optimal.

Several elements remain to be addressed. Considering risk averse users would change users strategies and affect the punctuality. It should be interesting to include congestion on road networks and in the bus. Congestion on the road would make the taxi journey longer and unpredictable, whereas congestion in the bus (understood as crowding) would accentuate the cost incurred by users. Finally, introducing the bus punctuality in a bus transit line with several stops and several buses (de Palma and Lindsey, 2001) will improve the modeling by introducing a snowball effect: if a bus is late, its lateness increases along its journey.

Appendices

Appendix A Proof of Proposition 1

We wish to compare expected costs of *Strategys* O , L and T , denoted $EC(O)$, $EC(L)$ and $EC(T)$ respectively, to define a choice rule for a commuter in *Group A*. From equations (1) and (2), we can write:

$$EC(O) = \kappa + (1 - P)(\eta + \gamma)x,$$

$$EC(L) = P\delta(\check{\alpha} + \tau) + (1 - P)(\kappa + \gamma x),$$

$$EC(T) = \delta(\check{\alpha} + \tau).$$

Therefore we have

$$EC(O) \leq EC(L) \text{ iff } \delta \geq \frac{\kappa + \left(\frac{1-P}{P}\right)\eta x}{\check{\alpha} + \tau} \equiv \delta_{L,O}^A, \quad (10)$$

$$EC(T) < EC(L) \text{ iff } \delta < \frac{\kappa + \gamma x}{\check{\alpha} + \tau} \equiv \delta_{T,L}^A, \quad (11)$$

$$EC(T) < EC(O) \text{ iff } \delta < \frac{\kappa + (1 - P)(\eta + \gamma)x}{(\check{\alpha} + \tau)} \equiv \delta_{T,O}^A. \quad (12)$$

We use A.1 and A.2 to rank $\delta_{L,O}^A$, $\delta_{T,L}^A$ and $\delta_{T,O}^A$:

(i) $\delta_{T,L}^A \geq \delta_{L,O}^A \iff \gamma \geq \left(\frac{1-P}{P}\right)\eta \iff \frac{\gamma}{\eta} \geq \frac{1-P}{P}$ which is true since $\frac{\gamma}{\eta} > 1 \geq \frac{1-P}{P}$ by A.1 and A.2.

(ii) $\delta_{T,L}^A \geq \delta_{T,O}^A \iff \gamma \geq (1 - P)(\eta + \gamma) \iff P \geq 1 - \frac{\gamma}{(\eta + \gamma)}$ which is true since

$\frac{\gamma}{(\eta+\gamma)} > \frac{1}{2}$ by A.2.

(iii) $\delta_{T,O}^A \geq \delta_{L,O}^A \iff \eta + \gamma \geq \frac{1}{P}\eta \iff P \geq \frac{\eta}{(\eta+\gamma)}$ which is true since $\frac{\eta}{(\eta+\gamma)} < \frac{1}{2}$ by A.2.

Therefore $\delta_{L,O}^A \leq \delta_{T,O}^A \leq \delta_{T,L}^A$.

Strategy B is chosen if and only if $\delta < \delta_{L,O}^A$ and $\delta \geq \delta_{T,L}^A$. As $\delta_{L,O}^A \leq \delta_{T,L}^A$, *Strategy B* is dominated and never chosen by commuter in *Group A*. Figure 10 summarizes results of the proof.

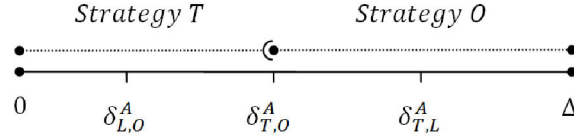


Figure 10: Strategy choice of a commuter in *Group A* depending on the taxi trip time δ .

Appendix B Proof of Proposition 2

We wish to compare expected costs of *Strategy O*, *L* and *T*, denoted $EC(O)$, $EC(L)$ and $EC(T)$ respectively, to define a choice rule for a commuter in *Group B*. From equations (1) and (2), we can write:

$$EC(O) = \kappa + [(1 - P)\eta + P\beta]x,$$

$$EC(L) = P\delta(\check{\alpha} + \tau) + (1 - P)\kappa,$$

$$EC(T) = \delta(\check{\alpha} + \tau).$$

Therefore we have

$$EC(O) \leq EC(L) \text{ iff } \delta \geq \frac{\kappa + \left(\frac{1-P}{P}\eta + \beta\right)x}{\check{\alpha} + \tau} \equiv \delta_{L,O}^B, \quad (13)$$

$$EC(T) < EC(L) \text{ iff } \delta < \frac{\kappa}{\check{\alpha} + \tau} \equiv \delta_{T,L}^B, \quad (14)$$

$$EC(T) < EC(O) \text{ iff } \delta < \frac{\kappa + [(1-P)\eta + P\beta]x}{\check{\alpha} + \tau} \equiv \delta_{T,O}^B. \quad (15)$$

We use A.1 and A.2 to rank $\delta_{L,O}^B$, $\delta_{T,L}^B$ and $\delta_{T,O}^B$:

$$(i) \delta_{T,L}^B \leq \delta_{L,O}^B \iff 0 \leq \left(\frac{1-P}{P}\eta + \beta\right) \text{ which is true.}$$

$$(ii) \delta_{T,L}^B \leq \delta_{T,O}^B \iff 0 \leq (1-P)\eta + P\beta \text{ which is true.}$$

$$(iii) \delta_{T,O}^B \leq \delta_{L,O}^B \iff (1-P)\eta + P\beta \leq \frac{1-P}{P}\eta + \beta \iff 1 \geq \frac{1}{P} \text{ which is true since } \frac{1}{P} \geq 1 \text{ by A.1.}$$

Therefore $\delta_{T,L}^B \leq \delta_{T,O}^B \leq \delta_{L,O}^B$. Figure 11 summarizes results of the proof.

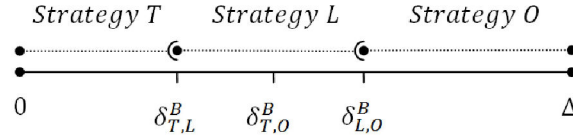


Figure 11: Strategy choice of a commuter in *Group B* depending on the taxi trip time δ .

Appendix C Proof of Proposition 3

We wish to show that P^e , the probability of the bus being on time at equilibrium, and κ^e , the bus fare, increase with τ the taxi fare. We first show that $\partial P^e / \partial \tau \geq 0$ (i) and that $\partial \kappa^e / \partial \tau \geq 0$ (ii). Then we check that boundaries of interval, c_1^e (iii) and c_2^e (iv), increase with τ . Let us recall expressions of equilibrium variables (see equations (6a) and (6b)):

$$\kappa^e = \frac{1}{2} \{ \check{\Delta}^e - [(1-P^e)\eta + (1-\theta)P^e\beta + \theta(1-P^e)\gamma]x \},$$

$$P^e = \begin{cases} \frac{1}{2} & \text{if } c > c_2^e, \\ \frac{\kappa^e \check{\eta} x}{c \check{\Delta}^e} & \text{if } c \in [c_1^e; c_2^e], \\ 1 & \text{if } c < c_1^e, \end{cases}$$

where $\check{\Delta}^e = \Delta(\check{\alpha} + \tau^e)$, $\check{\eta} = \eta - (1 - \theta)\beta + \theta\gamma$, $c_1^e \equiv \kappa^e \check{\eta}x / \check{\Delta}^e$ and where $c_2^e \equiv 2c_1^e$. By substituting κ^e in P^e , we obtain

$$P^e = \begin{cases} \frac{1}{2} & \text{if } c > c_2^e, \\ \frac{[\check{\Delta}^e - (\eta + \theta\gamma)x] \check{\eta}x}{2c\check{\Delta}^e - (\check{\eta}x)^2} & \text{if } c \in [c_1^e; c_2^e], \\ 1 & \text{if } c < c_1^e, \end{cases}$$

where $c_1^e = [\check{\Delta}^e - (1 - \theta)\beta x] \check{\eta}x / 2\check{\Delta}^e$ and where $c_2^e = (\check{\Delta}^e - \check{\eta}x) \check{\eta}x / 2\check{\Delta}^e$. We now derive P^e , κ^e , c_1^e and c_2^e on τ^e .

(i) $\frac{\partial P^e}{\partial \tau^e} = \Delta \check{\eta}x \frac{2c(\eta + \theta\gamma)x - (\check{\eta}x)^2}{[2c\check{\Delta}^e - (\check{\eta}x)^2]^2}$ so $\frac{\partial P^e}{\partial \tau^e} \geq 0$ if $c \geq \frac{(\check{\eta}x)^2}{2(\eta + \theta\gamma)x}$. Let us substitute c by c_1^e the minimal value of the interval $[c_1^e; c_2^e]$. Thus

$$c_1^e - \frac{(\check{\eta}x)^2}{2(\eta + \theta\gamma)x} = \frac{\check{\eta}x [\check{\Delta}^e - (\eta + \theta\gamma)x] (1 - \theta)\beta x}{2\check{\Delta}^e (\eta + \theta\gamma)x}.$$

Yet $\frac{\check{\eta}x(1-\theta)\beta x}{2\check{\Delta}^e(\eta+\theta\gamma)x} \geq 0$ and $\check{\Delta}^e - (\eta + \theta\gamma)x \geq 0$ by A.3. We therefore have $\frac{\partial P^e}{\partial \tau^e} \geq 0$;

$$(ii) \frac{\partial \kappa^e}{\partial \tau^e} = \frac{\Delta}{2} + \frac{\partial P^e}{\partial \tau^e} \frac{\check{\eta}x}{2} \geq 0 \text{ by A.2,}$$

$$(iii) \frac{\partial c_1^e}{\partial \tau^e} = \frac{(1-\theta)\beta x \check{\eta}x}{2\Delta(\alpha_{taxi} + \tau^e)^2} \geq 0,$$

$$(iv) \frac{\partial c_2^e}{\partial \tau^e} = \frac{(\check{\eta}x)^2}{2\Delta(\alpha_{taxi} + \tau^e)^2} \geq 0.$$

P^e , the probability of the bus being on time at equilibrium, and κ^e , the bus fare, increase well with τ^e the taxi fare.

Appendix D Proof: optimal bus and taxi fare

The social planner chooses the punctuality level P , the bus fare κ and taxi fare τ so as to minimize social cost. The first-order conditions for the socially optimal bus and taxi prices

are given by

$$\frac{\partial SC}{\partial \kappa} = \frac{\kappa(\check{\alpha} + d) - (\tau - d)\Gamma x}{\Delta(\check{\alpha} + \tau)^2} = 0, \quad (16a)$$

$$\frac{\partial SC}{\partial \tau} = \frac{(\tau - d)A - \kappa(\check{\alpha} + d)(\kappa + \Gamma x)}{\Delta(\check{\alpha} + \tau)^3} = 0, \quad (16b)$$

where $\Gamma = (1 - P)\eta + (1 - \theta)P\beta + \theta(1 - P)\gamma$, $A = \kappa\Gamma x + \chi$ and

$\chi = \theta[(1 - P)(\eta + \gamma)x]^2 + (1 - \theta)P[(\frac{1-P}{P}\eta + \beta)x]^2$. Therefore from (16a) and (16b)

$$\kappa^o = \frac{(\tau^o - d)\Gamma x}{(\check{\alpha} + d)}, \quad (17a)$$

$$\tau^o = \frac{\kappa^o(\check{\alpha} + d)(\kappa^o + \Gamma x)}{A\Delta(\check{\alpha} + \tau^o)^3} - d. \quad (17b)$$

By substituting (17a) into (17b), the first-best optimal bus and taxi prices can be written as¹²

$$\kappa^o = 0;$$

$$\tau^o = d.$$

Appendix E Proof of Proposition 4

We wish to show that the probability of the bus being on time is higher in the optimal situation than in equilibrium when $\theta = 1$. For that, we need to show that the result of $P_{\theta=1}^o - P_{\theta=1}^e$ is positive (i) and that the limits of the variation intervals are well sorted i.e. $c_{1;\theta=1}^o \geq c_{1;\theta=1}^e$ (ii) and $c_{2;\theta=1}^o \geq c_{2;\theta=1}^e$ (iii):

$$(i) P_{\theta=1}^o - P_{\theta=1}^e = \frac{c\check{\Delta}(\eta+\gamma)x[\check{\Delta}-(\eta+\gamma)x]}{\{2c\check{\Delta}-(\eta+\gamma)x\}^2\{c\check{\Delta}-(\eta+\gamma)x\}} \geq 0 \text{ by A.3;}$$

$$(ii) c_{1;\theta=1}^o - c_{1;\theta=1}^e = \frac{\check{\Delta}(\eta+\gamma)x}{2\check{\Delta}} \geq 0 \text{ by A.3;}$$

$$(iii) c_{2;\theta=1}^o - c_{2;\theta=1}^e = \frac{[2\check{\Delta}-(\eta+\gamma)x](\eta+\gamma)x}{2\check{\Delta}} \geq 0 \text{ by A.3.}$$

¹²Second-order conditions are satisfied as they require $(\alpha_{taxi} + d) \geq 0$ and $A \geq 0$.

The probability of the bus being on time well and truly is higher in the optimal situation than in equilibrium.

Appendix F Proof of Proposition 5

The idea of the proof is that the difference between optimal demand for the bus and equilibrium demand for the bus is a function of c , the bus punctuality cost and d the taxi operating cost. Throughout this proof we consider the extreme case where $\theta = 1$. Let us recall the expression of demand for the bus function:

$$D_{bus} = \int_{\delta_{T,O}^A}^{\Delta} f(\delta) d\delta,$$

where $\delta_{T,O}^A = [\kappa + (1 - P)(\eta + \gamma)x] / (\check{\alpha} + \tau)$. We can define

$$\begin{aligned} \bar{D} \equiv D_{bus}^o - D_{bus}^e &= 1 - \frac{\kappa^o + (1 - P^o)(\eta + \gamma)x}{\Delta(\check{\alpha} + \tau^o)} \\ &\quad - \left(1 - \frac{\kappa^e + (1 - P^e)(\eta + \gamma)x}{\Delta(\check{\alpha} + \tau^e)} \right), \end{aligned}$$

where $\kappa^o = 0$, $\tau^o = d$, $\kappa^e = \frac{1}{2}[\Delta(\check{\alpha} + \tau^e) - (1 - P^e)(\eta + \gamma)x]$ and $\tau^e = \check{\alpha} + 2d$. We therefore have

$$\bar{D} = \frac{2\Delta(\check{\alpha} + d) - (3 + P^e - 4P^o)(\eta + \gamma)x}{4\Delta(\check{\alpha} + d)}.$$

Since P^e and P^o are functions of c (equations (6b) and (9a)), we derive \bar{D} on c . For that, we need to know the order of c_1^e , c_2^e , c_1^o and c_2^o . We know that $c_1^e \leq c_2^e$ and $c_1^o \leq c_2^o$.

$$\begin{aligned} c_1^o - c_2^e &= (\eta + \gamma)x - \frac{2\kappa^e \check{\eta} x}{\check{\Delta}^e}, \\ \iff c_1^o - c_2^e &= (\eta + \gamma)x \left[1 - \frac{(1 - P)2\kappa^e}{\check{\Delta}^e} \right], \\ \iff c_1^o - c_2^e &= (\eta + \gamma)x \left[1 - (1 - P) \left(1 - \frac{\Gamma^e x}{\check{\Delta}^e} \right) \right] \geq 0. \end{aligned}$$

We therefore have $c_1^e \leq c_2^e \leq c_1^o \leq c_2^o$ and we distinguish between five sub-cases defined depending on the position of c relatively to c_1^e , c_1^o , c_2^e and c_2^o . Indeed the expression of the derivative is different according to the value of c .

(i) If $c \leq c_1^e$ then $P^e = P^o = 1$ and $\partial \bar{D} / \partial c = 0$.

(ii) If $c \in [c_1^e; c_2^e]$ then $P^o = 1$ and $\partial \bar{D} / \partial c \geq 0$.

(iii) If $c \in [c_2^e; c_1^o]$ then $P^o = 1$, $P^e = \frac{1}{2}$ and $\partial \bar{D} / \partial c = 0$.

(iv) If $c \in [c_1^o; c_2^o]$ then $P^e = \frac{1}{2}$ and $\partial \bar{D} / \partial c \leq 0$.

(v) If $c \geq c_2^o$ then $P^e = P^o = \frac{1}{2}$ and $\partial \bar{D} / \partial c = 0$.

Critical values of $\bar{D}(c)$ follow:

$$\bar{D}(c_1^e) = \frac{1}{2},$$

$$\bar{D}(c_2^e) = \frac{1}{2} + \frac{(\eta+\gamma)x}{8\bar{\Delta}(\bar{\alpha}+d)},$$

$$\bar{D}(c_2^o) = \frac{1}{2} - \frac{3(\eta+\gamma)x}{8\bar{\Delta}(\bar{\alpha}+d)},$$

where $\bar{\Delta} = \Delta(\alpha_{taxi} + d)$.

The variations of the difference between optimal demand for the bus and equilibrium demand for the bus are described in Table 3.

We know that $\bar{D}(c_1^e) \geq 0$ and $\bar{D}(c_1^o) \geq \bar{D}(c_1^e)$. According to Table 3, we can distinguish between two cases where the difference between optimal and equilibrium demand for the bus is positive. First, if the minimum value of the difference, $\bar{D}(c_2^e)$ is positive, the difference is positive. Second, if this minimum value of the difference is negative, then as $\bar{D}(c_2^e) \geq 0$ and $\bar{D}(c)$ strictly decreases between c_1^o and c_2^o , there exists a unique value of c denoted c^c

Table 3: Variation table of the difference between optimal demand for the bus and equilibrium demand for the bus depending on the cost of reliability.

c	0	c_1^e	c_2^e	c_1^o	c_2^o	∞	
$\frac{\partial \bar{D}}{\partial c}$	0		+	0	-	0	
$\bar{D}(c)$	\rightarrow	$\bar{D}(c_1^e)$	\nearrow	$\bar{D}(c_2^e)$	\searrow	$\bar{D}(c_2^o)$	\rightarrow

for which $D_{bus}^o \underset{c=c^c}{=} D_{bus}^e$. The difference is then positive if $c \leq c^c$.

One critical value of the taxi operating cost d_1^c may be defined such that

$$\begin{aligned} \bar{D}(c_2^o) \geq 0 &\iff 2\Delta(\tilde{\alpha} + d) - (3 + P^e - 4P^o)(\eta + \gamma)x \geq 0, \\ &\iff d \geq \frac{3(\eta + \gamma)x}{4\Delta} - \tilde{\alpha} \equiv d_1^c. \end{aligned}$$

We can now write

$$\bar{D} \begin{cases} \geq 0 & \text{if } d \geq d_1^c, \\ & \text{or if } d < d_1^c \text{ and } c \leq c^c, \\ < 0 & \text{if } d < d_1^c \text{ and } c > c^c. \end{cases}$$

Appendix G Proof of lemma 6

The idea of the proof is that the difference between optimal demand for the bus and equilibrium demand for the bus is a function of the cost of the bus reliability c and the operating cost of taxi d . We deal with the case where $\theta = 0$. Let us recall expressions of the demand function:

$$D_{bus} = (1 - P) \int_{\delta_{T,L}^B}^{\delta_{L,O}^B} f(\delta) d\delta + \int_{\delta_{L,O}^B}^{\Delta} f(\delta) d\delta,$$

where $\delta_{L,O}^B = [\kappa + (\frac{1-P}{P}\eta + \beta)x] / (\check{\alpha} + \tau)$ and $\delta_{T,L}^B = \kappa / (\check{\alpha} + \tau)$. We therefore have:

$$\bar{D} \equiv D_{bus}^o - D_{bus}^e = \frac{2\Delta(\check{\alpha} + d) + (4P_{\theta=0}^o - P_{\theta=0}^e)(\eta - \beta)x - 3\eta x}{4\Delta(\check{\alpha} + d)}.$$

Then

$$\bar{D} \geq 0 \iff d \geq \frac{3\eta x - (4P_{\theta=0}^o - P_{\theta=0}^e)(\eta - \beta)x}{2\Delta} - \check{\alpha}.$$

Considering $\max(4P_{\theta=0}^o - P_{\theta=0}^e) = \frac{7}{2}$ and $\min(4P_{\theta=0}^o - P_{\theta=0}^e) = 1$, we can define d_2^c and d_3^c such as if $d \leq d_2^c$ then $\bar{D} \leq 0$ and if $d \geq d_3^c$ then $\bar{D} \geq 0$. Consequently we have $d_2^c = -(\frac{1}{2}\eta - \frac{7}{2}\beta)x/2\Delta - \check{\alpha}$ and $d_3^c = (2\eta + \beta)x/2\Delta - \check{\alpha}$. We may write:

$$\bar{D} \begin{cases} \geq 0 & \text{if } d \geq d_3^c, \\ \leq 0 & \text{if } d \leq d_2^c. \end{cases}$$

Appendix H Discussion of Conjecture 2

With values specified in Table 1, we can draw the curve of the difference between optimal demand for the bus and equilibrium demand for the bus depending on the operating taxi cost d in Figure 12. When c is small, $P^o = P^e = 1$ and when c is large, $P^o = P^e = 1/2$. The lemma 6 is illustrated. \bar{D} functions are first negative then positive. Moreover they increase with d . The sign of \bar{D} between d_2^c and d_3^c depends on the values of P^o and P^e which depend on c (see Equations (6b) and (9b)). Therefore we conjecture that between d_2^c and d_3^c , \bar{D} is positive when $c \leq c_2^c$ and negative when $c > c_2^c$, where c_2^c is defined as the unique solution of $D_{\theta=0}^o = D_{\theta=0}^e$.

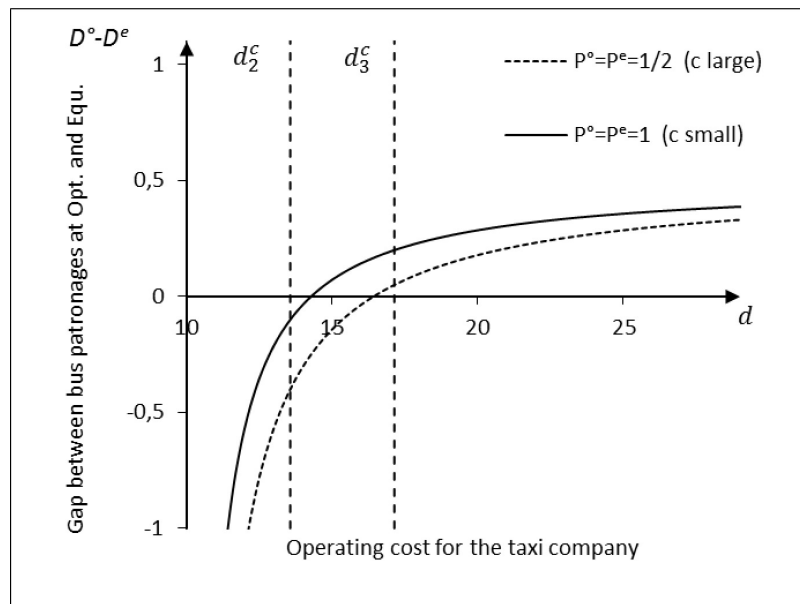


Figure 12: Difference between optimal demand for the bus and equilibrium demand for the bus depending on operating taxi cost d for *Group B*.

Chapter II

Economics of crowding in public transport

1 Introduction

Crowding is a growing problem in urban mass transit systems around the world.¹ A recent roundtable report by the International Transport Forum identifies crowding as a major source of inconvenience that increases the cost of travel (OECD, 2014). Several recent studies have documented the aggregate cost of crowding on transit networks. For example, Prud'homme et al. (2012) estimate that the 8% increase in densities² in the Paris subway between 2002 and 2007 imposed a welfare loss in 2007 of at least €75 million. Veitch et al. (2013) estimate the annual cost of crowding in Melbourne metropolitan trains in 2011 at €208 million. The costs of crowding are likely to grow as usage of public transit increases faster than transit investment.³

Transit crowding imposes disutility on riders in several ways. It increases waiting time (Oldfield and Bly, 1988) and in-vehicle travel time (Lin and Wilson, 1992), and reduces travel time reliability (Bates et al., 2001). Psychological studies find that crowding causes stress and feelings of exhaustion (Mohd Mahudin et al., 2012). A number of studies

¹See for example The Economist (2003), for London; Kelton (2012), for Adelaide; Wei (2011), for Beijing; and Chakraborty (2011), for Nagpur, India.

²Measured in passengers per square meter aboard trains.

³Allen and Levinson (2014) and King et al. (2014) document the rapid growth in usage of commuter rail services in North America.

document how disutility from in-vehicle time increases with the number of users (Li and Hensher, 2011; Wardman and Whelan, 2011; Kroes et al., 2014; Haywood and Koning, 2015; de Lapparent and Koning, 2015). Discomfort also occurs while entering and exiting vehicles, accessing stations on walkways and escalators, and so on. Extreme crowding is also a concern for emergency evacuation of facilities.

Crowding can affect peoples' transit choices in several ways including: departure time, transit line or route (Raveau et al., 2011) and destination. It can also influence their decisions whether to take transit, use another travel mode or not travel at all. Crowding is therefore an important factor in making investments in rolling stock and station capacity, and should be considered in cost-benefit analysis of transit projects (Parry and Small, 2009). Yet many policies that are designed to discourage driving and/or encourage use of public transport take into account road traffic congestion (Anderson, 2014) while ignoring the fact that transit capacity constraints and crowding limit the scope for shifting drivers off the road.

There is a large operations research literature on public transit system design (Vuchic, 2005). Beginning with Mohring (1972), an extensive economics literature has also developed on public transit capacity investments, service frequency, and optimal pricing and subsidy policy. There is also a rapidly growing literature on public transport crowding.⁴ These three branches of literature have made significant advances in understanding public transit systems and crowding. Nevertheless, most studies have employed static models that cannot account for travelers' time-of-use decisions and the large variations in ridership and crowding by time of day that are typically seen in transit systems. Trip-timing decisions are clearly important for transit. Users who travel during peak hours encounter crowded vehicles, and may be unable to board the first vehicle that arrives at the station or stop. Alternatively, users may choose to travel before or after the peak hour in a less crowded vehicle, but arrive earlier or later than they would like.⁵

The trade-off between crowding costs and scheduling costs is fundamental for understanding transit users' behavior and allocation of time to activities, particularly in the

⁴See Tirachini et al. (2013) for a review.

⁵Such behaviour is documented in Pepper et al. (2003) and Pownall et al. (2008).

short run (Peer et al., 2015). Automobile drivers face a similar trade-off between traffic congestion delays and scheduling costs that has been studied for many years. Vickrey (1969) was the first to do so by assuming that congestion takes the form of queuing behind a bottleneck, and that travellers incur a *schedule delay cost* if they complete their trip before or after their ideal arrival time. Building on Vickrey (1969), as well as the time allocation models of Becker (1965) and DeSerpa (1971), Small (1982) developed a utility-theoretic model of activity scheduling, and estimated individuals' preferences for when to commute to work when travel time varies with departure time. Henderson (1974, 1981) and Chu (1995) adopted the same demand-side specification as Vickrey (1969) and Small (1982), but instead of queuing assumed that travel delay manifests as flow congestion. As we will show, the public transit model we use has some of the same properties as the Henderson-Chu model.

A few studies of public transit systems have adopted models similar to the Vickrey and Henderson-Chu models. Kraus and Yoshida (2002) use the bottleneck model to analyze optimal pricing and capacity decisions for a rail service between a single origin and destination. They assume that trains run on a fixed timetable with a uniform headway, and travel time from origin to destination is independent of passenger load. Congestion takes the form of queuing delay. The number of people who board a train is limited by its capacity, and service discipline is First Come First Served. Users incur schedule delay costs from arriving early that are proportional to the amount of time they arrive early. Late arrivals are ruled out. In equilibrium, users who travel at a popular time have to wait in a queue at the origin station for several trains to pass before they can board. Those who take the earliest train avoid waiting, but incur the highest early-arrival costs.

Unlike Kraus and Yoshida (2002), Huang et al. (2005) assume that congestion is manifest as crowding rather than queuing. Instead of FCFS priority, travelers board trains in random order and everyone waiting for a train is able to get on. Huang et al. (2005) solve for the equilibrium distribution of users across trains, the fare and headway for the social optimum as well as two private ownership regimes. Several studies build on Huang et al. (2005). Huang et al. (2007) add mode choice by assuming that travellers can choose between taking a bus and driving. Buses and cars share the same right of way which is

subject to bottleneck queuing congestion. Tian et al. (2007) consider a multiple-origin-single-destination corridor with transit as the only travel mode. They establish several properties of the equilibrium times at which users board trains at each origin station. In an extension of their model they also make a distinction between the crowding or discomfort cost experienced by users who obtain a seat, and the costs incurred by users who have to stand. Tian et al. (2009b) solve the first-best optimum in the Tian et al. (2007) model, and derive train-dependent fares that supports the optimum. Xie and Fukuda (2014) do likewise using a different specification of scheduling preferences in which the costs of arriving early or late are quadratic rather than linear. Tian et al. (2009a) revert to a single origin-destination pair network, but distinguish between seated and standing passengers and apply the model to a light rail line in Beijing. Finally, de Palma et al. (2015) focus in more depth than earlier studies on the functional form of the crowding cost function for seated and standing passengers. They derive an optimal timetable and pricing scheme for several stylized settings.

Most of the studies reviewed in the previous paragraph take an engineering and/or operational research view of crowding, and do not explore all the economic aspects of the problem. In this paper we conduct a thorough economic analysis of trip-timing decisions and crowding on a transit line connecting a single origin and destination. Our work builds on the earlier studies in several directions. First, we examine and compare three fare regimes: a zero fare, an optimal uniform fare that is the same for all trains, and an optimal train-dependent fare. The optimal train-dependent fare internalizes crowding congestion and supports the socially optimal number and distribution of passengers on each train. We show how the properties of the equilibrium distribution depend on the curvature of the crowding cost function. Second, we compare passenger load patterns in the fare regimes and establish under what conditions loads are more even in the social optimum than with a uniform or zero fare. Third, we derive the welfare gain from implementing optimal train-dependent fares and examine how it depends on functional forms and parameter values.

Fourth, we derive the optimal number of trains and train capacity for each fare regime. Given linear crowding and schedule delay cost functions we are able to establish partial rankings across regimes. We show that for a given train capacity, more trains are operated

in the social optimum than in the optimal uniform-fare regime. (This is because usage is more spread out in the social optimum so that additional trains are more heavily used and valuable to add.) Similarly, if the number of trains is held constant, optimal train capacity is higher in the social optimum than the optimal uniform-fare regime. Finally, we calibrate the model to a portion of the Paris RER A mass transit system, and derive rough estimates of the potential welfare gains from introducing optimal train-dependent fares.

Throughout the analysis we compare results with those of the Vickrey and Henderson-Chu road traffic congestion models. Many of the results are intuitive, but a few are not. In particular, in the short run while capacity is fixed, the welfare gain from implementing train-dependent fares to internalize crowding cost externalities does not necessarily increase with the total number of users of the system. Indeed, if crowding costs grow at an increasing rate with the number of passengers aboard a train, the welfare gain actually decreases with the total number of users. This counterintuitive finding might help to explain why time-of-day dependent fare systems are still relatively rare.

The paper is organized as follows. Section 2 describes the general model and derives some properties of the no-fare equilibrium and the social optimum. Section 3 analyzes the case with linear crowding costs. Section 4 introduces elastic demand. Section 5 considers the long run in which the number of trains and train capacity are endogenous. Section 6 presents a numerical example based on the Paris RER A, and Section 7 concludes.

2 The general model with inelastic demand

In this section we describe and analyze a general model of public transit crowding which we will occasionally refer to as the “*PTC*” model. A transit line connects two stations without intermediate stops. The line operates on a timetable to which the operator adheres precisely. There are m trains, indexed in order of departure.⁶ Train k leaves the origin station at time t_k , $k = 1, \dots, m$. The timetable is left unspecified until Section 6 where it is assumed that the headway between successive trains is constant. Travel time aboard

⁶A notational glossary is provided in the Appendix.

a train is independent of both departure time and train occupancy, and without loss of generality it is normalized to zero. Each train therefore arrives at the destination station as soon as it leaves the origin station.

Each morning a fixed number, N , of identical individuals use the line to get to work. Users know the timetable and choose which train to take. Users are assumed to board a train in random order, and thus cannot increase their chance of securing a good seat by arriving at the origin station early. Users choose between trains based on the expected crowding disutility, $g(n)$, where n is the (known) number of users taking the same train. Function $g(n)$ is an average over possible states: securing a good seat, getting a bad seat, having to stand in the middle of the corridor, standing close to the door, etc.. Crowding disutility is assumed to be zero on an empty train (i.e., $g(0) = 0$), strictly increasing with n (i.e., $g'(\cdot) > 0$), and twice continuously differentiable. Several properties of the model derived later depend on the curvature of $g(n)$ which will be described by the elasticity of $g'(n)$ with respect to n : $\varepsilon(n) \equiv g''(n)n/g'(n)$. The elasticity is respectively positive, zero or negative as $g(n)$ is convex, linear or concave.

Because trains are costly to procure and operate, it is natural to assume that all m trains are used. Letting n_k denote the number of users on train k we thus assume that N is large enough that $n_k > 0$, $k = 1 \dots m$. Since $g(n) > 0$ for $n > 0$, this implies that $g(n_k) > 0$, $k = 1 \dots m$: users incur a crowding disutility on every train.

Since travel time is normalized to zero, an individual is either at home or at work. Time at home yields a time-varying utility $u_h(t)$, and time at work a time-varying utility of $u_w(t)$. Let (t_B, t_E) denote the time interval during which all travel takes place. During this interval, $u_h(t)$ is assumed to be weakly decreasing and $u_w(t)$ is weakly increasing. The functions intersect at time t^* which is the *desired arrival time* (i.e., $u_h(t^*) = u_w(t^*)$). A user taking train k gains a total utility:

$$U_k = \int_{t_B}^{t_k} u_h(t) dt + \int_{t_k}^{t_E} u_w(t) dt - g(n_k).$$

If there were no need for travel from home to work, the user would switch from home to work at t^* without suffering crowding disutility. As a consequence, his utility would be

maximal and equal to

$$U^{\max} = \int_{t_B}^{t^*} u_h(t) dt + \int_{t^*}^{t_E} u_w(t) dt.$$

We define the *user travel cost*, c_k , as the difference between this hypothetical utility and the actual utility of taking train k :

$$c_k = U^{\max} - U_k = g(n_k) + \begin{cases} \int_{t_k}^{t^*} (u_h(t) - u_w(t)) dt & \text{if } t_k < t^* \\ \int_{t^*}^{t_k} (u_w(t) - u_h(t)) dt & \text{if } t_k \geq t^* \end{cases}.$$

Note that maximizing U_k is equivalent to minimizing c_k . It is convenient to define the *schedule delay cost*, $\delta(t_k)$:

$$\delta(t_k) = \begin{cases} \int_{t_k}^{t^*} (u_h(t) - u_w(t)) dt & \text{if } t_k < t^* \\ \int_{t^*}^{t_k} (u_w(t) - u_h(t)) dt & \text{if } t_k \geq t^* \end{cases}.$$

The schedule delay cost is the disutility accumulated while an individual is not where his utility is greatest. When the individual arrives at work before t^* , disutility is incurred because utility from being at home before t^* is higher than utility at work. Similarly, utility is foregone when arriving at work after t^* because time is more valuable at work than at home. Function $\delta(t)$ is weakly decreasing for $t < t^*$ and weakly increasing for $t > t^*$.

In Section 3 it is assumed that $\delta(t)$ has a piecewise linear form:

$$\delta(t_k) = \begin{cases} \beta(t^* - t_k) & \text{if } t_k < t^* \\ \gamma(t_k - t^*) & \text{if } t_k \geq t^* \end{cases},$$

where β and γ are respectively marginal disutilities from arriving early and late. This specification, called “step preferences”, is used in most studies of road traffic congestion and public transit crowding.⁷

This piecewise linear form arises when the utilities flows from being at home and at

⁷Xie and Fukuda (2014) have recently explored an alternative specification called “step preferences” in which $u_h(t)$ is a linear decreasing function, and $u_w(t)$ is a linear increasing function. They estimate both specifications using Japanese data and do not find a clear advantage between them in terms of goodness-of-fit.

work satisfy⁸

$$u_h(t) - u_w(t) = \begin{cases} \beta & \text{if } t_k < t^* \\ -\gamma & \text{if } t_k \geq t^* \end{cases} .$$

In the general case, a user taking train k with n_k users incurs a combined schedule delay and crowding disutility of

$$c_k = \delta(t_k) + g(n_k), \quad k = 1, \dots, m. \quad (18)$$

To economize on writing, $\delta(t_k)$ is henceforth written δ_k unless time dependence is required for clarity. Trains that arrive close to t^* have small values of δ_k , and will sometimes be called *timely trains*. As shown in the next subsection, timely trains are more heavily used than trains with larger values of δ_k .

2.1 User equilibrium

In this subsection we derive and characterize user equilibrium when there is no fare. With N fixed, a fare would not affect either the division of users between trains or crowding costs. A time-varying fare will be considered for the social optimum in subsection 2.2, and a uniform fare (i.e., independent of k) will be introduced for the analysis of user equilibrium with elastic demand in Section 4.

Let superscript “ e ” denote the no-fare or user equilibrium (UE), and c^e the equilibrium trip cost. In UE, users distribute themselves between trains so that the user cost on every train is c^e . Hence

$$\delta_k + g(n_k^e) = c^e, \quad k = 1, \dots, m. \quad (19)$$

Given $g'(\cdot) > 0$, the inverse function $g^{-1}(\cdot)$ exists, with $g^{-1}(0) = 0$ and $(g^{-1})'(\cdot) > 0$. Eq. (19) can therefore be solved for the n_k^e as a function of c^e :

$$n_k^e = g^{-1}[c^e - \delta_k], \quad k = 1, \dots, m. \quad (20)$$

⁸This property is discussed by Tseng and Verhoef (2008).

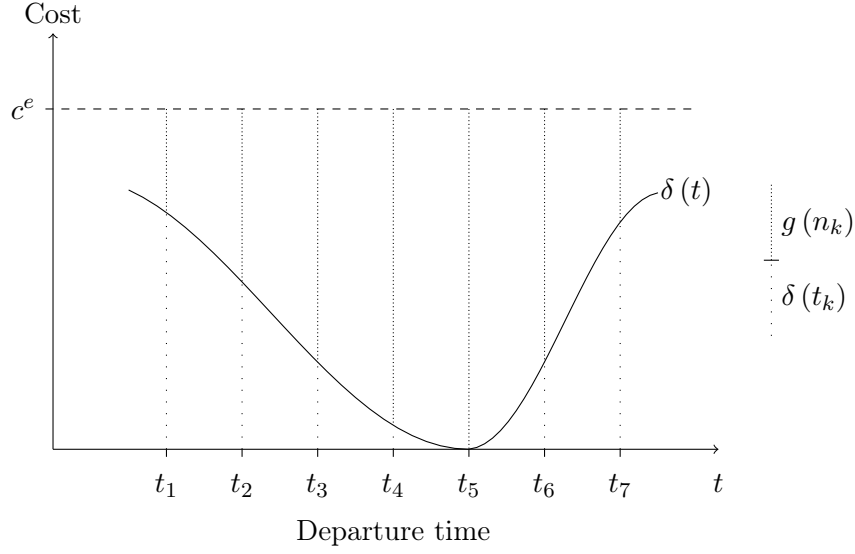


Figure 13: Schedule delay $\delta(t_k)$, crowding cost $g(n_k)$ and equilibrium cost c_e for seven trains, $t_5 = t^*$

Since every user has to take some train,

$$\sum_{k=1}^m n_k^e = N, \quad (21)$$

or

$$\sum_{k=1}^m g^{-1}[c^e - \delta_k] - N = 0. \quad (22)$$

Eq. (22) implicitly determines c^e . Figure 13 depicts a UE for seven trains ($m = 7$). Train $k = 5$ arrives on time and carries the most users. Less timely trains carry fewer passengers because users incur schedule delay costs.

Comparative statics properties of UE with respect to N are easily derived. Equilibrium cost increases with the total number of passengers:

$$\frac{\partial c^e}{\partial N} = \frac{1}{\sum_{k=1}^m \frac{1}{g'(u_k)}} > 0, \quad (23)$$

where $u_k \equiv g^{-1}[c^e(N) - \delta_k]$. The second derivative is:

$$\frac{\partial^2 c^e}{\partial N^2} \propto \sum_{k=1}^m \frac{g''(u_k)}{[g'(u_k)]^3}. \quad (24)$$

The sign of expression (24) depends on whether $g(\cdot)$ is convex or concave. Properties (23) and (24) of UE are summarized in the following proposition.

Proposition 7. *In equilibrium, user cost is an increasing function of N . It is convex, linear or concave if $g(\cdot)$ is respectively convex, linear or concave.*

Similar to the models of Vickrey (1969), Henderson (1981), Chu (1995) and Kraus and Yoshida (2002), user cost in the *PTC* model is an increasing function of total patronage. All the models have this property because they feature congestion in some form. However, the curvature of $c^e(N)$ differs across the models. In Vickrey's model the curvature of $c^e(N)$ matches that of the schedule delay cost function. With step preferences, the schedule delay cost function is linear and $c^e(N)$ is also linear. If the schedule delay cost function is convex (resp. concave), then $c^e(N)$ is convex (resp. concave). It is straightforward to show that the Kraus and Yoshida (2002) model also has these properties. By contrast, in the Henderson/Chu model $c^e(N)$ is concave if the schedule delay cost function is either linear or concave. The Henderson/Chu model behaves differently because, with flow congestion, there is a trade-off between travel speed and flow. As demand increases, the arrival rate of vehicles at the destination increases and the duration of the travel period increases less than proportionally with N . In effect, the road (or other facility) has more flexibility to accommodate additional traffic than the bottleneck in Vickrey's model.

In contrast to the other models, in the *PTC* model the curvature of $c^e(N)$ depends on the crowding cost function rather than the schedule delay cost function. This is because the train timetable is fixed in the short run, and users cannot travel earlier or later in response to growing demand. Furthermore, since each train's arrival time is fixed, the schedule delay cost incurred when taking a given train does not depend on N . The only way the service can accommodate additional demand is for each train to carry more passengers. Equilibrium user cost therefore increases at an increasing (resp. decreasing) rate with N if the marginal cost of crowding aboard a train increases (resp. decreases) with ridership.

User equilibrium in the *PTC* model is clearly inefficient because users impose an external crowding cost on each other. The marginal social cost of a trip, MSC , is determined

by differentiating the equilibrium total cost function, $TC^e = N \times c^e$, with respect to N :

$$MSC^e \equiv \frac{\partial TC^e}{\partial N} = c^e + N \frac{\partial c^e}{\partial N}.$$

The average marginal external cost of a trip is therefore

$$MEC^e \equiv MSC^e - c^e = N \frac{\partial c^e}{\partial N}. \quad (25)$$

With elastic demand (Section 4) it is optimal to charge a uniform fare equal to the average external cost to avoid over-use of public transport:

$$\tau^u = N \frac{\partial c^e}{\partial N},$$

where superscript “ u ” denotes the optimal uniform fare. Total revenue from this fare is

$$R^u = \tau^u N = N^2 \frac{\partial c^e}{\partial N}. \quad (26)$$

The optimal uniform fare does not support the social optimum because the marginal external cost of crowding varies with train occupancy and it is larger on more heavily used trains. As explained in the next subsection, the social optimum can be achieved by levying train-specific fares.

2.2 Social Optimum

The social optimum (SO) differs from the UE because users are distributed between trains to equalize the marginal social costs of trips rather than their private costs. The user cost of a trip on train k is given by Eq. (18). The marginal social cost of using train k is

$$MSC_k = \frac{\partial (c_k n_k)}{\partial n_k} = \delta(t_k) + v(n_k), \quad k = 1, \dots, m, \quad (27)$$

where $v(n_k) \equiv g(n_k) + n_k g'(n_k)$ is the marginal social crowding cost on train k . The marginal external cost equals the difference between the right-hand sides of Eq. (27) and

Eq. (18):

$$MEC_k = MSC_k - c_k = n_k g'(n_k), \quad k = 1, \dots, m. \quad (28)$$

Let superscript “o” denote the SO. Total costs in the SO are $TC^o = \sum_{k=1}^m c_k n_k$, and the marginal social cost of a trip is $MSC^o = \frac{\partial TC^o}{\partial N}$. At the optimum, users are distributed across trains so that $MSC_k = MSC^o$ for every train:

$$\delta_k + v(n_k^o) = MSC^o, \quad k = 1, \dots, m. \quad (29)$$

Since $g'(\cdot) > 0$ for $n > 0$, the marginal social crowding cost is always positive. In practice, it may not increase monotonically at all levels of ridership.⁹ To facilitate analysis, however, we assume that $v'(\cdot) > 0$. This is equivalent to assuming that $g''(n) > -2g'(n)/n$, or $\varepsilon(n) > -2$.

Assumption 4. *The elasticity of $g'(n)$ with respect to n exceeds -2 : $\varepsilon(n) > -2$.*

Assumption 4 is satisfied for all convex $g(\cdot)$ functions and for all power function $g(n) \propto n^r$, $r < 0$. It implies that the marginal social cost of usage increases with the number of users on a train. Given Assumption 4, the inverse function $v^{-1}(\cdot)$ exists, and Eq. (29) yields

$$n_k^o = v^{-1}[MSC^o - \delta_k]. \quad (30)$$

Since all users must take some train in the SO, $\sum_{k=1}^m n_k^o = N$ and Eq. (30) implies:

$$\sum_{k=1}^m v^{-1}[MSC^o - \delta_k] - N = 0. \quad (31)$$

Because functions $g(\cdot)$ and $v(\cdot)$ are both positive and increasing, (31) has the same qualitative properties as (22) describing the UE, and the following counterpart to Prop. 7 can be stated:

Proposition 8. *In the social optimum, the marginal social cost of a trip is an increasing function of N . It is convex, linear or concave if $v(\cdot)$ is respectively convex, linear or*

⁹For example, $v(\cdot)$ may drop when all seats are occupied and additional riders have to stand; see de Palma et al. (2015).

concave.

Comparing Prop. 8 with Prop. 7 it is clear that $v(\cdot)$ plays the same role in shaping the SO as $g(\cdot)$ does for the UE.¹⁰ Prop. 8 contrasts again with the corresponding properties of the SO in the Vickrey (1969), Henderson (1981), Chu (1995) and Kraus and Yoshida (2002) models. For example, with linear schedule delay costs the marginal social cost of a trip in the Vickrey model is a linear function of N . In the *PTC* model it instead depends on the crowding cost function.

We now consider the distribution of ridership over trains. Intuition might suggest that passenger loads are spread more evenly in the SO than the UE because smoother loads should reduce the total costs of crowding as discussed in de Palma et al. (2015). In fact, this is not invariably true but rather depends on how the marginal external crowding cost varies with usage. Now

$$\frac{dMEC}{dn} = \frac{d(ng'(n))}{dn} = g'(n) + ng''(n) = g'(n)(1 + \varepsilon(n)).$$

The marginal external crowding cost increases with usage if $\varepsilon(n) > -1$, and decreases with usage if $\varepsilon(n) < -1$. The load patterns in the SO and UE are compared in the following proposition.

Proposition 9. *If $\varepsilon(n) > -1$ ($\varepsilon(n) < -1$, respectively) the social optimum distribution of users across trains is a mean-preserving spread (respectively contraction) of the user equilibrium distribution of users across trains.*

Proof: see Appendix A. The mean-preserving spread concept has been defined by Rothschild and Stiglitz (1970).¹¹ The SO load pattern is a mean-preserving spread of the UE load pattern if the SO load pattern has more weight in the tails than the UE load pattern. It has $\varepsilon(n) > -1$ applies if the marginal external crowding cost increases monotonically with passenger load.¹² If so, the marginal social costs of trips on two trains

¹⁰Tian et al. (2009b) also remark on this property.

¹¹The case of interest for our paper is p. 229 in subsection II.2 *Mean Preserving Spreads: Discrete Distributions.*

¹²Similar to Assumption 4, which is weaker, $\varepsilon(n) > -1$ is satisfied for all convex crowding cost functions, and crowding cost functions that belong to the class of power functions: $g(n) \propto n^r$, $r > 0$.

with unequal loads differ by more than their user costs. Consequently, the SO balance between crowding costs and schedule delay costs calls for a smaller range of train loads than in the UE. Conversely, if $\varepsilon(n) < -1$, which is possible only if $g(\cdot)$ is sufficiently concave,¹³ then passenger loads are more peaked in the SO than the UE.

In summary, the difference in train loads between the SO and UE is determined by the curvature of the crowding cost function. According to most empirical studies $g(\cdot)$ is linear or convex (Wardman and Whelan, 2011; Haywood and Koning, 2015; de Palma et al., 2015). As a consequence, empirical evidences suggest that $\varepsilon(n) \geq 0$, and that ridership in the UE is too concentrated on timely trains and should be spread out.

Regardless of whether the SO is more or less peaked than the UE, the SO usage pattern can be decentralized by charging a fare on train k equal to the marginal external cost of usage.¹⁴ We will call the fare pattern the *SO-fare*. Given Eq. (28) the SO-fare is:

$$\tau_k^o = n_k^o g'(n_k^o), \quad k = 1, \dots, m. \quad (32)$$

With this fare structure in place, users of train k incur a private cost equal to the social cost of a trip:

$$p_k^o = c_k^o + \tau_k^o = MSC^o, \quad k = 1, \dots, m. \quad (33)$$

The SO is more efficient than the UE because users are better distributed between trains. However, inclusive of the SO-fare users incur a higher private cost in the SO. To see this, note that at least one train is more crowded in the SO than the UE. Compared to the UE, in the SO a rider of that train incurs the same schedule delay cost but a higher crowding cost and a positive fare. Since all users incur the same private cost in the UE, and all users incur the same private cost in the SO, private costs are higher in the SO.¹⁵

Unless fare revenues are used to improve service in some way, charging fares to price crowding costs in the *PTC* model leaves users worse off. This is also true of pricing road traffic congestion in the Henderson/Chu model although the physical effects of tolling

¹³For example, Case 2 of Prop. 9 holds for the function $g(n) = c_0 + c_1 \ln(n) - kn$ for $c_0 > k$ and $n \in [1, c_1/k]$.

¹⁴The fare is set according to Pigouvian principles. Revenue generation or other goals are ruled out.

¹⁵The difference in private cost is, however, smaller than the average fare paid because the social (i.e., resource) costs of travel are lower in the SO.

differ. In their model, tolling causes the departure period to spread out and the first and last users incur higher schedule delay costs than in the UE. Because the first and last users incur no congestion delay in either the UE or the SO, their costs are higher in the SO. Since all users incur the same private costs in the UE and SO, equilibrium private user costs are increased by tolling. By contrast, in the Vickrey and Kraus-Yoshida models congestion pricing leaves private costs unchanged because the travel period is not affected.

The Henderson/Chu and Vickrey/Kraus-Yoshida models therefore differ in the incidence of tolling costs. However, in all four models congestion toll revenues increase with N . This is because the average congestion externality increases with N , and hence so does the average toll. To determine how fare revenues in the *PTC* model vary with N , let R^o denote total revenue from the SO-fare. Revenues from SO-fare equal $R^o = \sum_{k=1}^m n_k^o \tau_k^o$ with n_k^o given in Eq. (30) and τ_k^o in Eq. (32), and revenues from optimal uniform fare are given by Eq. (26).

Proposition 10. *Let $i = u, o$ index the pricing regime. Then,*

$$\frac{\partial R^i}{\partial N} = N \frac{\partial MSC^i}{\partial N}.$$

Proof: see Appendix B. Prop 10 reveals that fare revenues increase if the marginal social cost of a trip increases with total usage. This will be the case unless the crowding cost function is sufficiently concave.

Next we examine how the welfare gain from implementing the SO-fare varies with usage. Let $G^{eo} \equiv TC^e - TC^o$ denote the welfare gain in shifting from the UE to the SO. Intuition suggests that G^{eo} increases with N : first because crowding becomes more onerous for users on average, and second because more users suffer the increased cost. However, we already know that the rate at which the cost of crowding increases with load depends on the curvature of the crowding cost function. It turns out that properties of the crowding cost function also govern how G^{eo} depends on N .

Consider the following assumptions:

Assumption 5. *The marginal external cost of crowding increases with load: $\varepsilon(n) > -1$.*

Assumption 6a. *The marginal social cost of crowding is a strictly convex function of load (i.e., $v''(n) > 0$), and $\varepsilon(n)$ is a nonincreasing function of load (i.e., $\frac{d\varepsilon(n)}{dn} \leq 0$).*

Assumption 6b. *The marginal social cost of crowding is a strictly concave function of load (i.e., $v''(n) < 0$), and $\varepsilon(n)$ is a nondecreasing function of load (i.e., $\frac{d\varepsilon(n)}{dn} \geq 0$).*

Assumption 6a holds if $g(n)$ is convex, and Assumption 6b holds if $g(n)$ is concave. The effect of total ridership on the welfare gain from the SO-fare is described in the following proposition.

Proposition 11. *Let Assumption 5 hold. The welfare gain from the SO-fare, G^{eo} , decreases with N , increases with N or is independent of N if Assumption 6a holds, Assumption 6b holds or if $g(\cdot)$ is linear, respectively.*

Proof: see Appendix C. Proposition 11 identifies conditions under which G^{eo} increases, decreases, or is independent of total ridership. Since the conditions are not collectively exhaustive, Prop. 11 does not establish the direction of change for all cases. Nevertheless, the conditions span a broad class of functions.

As noted earlier, most empirical studies find that $g(\cdot)$ is linear or convex. According to Prop. 11, G^{eo} is then either constant or a decreasing function of N . This is a surprise since it goes against the intuition described above. To understand why, note that the welfare gain derives from reallocating users between trains. If $g(\cdot)$ is convex, Case 1 of Prop. 9 applies and users are reallocated more evenly. Since the difference in crowding costs between two successive trains equals the difference in schedule delay costs as per Eq. (19), the marginal benefit from starting to reallocate users is independent of N . However, as N increases the marginal crowding cost on each train becomes higher and the UE and SO train loads become more similar. Consequently, the *amount* of reallocation decreases, and the total welfare gain therefore falls as well. The argument acts in reverse if $g(\cdot)$ is concave since the optimal amount of reallocation then increases with N .

Another way to view Prop. 11 is in terms of the marginal social cost of usage, which is MSC^e in the UE and MSC^o in the SO. If $MSC^o < MSC^e$, an additional user causes total costs to rise by less in the SO than the UE, and G^{eo} rises. Conversely, if $MSC^o > MSC^e$, total costs rise more in the SO and G^{eo} falls. Thus, if $g(\cdot)$ is convex an additional user is,

paradoxically, more costly to accommodate in the SO than in the UE even though users are distributed optimally between trains in the SO. If $g(\cdot)$ is linear, $MSC^o = MSC^e$ and the difference in total costs between UE and SO is independent of N . In effect, the benefits of internalizing the crowding cost externality are exhausted once total usage is large enough for all trains to be used. We illustrate this case diagrammatically in the next section.

For most of the balance of the paper we limit attention to a particular instance of the model in which the crowding cost function is linear. In Sections 5 and 6 we assume that the schedule delay cost function is linear as well. We choose linear functions for three reasons. First, it enhances analytical tractability and the model can be extended to allow elastic demand. The optimal number of trains and train capacity can be characterized as well. Second, linearity facilitates comparisons with the bottleneck model. Third, empirical studies often find that schedule delay and crowding costs are close to linear (see Wardman et al., 2012, for scheduling cost, and Wardman, 2004; Wardman and Whelan, 2011; Haywood and Koning, 2015, for crowding cost).

3 Linear crowding costs

We now assume that the crowding cost function is linear in usage: $g(n) = \lambda \frac{n}{s}$, where $s > 0$ is a measure of train capacity, and $\lambda > 0$. The marginal social crowding cost function is $v(n) = 2\lambda \frac{n}{s}$; i.e. twice the private crowding cost. From Eq. (18) the cost of taking train k is therefore

$$c_k = \delta_k + \lambda \frac{n_k}{s}, \quad k = 1, \dots, m. \quad (34)$$

3.1 User equilibrium with a uniform fare

In the uniform-fare regime the same fare is charged for all trains. The user equilibrium analyzed in subsection 2.1 is a special case where the fare is zero. We begin this subsection by setting the fare to zero, and then derive the optimal uniform fare for use in Section 4.

Using $g(n) = \lambda \frac{n}{s}$ in Eqs. (20) and (22) gives

$$c^e = \bar{\delta} + \frac{\lambda N}{ms}, \quad (35)$$

and

$$n_k^e = \frac{N}{m} + \frac{s}{\lambda} [\bar{\delta} - \delta_k], \quad k = 1, \dots, m, \quad (36)$$

where $\bar{\delta} \equiv \frac{1}{m} \sum_{k=1}^m \delta_k$ is the unweighted average scheduling cost for trains. As in the general model, timely trains carry more users than other trains. The difference in passenger loads between two successive trains is proportional to parameter s , and inversely proportional to λ . Because the first (or last) train carries the fewest passengers, the solution satisfies all the non-negativity constraints $n_k^e \geq 0$ if $n_1^e \geq 0$ and $n_m^e \geq 0$:

$$N > \frac{ms}{\lambda} [\max[\delta_1, \delta_m] - \bar{\delta}]. \quad (37)$$

Since service is costly to provide, condition (37) is satisfied when m and s are chosen optimally as in Section 5.

According to Eq. (35), for given values of m and s equilibrium trip cost is a linear increasing function of ridership, N . This property of the solution is formalized together with Eq. (36) as:

Proposition 12. *In the uniform-fare equilibrium, all m trains are used if Condition (37) is satisfied. Train k carries a load $n_k^e = \frac{N}{m} + \frac{s}{\lambda} [\bar{\delta} - \delta_k]$. User cost is $c^e = \bar{\delta} + \frac{\lambda N}{ms}$, which is a linear increasing function of N .*

As in the bottleneck model, the equilibrium cost is a linear function of N . This is a consequence of the assumptions here that the train timetable is fixed and crowding costs are linear.

We now determine aggregate travel costs. Let SDC denote total schedule delay costs, TCC total crowding costs, and TC total travel costs net of the fare.

Proposition 13. *In the uniform-fare equilibrium, total schedule delay costs, SDC , total crowding costs, TCC , and total travel costs net of the fare, TC are given by*

$$\begin{aligned} SDC^e &= \bar{\delta}N - 4RV^o, \\ TTC^e &= \frac{\lambda N^2}{ms} + 4RV^o, \\ TC^e &= \bar{\delta}N + \frac{\lambda N^2}{ms}, \end{aligned}$$

where $RV^o = \frac{s}{4\lambda} \left(\sum_{k=1}^m [\delta_k]^2 - \frac{1}{m} [\sum_{k=1}^m \delta_k]^2 \right)$.¹⁶

Note that $RV^o > 0$ by the Cauchy-Schwarz inequality. Crowding costs are analogous to travel time costs (TTC) in traffic congestion models. In the bottleneck model equilibrium, $SDC^e = TTC^e$ for all values of N . In the Henderson (1974, 1981) and Chu (1995) models, $SDC^e < TTC^e$. The behavior of SDC^e and TCC^e is more complicated in the PTC model. Total schedule delay costs are lower than if users were equally distributed across trains (in which case $SDC^e = \bar{\delta}N$), and total crowding costs are higher by the same amount. This is because users crowd onto timely trains that arrive closer to t^* . For small values of N only one train is used. Schedule delay costs are zero (if $t_1 = t^*$) while crowding costs are proportional to N^2 . For a given value of m , $m > 1$, SDC^e is a linear increasing function of N with a negative intercept, while TCC^e increases with N^2 and has a positive intercept.

We now derive the optimal uniform fare. The equilibrium private cost of a trip, p^e , equals the user cost plus the fare, τ :

$$p^e = \bar{\delta} + \frac{\lambda N}{ms} + \tau. \quad (38)$$

From Prop. 13, the marginal social cost of a trip is:

$$MSC^e = \frac{\partial TTC^e}{\partial N} = \bar{\delta} + \frac{2\lambda N}{ms}. \quad (39)$$

The first term in (39) is the same as in (35). The second term is proportional to N , and it is twice the corresponding term in (35) because the marginal social cost of crowding, $v(n)$, is twice the average cost. From Eq. (25), the average external cost is

$$MEC^e = \frac{\lambda N}{ms}.$$

If the demand is elastic (Section 4), the optimal uniform fare is:

$$\tau^u = \frac{\lambda N}{ms}. \quad (40)$$

¹⁶ RV^o is *variable revenue* from the SO-fare, defined in the next subsection. For ease of reference, we introduce the notation here.

This fare yields a total revenue of

$$R^u = \tau^u N = \frac{\lambda N^2}{ms}. \quad (41)$$

The external cost varies from train to train because it depends on train occupancy, which is higher on more timely trains. The social optimum is examined in the following subsection.

3.2 Social Optimum

The social optimum is readily derived using results for the general model in Section 2.2. Given $v(n) = 2\lambda \frac{n}{s}$, $v^{-1}(x) = \frac{s}{2\lambda}x$. From Eq. (31),

$$MSC^o = \bar{\delta} + \frac{2\lambda N}{ms}.$$

From Eq. (30),

$$n_k^o = \frac{N}{m} + \frac{s}{2\lambda} [\bar{\delta} - \delta_k], \quad k = 1, \dots, m. \quad (42)$$

Assumption 5 holds for the linear crowding cost function. Hence, by Prop. 8 train loads are more evenly distributed in the social optimum than the uniform-fare equilibrium. Equations (42) and (36) reveal that the difference in loads between successive trains is only half as large. The non-negativity constraint on usage of all trains is satisfied if

$$N > \frac{ms}{2\lambda} [\max[\delta_1, \delta_m] - \bar{\delta}]. \quad (43)$$

Condition (43) is satisfied if condition (37) is satisfied for the uniform-fare equilibrium.

The optimal usage pattern can be decentralized by charging a fare for train k equal to the marginal external cost of usage as given by Eq. (32). Using Eq. (42):

$$\tau_k^o = \lambda \frac{n_k^o}{s} = \frac{\lambda N}{ms} + \frac{\bar{\delta} - \delta_k}{2}, \quad k = 1, \dots, m. \quad (44)$$

Compared to the uniform fare in Eq. (40), the fare is lower on the earliest and latest trains

with $\delta_k > \bar{\delta}$, and higher on timely trains with $\delta_k < \bar{\delta}$. At the optimum, users of all trains incur a private cost equal to the social cost of a trip as per Eq. (33):

$$p_k^o = c_k^o + \tau_k^o = MSC^o = \bar{\delta} + \frac{2\lambda N}{ms}, k = 1, \dots, m.$$

The characteristics of the social optimum are summarized in the following proposition:

Proposition 14. *In the social optimum, all m trains are used if Condition (43) is satisfied. Train k carries a load of $n_k^o = \frac{N}{m} + \frac{s}{2\lambda} [\bar{\delta} - \delta_k]$. The optimum can be decentralized by charging a fare for train k of $\tau_k^o = \lambda \frac{n_k^o}{s}$. Private and marginal social costs of trips are the same for all trains and equal to $\bar{\delta} + \frac{2\lambda N}{ms}$ which is a linear increasing function of N .*

Given Eqs. (30) and (32), total revenue from the SO-fare is

$$R^o = \sum_{k=1}^m \tau_k^o n_k^o = \frac{\lambda N^2}{ms} + \frac{s}{4\lambda} \left(\sum_{k=1}^m \delta_k^2 - m\bar{\delta}^2 \right). \quad (45)$$

The first term in (45) matches revenue from the optimal uniform fare, R^u , in (41). The second term is extra revenue (when $m > 1$) due to variation of the fare. As noted in subsection 3.1 this is called *variable revenue*

$$RV^o \equiv \frac{s}{4\lambda} \left(\sum_{k=1}^m \delta_k^2 - m\bar{\delta}^2 \right). \quad (46)$$

A notable feature of (46) is that variable revenue is independent of N . This property is discussed below in connection with the welfare gains from imposing the SO-fare.

Proposition 15. *In the decentralized social optimum, total schedule delay costs, SDC , total crowding costs, TCC , and total travel costs net of the fare, TC , are given by*

$$\begin{aligned} SDC^o &= \bar{\delta}N - 2RV^o = SDC^e + 2RV^o, \\ TTC^o &= \frac{\lambda N^2}{ms} + RV^o = TTC^e - 3RV^o, \\ TC^o &= \bar{\delta}N + \frac{\lambda N^2}{ms} - RV^o = TC^e - RV^o. \end{aligned}$$

If there is only one train, there is no scope to spread usage between trains. Variable

revenue is zero, and aggregate costs are the same in the social optimum and the optimal uniform-fare equilibrium. With $m > 1$, total schedule delay costs are higher in the social optimum than the optimal uniform-fare equilibrium, but crowding costs are smaller by 1.5 as much and total costs are lower by an amount equal to variable fare revenue.

With linear crowding costs, the welfare gain from imposing the SO-fare is therefore equal to variable revenue

$$G^{eo} = RV^o. \quad (47)$$

Prop. 15 can be compared with the corresponding formulas in the bottleneck model, denoted with a subscript Bn :

$$\begin{aligned} SDC_{Bn}^o &= SDC_{Bn}^e, \\ TTC_{Bn}^o &= 0, \\ RV_{Bn}^o &= TTC_{Bn}^e, \\ TC_{Bn}^o &= TC_{Bn}^e - RV_{Bn}^o. \end{aligned}$$

Tolling in the bottleneck model eliminates queuing (the counterpart to crowding in the PTC model) without increasing total schedule delay costs. Variable revenue matches total queuing costs in the UE, and total costs are reduced by variable revenue. Thus, in both models variable revenue measures the welfare gain from tolling, but tolling is more effective in the bottleneck model because it eliminates the external cost of congestion without causing schedule delay costs to increase.

The numerical example in Section 6 features linear schedule delay costs and a constant headway between trains of h . Given Eq. (47) it is straightforward to show that for large values of m ,

$$G^{eo} \simeq \frac{s}{48\lambda} \left(\frac{\beta\gamma}{\beta + \gamma} \right)^2 h^2 m (m^2 - 1). \quad (49)$$

Eq. (49) offers a transparent view of how the welfare gain from the SO-fare varies with parameters. First, as noted in discussing Prop. 11 above, G^{eo} is independent of total usage, N . To see why, consider a simple case with two trains. As per Eq. (34), the cost of using train k is $c_k = \delta_k + gn_k$ where $g = \frac{\lambda}{s}$ measures the rate at which crowding costs

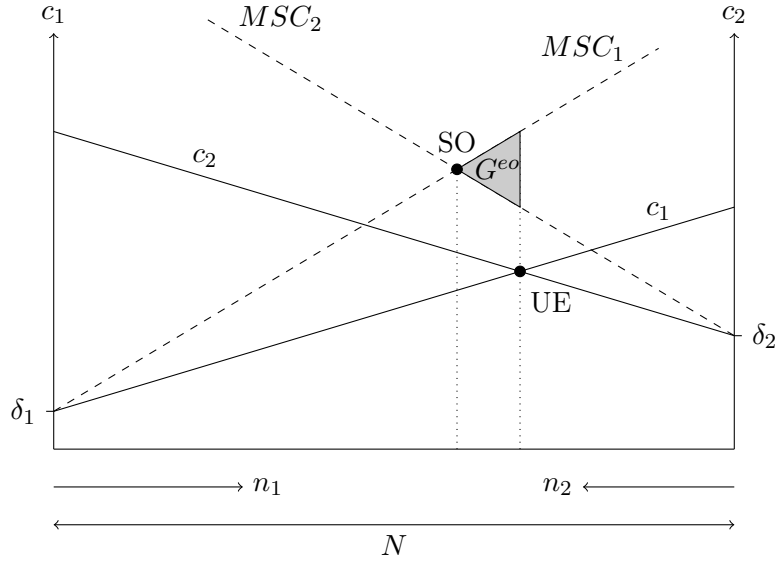


Figure 14: User equilibrium (UE), social optimum (SO) and welfare gain (G^{eo}) with two trains

increase with train load. Figure 14 depicts the UE and SO using a diagram with two vertical axes separated by N . Usage of train 1 is measured to the right from the left-hand axis, and usage of train 2 to the left from the right-hand axis. By assumption, $\delta_2 > \delta_1$ so that train 1 is overused in the UE. The welfare gain in shifting users from train 1 to train 2 is shown by the triangular shaded area. The height of the triangle is $\delta_2 - \delta_1$, and the width of the triangle is $\frac{\delta_2 - \delta_1}{2g}$. The area of the triangle is therefore $\frac{(\delta_2 - \delta_1)^2}{4g}$. It does not depend on N because neither dimension of the triangle depends on N . The height of the triangle equals the difference in marginal external costs of using the two trains in the UE. This is determined by the difference in their attractiveness, $\delta_2 - \delta_1$, not N . The width of the triangle is the optimal number of users to redistribute between trains which is proportional to $\delta_2 - \delta_1$, and inversely proportional to g . This, too, is independent of N .

Eq. (49) also reveals that G^{eo} varies with the square of β and γ together. This follows immediately from the quadratic dependence on the schedule delay costs, δ_1 and δ_2 . G^{eo} varies with the square of the headway, h , for the same reason. G^{eo} varies inversely with the ratio g because the scope to alleviate crowding by redistributing riders between trains decreases if trains become crowded more quickly.

Finally, G^{eo} varies approximately with the cube of the number of trains. This highly

nonlinear dependence is due to two multiplicative factors. First, the average schedule delay cost of trains varies proportionally with m . The average difference in schedule delay costs is therefore proportional to m , and the welfare gain from redistributing passengers between two trains varies with m^2 . Second, the number of trains between which passenger loads can gainfully be redistributed is approximately proportional to m . Hence, the overall welfare gain varies approximately with m^3 .

In the introduction to the paper we noted that, similar to other congestible facilities, the distribution of passengers between trains is governed by the trade-off users make between scheduling costs and crowding costs. It is therefore perhaps surprising that the parameters measuring the strength of these two costs have such divergent effects on the gains from congestion pricing. According to Eq. (49), doubling the unit costs of schedule delay, β and γ , increases the gains four-fold. By contrast, doubling the crowding cost parameter, λ , reduces the gains by half. In assessing the potential benefits from implementing congestion pricing it is therefore important to predict how these two costs will evolve over time.

4 Elastic demand

So far it has been assumed that transit ridership, N , is exogenous. Yet in practice, travelers can often use other transport modes. They may also choose to forego travel if it is too costly. To admit these possibilities we now assume that demand for public transport trips is a smooth and decreasing function of the private cost:

$$N = N(p), \quad \frac{\partial N}{\partial p} < 0. \quad (50)$$

Consumers' surplus from trips is

$$CS(p) = \int_p^\infty N(u) du,$$

and social surplus (gross of capacity costs) is the sum of consumers' surplus and fare revenue:

$$SS(p, \tau) = CS(p) + R.$$

In this section we continue to treat m and s as fixed. We first consider uniform fares (which include no fare and the optimal uniform fare as special cases), and then the SO-fare. The main goal of the section is to rank equilibrium prices and numbers of trips in the pricing regimes.

4.1 Uniform-fare regimes

With a uniform fare, the private cost of a trip is given by (38) which is reproduced here for ease of reference:

$$p^e = \bar{\delta} + \frac{\lambda N}{ms} + \tau. \quad (51)$$

Eq. (51) serves as a supply function for trips. Solving (51) and the demand function (50) yields the equilibrium private cost and number of trips, \widehat{p}^e and \widehat{N}^e , where a “ $\widehat{}$ ” denotes an equilibrium value with elastic demand. If the fare is zero, the equilibrium price is

$$\widehat{p}^n = \bar{\delta} + \frac{\lambda \widehat{N}^n}{ms}, \quad (52)$$

where superscript n denotes the no-fare regime. Social surplus equals consumers’ surplus:

$$\widehat{SS}^n = \widehat{CS}^n = \int_{\widehat{p}^n}^{\infty} N(u) du.$$

The optimal uniform fare is given by Eq. (40):

$$\widehat{\tau}^u = \frac{\lambda \widehat{N}^u}{ms},$$

and fare revenue is given by Eq. (41):

$$\widehat{R}^u = \tau^u \widehat{N}^u = \frac{\lambda (\widehat{N}^u)^2}{ms}. \quad (53)$$

The efficient price of a trip equals marginal social cost:

$$\widehat{p}^u = \widehat{MSC}^n = \widehat{c}^u + \widehat{\tau}^u = \bar{\delta} + \frac{2\lambda \widehat{N}^u}{ms}. \quad (54)$$

Social surplus is equal to

$$\widehat{SS}^u = \int_{\widehat{p}^u}^{\infty} N(u) du + \tau^u \widehat{N}^u.$$

Finally, the welfare gain in switching from no fare to the optimal uniform fare is

$$G^{eu} = \widehat{SS}^u - \widehat{SS}^n.$$

4.2 Social optimum

The social optimum can be supported by imposing train-specific fares as per Eq. (44).

Total travel costs are derived by substituting Eq. (46) into the expression for TC^o given in Prop. 15:

$$\widehat{TC}^o = \bar{\delta} \widehat{N}^o + \frac{\lambda (\widehat{N}^o)^2}{ms} - RV^o.$$

Since variable revenue in Eq. (46) does not depend on the number of trips, the marginal social cost of a trip is

$$\widehat{MSC}^o = \bar{\delta} + \frac{2\lambda \widehat{N}^o}{ms}.$$

Similar to the optimal uniform-fare regime, the efficient price of a trip equals marginal social cost so that:

$$\widehat{p}^o = \widehat{MSC}^o (\widehat{N}^o) = \bar{\delta} + \frac{2\lambda \widehat{N}^o}{ms}. \quad (55)$$

Eqs. (55) and (54) reveal that the optimal price is the same function of usage in regimes u and o . This is consistent with Prop. 11 and the observation that, if the crowding cost function is linear, the marginal social cost of trips is the same in the SO and UE. Social surplus is equal to

$$\widehat{SS}^o = \int_{\widehat{p}^o}^{\infty} N(u) du + R^o (\widehat{N}^o) = \int_{\widehat{p}^o}^{\infty} N(u) du + \frac{\lambda (N^o)^2}{ms} + RV^o.$$

The welfare gain in switching from no fare to the SO-fare is

$$G^{eu} = \widehat{SS}^o - \widehat{SS}^e,$$

and the welfare gain in switching from the optimal uniform fare to the SO-fare is

$$G^{uo} = \widehat{SS}^o - \widehat{SS}^u.$$

4.3 Comparison of the regimes

Private costs in regimes n , u and o are given by Eqs. (52), (54), and (55) respectively. For given values of m , s , and N , it is clear that private costs are the same in regimes u and o , and lower in regime n . With elastic demand this implies that equilibrium usage is the same in regimes u and o , and higher in regime n . Correspondingly, the equilibrium private cost and consumers' surplus is the same in regimes u and o , and higher in regime n . Social surplus is highest in regime o , lowest in regime n , and intermediate in regime u . These results are summarized in the following proposition.

Proposition 16. *For given values of m and s , and elastic demand, equilibrium private costs are the same in the optimal uniform-fare and SO-fare regimes, and lower in the no-fare regime: $\widehat{p}^o = \widehat{p}^u > \widehat{p}^n$.*

Equilibrium usage is the same in the optimal uniform-fare and SO-fare regimes, and higher in the no-fare regime: $\widehat{N}^o = \widehat{N}^u < \widehat{N}^n$.

Consumers' surplus is the same in the optimal uniform-fare and SO-fare regimes, and higher in the no-fare regime: $\widehat{CS}^o = \widehat{CS}^u < \widehat{CS}^n$.

Social surplus is highest in the SO-fare regime, intermediate in the optimal uniform-fare regime, and lowest in the no-fare regime: $\widehat{SS}^o > \widehat{SS}^u > \widehat{SS}^n$. Consequently, $\widehat{G}^{eo} > \widehat{G}^{eu} > 0$.

The results in Prop. 16 differ from those in the bottleneck model. In the bottleneck model, the equilibrium price of a trip for a given N is the same in the social optimum and no-toll user equilibrium, and higher in the uniform-toll equilibrium. Consequently, $\widehat{p}_{Bn}^u > \widehat{p}_{Bn}^o = \widehat{p}_{Bn}^n$ which contrasts with $\widehat{p}^u = \widehat{p}^o > \widehat{p}^n$ in Prop. 16. The rankings of usage and consumers' surplus also differ, and only the rankings of social surplus and the welfare gains of pricing are the same.

5 Optimal transit service

We now turn attention to the long run when the transit authority can choose m , s , and the timetable for the m trains. For tractability, we continue to assume that schedule delay costs are linear, and we now assume that the headway between trains is a constant, h . First we derive the optimal timetable for given values of m and s . Next, we derive properties of the optimal m and s for a general capacity cost function. Finally, we adopt a specific capacity function and derive analytical formulas for the optimal m and s while treating m as a continuous variable.

5.1 Optimal timetable

The optimal timetable is derived by minimizing users' total costs. In general, the optimal timetable for given m and s is not the same for the UE and the SO because their load patterns differ. However, the timetables are equal given linear schedule delay costs and a uniform headway as assumed here.

Since the timetable consists of m successive trains with a constant headway h , the timetable is fully described by the arrival time of the last train, t_m . The optimal value of t_m , which is derived in Appendix D, is described in the following proposition:

Proposition 17. *With the optimal timetable the last train leaves at time $t_m^o = t^* + h \left(m - \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor - 1 \times I_{\frac{\gamma m}{\beta + \gamma} > \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor} \right)$. Train k , $k = \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor + 1 \times I_{\frac{\gamma m}{\beta + \gamma} > \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor}$ arrives on time at t^* . The unweighted average schedule delay cost is $\bar{\delta} \simeq \frac{\beta \gamma}{\beta + \gamma} \frac{h}{2}$.*

According to Prop. 17, the higher is the unit cost of late arrival (γ) relative to early arrival (β) the earlier train service begins. The number of trains that arrive before t^* is approximately $\frac{\gamma}{\beta + \gamma}$. This formula is approximate because the number of trains is integer-valued. For the same reason, the formula for average schedule delay cost, $\bar{\delta}$, is approximate too.

5.2 General capacity function

Let $K(m, s)$ denote the cost of providing service including capital, operations and maintenance.¹⁷ To facilitate analysis, for the remainder of this section we treat m as a continuous variable. (The formula for $\bar{\delta}$ in Prop. 17 is then exact.) Function $K(m, s)$ is assumed to be a strictly increasing and differentiable function of m and s . As in Section 4, we first consider the uniform-fare regimes and then the social optimum.

Uniform-fare regimes

Let $p(N)$ denote the inverse demand curve corresponding to demand function (50). With a uniform fare, social surplus net of capacity costs is

$$SS^e = \int_{n=0}^N p(n) dn - \left(\bar{\delta}N + \frac{\lambda N^2}{ms} + K(m, s) \right). \quad (56)$$

The transit authority chooses m and s to maximize (56). To economize on notation, let K_m and K_s denote the derivatives of $K(m, s)$ with respect to m and s respectively. First-order conditions for a maximum are¹⁸

$$\frac{\lambda N^2}{ms^2} \cdot \frac{p_N N - \tau - \frac{d\tau}{dN} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} + \frac{\left(\tau - \frac{\lambda N}{ms} \right) \frac{d\tau}{ds} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} = K_s, \quad (57a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N \cdot \frac{p_N N - \tau - \frac{d\tau}{dN} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} + \frac{\left(\tau - \frac{\lambda N}{ms} \right) \frac{d\tau}{dm} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} = K_m. \quad (57b)$$

The right-hand side of (57a) is the marginal cost of expanding train capacity, and the left-hand side is the marginal benefit. Similarly, the right-hand side of (57b) is the marginal cost of adding a train, and the left-hand side is the marginal benefit.¹⁹ The derivative $\partial \bar{\delta} / \partial m$ given in Prop. (17) is constant.

¹⁷System costs are assumed to be independent of usage. Adding N as an argument of the service cost function would not affect results of interest.

¹⁸See Appendix E for the derivation.

¹⁹Since $K_m > 0$, the RHS of Eq. (57b) is positive. Provided $\left(\tau - \frac{\lambda N}{ms} \right) \frac{d\tau}{dm} > 0$, this guarantees that $\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} > 0$, or $N > \frac{\partial \bar{\delta}}{\partial m} \cdot \frac{sm^2}{\lambda}$. It is not clear that this condition is sufficient to guarantee condition (37), $n_k^e \geq 0$, for all k . However, m is treated in this section as a continuous variable. If m is restricted to integer values, the optimal number of trains is derived by increasing m in steps of one until the incremental net benefit becomes negative. With such a procedure, condition (37) is satisfied at the optimum.

No fare Setting the fare to zero in Eqs. (57a) and (57b) leads to the following proposition.

Proposition 18. *In the no-fare regime, optimal train capacity, s , and number of trains, m , are defined by the conditions*

$$\frac{\lambda N^2}{ms^2} \cdot \frac{p_N N}{p_N N - \frac{\lambda N}{ms}} = K_s, \quad (58a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N \cdot \frac{p_N N}{p_N N - \frac{\lambda N}{ms}} = K_m. \quad (58b)$$

The first term of the product on the left-hand side of (58a) is the marginal benefit from expanding train capacity if usage remained fixed. The cost of crowding would decrease by $\frac{\lambda N}{ms^2}$ for each of the N users. The actual reduction in crowding is smaller than this because the improved service quality attracts new users. Because usage is underpriced, the increase in usage is welfare-reducing which shrinks the benefit from greater capacity. This latent demand effect accounts for the second term of the product on the left-hand side which is less than 1. In the limit of perfectly elastic demand (i.e., $p_N \rightarrow 0$), the potential benefit from expanding capacity is completely dissipated. In the opposite limit of fixed demand considered in Section 3 (i.e., $p_N \rightarrow -\infty$), the second term converges to 1, and there is no dilution of benefit.

Eq. (58b) for m is interpreted similarly. The first term inside the brackets on the left-hand side is the marginal benefit per user from less crowding. The second term inside the brackets is the marginal disbenefit due to greater average schedule delay costs. This net benefit is diluted by the same factor as in Eq. (58a).

Optimal uniform fare With the optimal uniform fare, $\tau = \frac{\lambda N}{ms}$. Eqs. (57a) and (57b) lead to the following proposition:

Proposition 19. *In the optimal uniform-fare regime, optimal train capacity, s , and num-*

ber of trains, m , are defined by the conditions

$$\frac{\lambda N^2}{ms^2} = K_s, \quad (59a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N = K_m. \quad (59b)$$

In contrast to Eqs. (58a) and (58b), the marginal benefits from expanding train capacity and the number of trains in (59a) and (59b) are not diluted by additional usage because usage is priced efficiently. This might suggest that the optimal values of s and m , s_*^u and m_*^u , are larger than their counterparts with a zero fare, s_*^n and m_*^n . However, at least for given values of s and m , usage is higher in the no-fare regime as per Prop. 16. This leaves the rankings of s_*^u and s_*^n , and m_*^u and m_*^n , ambiguous in general. Moreover, unlike in the bottleneck model it is not possible as in Arnott et al. (1993) to derive simple rankings in terms of the elasticity of demand. This is because capacity has two dimensions (m and s) rather than one, and also because the user cost in (35) has a fixed component that is independent of usage.

Social optimum

With the SO-fare, social surplus net of capacity costs is given by

$$SS^o = \int_{n=0}^N p(N) - \left(\bar{\delta} N + \frac{\lambda N^2}{ms} + K(m, s) \right) + RV^o(m, s). \quad (60)$$

Eq. (60) is the same as Eq. (56) except for the last term, RV^o , which is a function of m and s , but does not depend on usage. In effect, net financial system costs in the social optimum are $K(m, s) - RV^o(m, s)$. Since usage is priced efficiently in both the SO-fare and optimal uniform-fare regimes, the first-order conditions for s_*^o and m_*^o are the same as (59a) and (59b) for s_*^u and m_*^u , with the derivatives of $K(m, s) - RV^o(m, s)$ in place of the derivatives of $K(m, s)$.

Proposition 20. *In the SO-fare regime, optimal train capacity, s , and number of trains,*

m , are defined by the conditions

$$\frac{\lambda N^2}{ms^2} = K_s - RV_s^o, \quad (61a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N = K_m - RV_m^o, \quad (61b)$$

where RV_s^o and RV_m^o denote the derivatives of $RV^o(m, s)$ with respect to s and m respectively.

The right-hand sides of Eqs. (61a) and (61b) are smaller than their counterparts for the optimal uniform toll, (59a) and (59b). The generation of variable revenue from the SO-fare effectively reduces the marginal financial cost of expanding either s or m . In the case of (61a) this implies that optimal train capacity conditional on the values of m and N is larger in the social optimum: $s_*^o(m, N) > s_*^u(m, N)$. Similarly, Eq. (61b) implies that the optimal number of trains conditional on the values of s and N is also larger in the social optimum: $m_*^o(s, N) > m_*^u(s, N)$.

These rankings may seem surprising given that total system costs are lower in the social optimum than the uniform-fare regime. Inequality $m_*^o(s, N) > m_*^u(s, N)$ is explained by the fact that ridership is distributed more evenly across trains in the social optimum. More users take the earliest and latest trains in the social optimum which makes adding extra trains more beneficial. To understand the inequality $s_*^o(m, N) > s_*^u(m, N)$, recall from Eq. (49) that in the uniform-fare regime the deadweight loss from imbalanced ridership between trains *increases* with s . Expanding capacity is therefore more valuable in the social optimum.

Despite the inequalities $m_*^o(s, N) > m_*^u(s, N)$ and $s_*^o(m, N) > s_*^u(m, N)$, there is no guarantee that the unconditionally optimal values (s_*^o, m_*^o) in the social optimum are both larger than their counterparts (s_*^u, m_*^u) . One reason is that $s_*^u(m, N)$ is a decreasing function of m , and $m_*^u(s, N)$ is a decreasing function of s , and one function can shift much more than the other. The other reason is that usage generally differs in the two regimes; i.e. $N_*^o \neq N_*^u$. To proceed further, we now adopt a specific capacity function.

5.3 A specific capacity function

Kraus and Yoshida (2002) distinguish in their model between the number of train runs and the number of train sets (a train set can make more than one run). They also account for the time required for a train set to make a round trip. These variables are absent from our model, and we adopt a simpler service cost function for transit of the form:

$$K(m, s) = (\nu_0 + \nu_1 s)m + \nu_2 s, \quad (62)$$

where ν_0 , ν_1 , and ν_2 are all non-negative parameters. The term $\nu_0 + \nu_1 s$ in (62) is the incremental capital and operating costs of running an additional train. It is a linear increasing function of train capacity. If $\nu_0 > 0$, there are scale economies with respect to train size. The second term in (62), $\nu_2 s$, accounts for costs that depend on train capacity but not the number of trains. Kraus and Yoshida (2002) interpret this term as capital costs for terminals.²⁰

In this subsection we focus on the optimal uniform fare and SO-fare because in these regimes the slope of the demand function does not affect the optimal values of m or s , and properties of the solution can be derived while treating N parametrically.

Optimal uniform fare

With the optimal uniform fare, the first-order conditions for s and m are Eqs. (59a) and (59b). Given the service cost function (62), these equations become

$$\frac{\lambda N^2}{ms^2} = \nu_1 m + \nu_2, \quad (63a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N = \nu_0 + \nu_1 s. \quad (63b)$$

Before solving (63a) and (63b) simultaneously, it is instructive to consider each equation by itself. Eq. (63b) can be solved for the conditionally-optimal number of trains considered

²⁰They note that the linear specification is valid if terminal cost is proportional to terminal area, and terminal area is proportional to train capacity.

in the previous subsection:

$$m_*^u(s, N) = \sqrt{\frac{\lambda}{s \left(\frac{\partial \bar{\delta}}{\partial m} N + \nu_0 + \nu_1 s \right)}} N. \quad (64)$$

Eq. (64) is of practical interest if the transit authority cannot adjust train capacity — perhaps because train platforms cannot be lengthened.²¹ As expected, the optimal number of trains decreases with train capacity. The number of trains increases with demand at a rate faster than \sqrt{N} , but slower than N . Service quality degrades because both the duration of the travel period and average train occupancy increase.

First-order condition (63a) can be solved to obtain a formula for conditionally-optimal train capacity:

$$s_*^u(m, N) = \sqrt{\frac{\lambda}{\nu_1 m^2 + \nu_2 m}} N. \quad (65)$$

Optimal train capacity decreases with m at a rate faster than $m^{-1/2}$. It also varies proportionally with N . Thus, if the transit authority cannot add trains but can introduce bigger trains, it adds sufficient capacity to maintain a given level of crowding on each train. Service quality remains constant. In this sense, users fare better if the transit authority can only expand train capacity than if it can only add more trains.²²

To solve simultaneously for the unconditionally optimal values, s_*^u and m_*^u , it is convenient to define the intermediate variable $Z \equiv \frac{\partial \bar{\delta}}{\partial m} N + \nu_0$. Eqs. (63a) and (63b) can be solved jointly to obtain a quartic equation for s_*^u :

$$\nu_1 (s_*^u)^4 + Z (s_*^u)^3 - \frac{\lambda Z^2}{\nu_2^2} N^2 = 0, \quad (66)$$

and another quartic equation for m_*^u :²³

$$\nu_1 (m_*^u)^4 + \nu_2 (m_*^u)^3 - \frac{\lambda \nu_2^2}{Z^2} N^2 = 0. \quad (67)$$

The characteristics of the solution depend on the relative magnitudes of parameters

²¹Train platforms may also have to be adjusted to accommodate wider trains: a problem that the French rail network, SNCF, has overlooked (Willsher, 2014).

²²This might not be true if the headway between trains can be reduced.

²³Note that Eq. (67) in m_*^u is the same as Eq. (66) in s_*^u if ν_2 and Z are interchanged.

ν_0 , ν_1 , and ν_2 . If $\nu_2 = 0$, there would be no fixed costs of expanding train capacity such as train station infrastructure. Costs would be minimized by reducing m toward zero while increasing s proportionally to maintain ms constant. Given the integer constraint on m this would imply operating one train large enough to accommodate all passengers and setting the timetable so that everyone arrives on time. Such a degenerate solution is implausible, and would render the model meaningless.

If $\nu_1 = 0$, the cost of a train is independent of its capacity. Eqs. (66) and (67) then have closed-form solutions:

$$s_*^u(N) = \left(\frac{\partial \bar{\delta}}{\partial m} \frac{\lambda}{\nu_2^2} N^3 + \frac{\lambda \nu_0}{\nu_2^2} N^2 \right)^{1/3}, \quad (68a)$$

$$m_*^u(N) = \frac{\lambda}{\nu_2 \left(\frac{\partial \bar{\delta}}{\partial m} \frac{\lambda}{\nu_2^2} + \frac{\lambda \nu_0}{\nu_2^2} \frac{1}{N} \right)^{2/3}}. \quad (68b)$$

According to Mohring's square-root rule, both optimal service frequency and the number of passengers carried per train (or bus) increase with \sqrt{N} . In the *PTC* model, service frequency is constant because headway is fixed. Eq. (68a) shows that s_*^u rises with N at a rate faster than $N^{2/3}$.²⁴ Eq. (68b) shows that m_*^u grows at a rate slower than $N^{2/3}$, but since it does increase with N the duration of the travel period increases. As N becomes very large, m_*^u approaches a constant value and s_*^u increases approximately linearly with N .^{25, 26}

With $\nu_1 = 0$, it is possible to show that equilibrium user cost is a U-shaped function of N with a minimum at $N = \nu_0 \left(\frac{\partial \bar{\delta}}{\partial m} \right)^{-1}$. However, both the equilibrium price, p , and the average system cost, $c^u(m_*^u, s_*^u) + \frac{1}{N} K(m_*^u, s_*^u)$, decline monotonically with N . This is attributable to the fact that, with $\nu_1 = 0$, the service cost function has constant returns to scale while the user cost function has increasing returns.

The final limiting case $\nu_0 = 0$ applies if there are no scale economies with respect to

²⁴In Kraus and Yoshida's (2002) model the effect of N on s is ambiguous. Nevertheless, they remark (p.178) that "with realistic parameters" s is likely to increase with N .

²⁵Eventually a physical limit to train capacity would be reached due to constraints on platform size or tractive power.

²⁶If $\nu_0 = 0$ as well as $\nu_1 = 0$, m_*^u is independent of N and s_*^u rises proportionally with N for all values of N . This is a highly unrealistic case since it means that procuring and operating trains is costless.

train size. Eqs. (66) and (67) simplify to:

$$\nu_1 + \frac{\partial \bar{\delta}}{\partial m} \frac{N}{s_*^u} - \frac{\lambda}{\nu_2^2} \left(\frac{\partial \bar{\delta}}{\partial m} \right)^2 \left(\frac{N}{s_*^u} \right)^4 = 0, \quad (69a)$$

$$\nu_1 (m_*^u)^4 + \nu_2 (m_*^u)^3 - \frac{\lambda \nu_2^2}{\left(\frac{\partial \bar{\delta}}{\partial m} \right)^2} = 0. \quad (69b)$$

Eq. (69a) solves for a unique value of N/s_*^u which implies that train capacity is chosen proportional to ridership. Eq. (69b) solves for a unique value of m_*^u which implies that the number of trains is independent of ridership. These properties immediately imply that equilibrium user cost, c^u , price, p^u , and average system cost are all constant. Hence, unlike in Mohring's model (in which $\nu_1 = \nu_2 = 0$, $\nu_0 > 0$) there are no scale economies with respect to traffic density. However, the case $\nu_0 = 0$ is similar to the bottleneck model in which optimal capacity is proportional to usage, and equilibrium user cost, price, and average system cost are constants (see Arnott et al. (1993)).

The degree of cost recovery from fare revenues is easily derived. Fare revenues are given by Eq. (53): $R^u = \frac{\lambda N^2}{m_u s_u}$. Given first-order condition (63a) this implies

$$R_*^u = (\nu_1 m_*^u + \nu_2) s_*^u.$$

The cost recovery ratio is therefore

$$\rho = \frac{R_*^u}{K(m_*^u, s_*^u)} = \frac{(\nu_1 m_*^u + \nu_2) s_*^u}{\nu_0 m_*^u + (\nu_1 m_*^u + \nu_2) s_*^u} \leq 1.$$

If there are no scale economies with respect to train size (i.e., $\nu_0 = 0$), fare revenues fully cover capacity costs. Otherwise, costs are only partially recovered and the service runs a deficit.

Social optimum

For the social optimum the first-order conditions for s and m are given in Eqs. (61a) and (61b). With the service cost function (62), the equations become

$$\frac{\lambda N^2}{ms^2} = \nu_1 m + \nu_2 - RV_s^o, \quad (70a)$$

$$\left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N = \nu_0 + \nu_1 s - RV_m^o. \quad (70b)$$

Unlike Eqs. (63a) and (63b), (70a) and (70b) cannot be solved to obtain useful expressions for s_*^o and m_*^o . As noted above, there is no guarantee in general that service quality is better in the social optimum than the uniform-fare regime in the sense that $s_*^o(N) > s_*^u(N)$ and $m_*^o(N) > m_*^u(N)$. Indeed, in the numerical example of Section 6 it turns out that $s_*^o(N) < s_*^u(N)$. However, with capacity function (62), $m_*^o(N) > m_*^u(N)$. This result is proved in Appendix F and recorded as:

Proposition 21. *For a given usage level, the optimal number of trains is greater in the social optimum than the optimal uniform-fare regime.*

Unlike for the optimal uniform-fare regime, there is no simple formula for the degree of cost recovery from SO-fare revenues. To derive further insights, and to rank m , s , and N for the three fare regimes, we now consider a numerical example.

6 A numerical example

The numerical example draws on recent empirical estimates of crowding costs, and it is calibrated to describe service on the Paris RER A line during the morning peak.²⁷ Base-case parameter values are: $\beta = 7.4$ [€/ (hr·user)], $\gamma = 17.2$ [€/ (hr·user)], $\lambda = 4.4$ [€/user], and $h = 2.5$ [min/train]. The demand function (50) is assumed to have a constant-elasticity form $N = N_0 p^\eta$ with $\eta = -1/3$.²⁸ Parameter N_0 and parameters ν_0 , ν_1 , and ν_2 of the capacity cost function are chosen to yield equilibrium values for the optimal

²⁷Parameter values are explained in Appendix G.

²⁸An elasticity of $-1/3$ is in the mid-range of empirical estimates (Oum et al., 2008, p.249; Litman, 2013). Consumers' surplus is infinite with $\eta > -1$. To enable comparisons of consumers' surplus between regimes, the area to the left of the demand curve is computed only for $p \leq \text{€}100$.

Table 4: Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: base-case parameter values

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	25.26	24	26.70
s	1,762	1,733	1,710
N	37,173	32,600	32,907
p	6.40	9.48	9.22
$Rev/user$	0	3.45	3.39
TCC	161,558	133,499	111,520
SDC	76,210	63,244	80,376
TC	237,768	196,743	191,896
K	138,270	134,889	136,528
R	0	112,407	111,520
ρ	0	0.833	0.817
CS	1,873,288	1,766,213	1,774,816
SS	1,735,018	1,743,732	1,749,807
$Total\ gain$		8,714	14,789
$Gain/user$		0.27	0.45
$Rel.eff$	0	0.59	1

uniform-fare equilibrium of $N^u = 32,600$, $m_*^u = 24$, $s_*^u = 1,733$, and a cost recovery rate of 5/6. The resulting values are: $N_0 = 69,003$ [users], $\nu_0 = 936.7$ [€/train], $\nu_1 = 0.1344$ [€/user], and $\nu_2 = 61.63$ [€·train/user]. Results for the three fare regimes are reported in Table 4.

6.1 Base-case results

No fare

With no fare, the equilibrium private cost (which equals the equilibrium user cost) is €6.40. There are $N^n = 37,173$ users who are accommodated in $m_*^n = 25.26$ trains with nominal capacities of $s_*^n = 1,762$. Total crowding costs (TCC^n) are more than double total schedule delay costs (SDC^n). Capital costs (K^n) are about 58% as large as total user costs (TC^n). Given no fare, the degree of cost recovery is zero.

Optimal uniform fare

The optimal uniform fare works out to $\tau^u = \text{€}3.45$. It boosts the equilibrium private cost to $p^u = \text{€}9.48$ which is $\text{€}3.08$ above the no-fare equilibrium price. Ridership drops to $N^u = 32,600$: about 12% below the no-fare level. Both the number of trains and train capacity are lower than with no fare although capacity costs are reduced by only 2.4%. Total crowding costs and total schedule delay costs are also lower than with no fare. By design, fare revenue of $R^u = 112,407$ covers a fraction $\rho^u = 0.833$ of capacity costs. Consumers' surplus is lower than with no fare, but social surplus is higher by $\text{€}8,714$ or about $\text{€}0.27$ per rider in the uniform-fare equilibrium. The relative efficiency of the optimal uniform fare can be measured by taking the no-fare and social optimum regimes as polar benchmarks and using the index

$$Eff_u = \frac{\widehat{SS}^u - \widehat{SS}^n}{\widehat{SS}^o - \widehat{SS}^n}.$$

With the base-case parameter values, $Eff_u \simeq 0.59$ so that the optimal uniform fare yields nearly 3/5 of the efficiency gains from the SO-fare.

Social optimum

The social optimum calls for more trains than either the no-fare or the uniform-fare regime. This is consistent with the result $m_*^o(N) > m_*^u(N)$ established for fixed demand in Prop. 21. However, train capacity is slightly lower than in the other two regimes. Ridership and consumers' surplus are slightly higher than with a uniform fare. Price, revenue per user and cost recovery are slightly lower. Crowding costs are significantly lower than in the other regimes, but schedule delay costs are higher because the SO-fare spreads usage more evenly over trains. Capacity costs are intermediate between the other regimes. Social surplus is higher than with no fare by about $\text{€}0.45$ per rider.

Short-run versus long-run gains from pricing

In Table 4, capacity is chosen optimally for each fare regime. Because rail transit capacity can take years to adjust, it is of practical interest to compare fare regimes in the "short run"

when capacity is fixed. If pricing is assumed to become more efficient over time, there are three cases to consider: regime u with capacity fixed at (m_*^n, s_*^n) , regime o with capacity fixed at (m_*^n, s_*^n) , and regime o with capacity fixed at (m_*^u, s_*^u) . Let G_x^{xy} denote the welfare gain in shifting from regime x to regime y when capacity remains fixed at its optimal level for regime x . With the base-case parameters one obtains $G_n^{nu} = \text{€}8,336$, $G_u^{uo} = \text{€}5,273$ and $G_n^{no} = \text{€}14,589$. By comparison, from Table 1 the long-run welfare gains when capacity is adjusted optimally are $G^{nu} = \text{€}8,714$, $G^{uo} = \text{€}6,076$ and $G^{no} = \text{€}14,788$. The long-run gains are higher by 4.5%, 15.2% and 1.4% respectively. The difference between short-run and long-run gains is appreciable only for G^{uo} . This is mainly because regimes u and o differ the most in terms of optimal number of trains²⁹

6.2 Sensitivity analysis

Integer-valued number of trains

The number of trains, m , has been treated as a continuous variable although it is discrete in reality. An integer constraint can be imposed by fixing m , and then choosing s using first-order conditions (58a), (59a) and (61a) for regimes n , u and o respectively. To assess how the integer constraint affects results, m was first set to the largest integer smaller than the real-valued solution and then the next integer larger. Thus, for the no-fare regime m was first set to $\lfloor m^n \rfloor$ and then $\lfloor m^n \rfloor + 1$. Since m^u was calibrated to be an integer value this was unnecessary for regime u . The integer value yielding the higher social surplus was then selected. The results changed very little, and social surplus was virtually unchanged. Integer constraints also had little effect for a range of other parameter values.

Demand elasticity

If the price elasticity of demand is reduced to $\eta = 0$, ridership is the same in the three fare regimes. With $p_N = -\infty$, the first-order conditions (57a) and (57b) for s and m are the same for regimes n and u so that $s_*^u = s_*^n$, and $m_*^u = m_*^n$. Imposing the uniform fare yields

²⁹Note that by Prop. 16, usage with the SO-fare and capacity fixed at (m_*^u, s_*^u) is the same as usage with the optimal uniform fare. Regimes u and o thus differ only in how passengers are distributed between trains.

Table 5: Comparison of no-fare, optimal uniform fare, and SO-fare (i.e., social optimum) regimes: $\eta = -2/3$

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	26.34	24	26.75
s	1,764	1,733	1,725
N	41,006	32,600	33,220
p	6.72	9.48	9.22
$Rev/user$	0	3.45	3.39
TCC	187,604	133,499	112,503
SDC	88,044	63,244	81,248
TC	275,648	196,743	193,751
K	139,632	134,889	137,558
R	0	112,407	112,503
ρ	0	0.833	0.818
CS	1,206,851	1,106,343	1,115,033
SS	1,067,219	1,083,862	1,089,978
$Totalgain$		16,643	22,759
$Gain/user$	0	0.51	0.70
$Rel.eff$	0	0.73	1

no welfare gain at all, and merely transfers money from users to the transit authority. The SO-fare does yield a welfare gain although (with ridership fixed at 32,600) it is only €0.185 compared to €0.45 in the base case.

To examine the effects of a higher price elasticity, η was doubled in magnitude to $-2/3$.³⁰ To maintain equilibrium ridership at 32,600 in the optimal uniform-fare regime, parameter N_0 was increased to 146,056. The results are shown in Table 5. With the higher price elasticity, consumers' surplus and social surplus in each regime are lower than with the base-case parameters. Regime u is otherwise unaffected. However, the welfare gain per rider nearly doubles from €0.27 to €0.51. The welfare gain per rider in the social optimum increases from €0.45 to €0.70, but by a smaller percentage so that the relative efficiency of regime u increases.

³⁰Few transit services are likely to face such a high elasticity – especially during peak travel times.

Table 6: Effects of increasing parameters β and γ (or parameter h) by 10%

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	25.26 \rightarrow 24.00 -4.98%	24.00 \rightarrow 22.80 -4.98%	26.70 \rightarrow 25.51 -4.44%
s	1,763 \rightarrow 1,792 +1.69%	1,733 \rightarrow 1,763 +1.70%	1,710 \rightarrow 1,738 +1.64%
N	37,173 \rightarrow 36,774 -1.07%	32,600 \rightarrow 32,278 -0.99%	32,907 \rightarrow 32,589 -0.96%
<i>Gains</i>		$G^{eu} : 8,714 \rightarrow 8,777$ +0.7%	$G^{uo} : 6,076 \rightarrow 6,458$ +6.3%

Schedule delay costs, headway and crowding costs

In all three fare regimes the equilibrium price is an increasing function of parameters β , γ , λ and h . Equilibrium usage decreases correspondingly. Because capacity is endogenous in this section, changing parameters β , γ , λ or h induces changes in s and m as well, and to determine the size of the effects it is necessary to solve for the new equilibria.

As a first experiment, parameters β and γ were both increased by 10%. The results are shown in Table 6. In each fare regime the number of trains drops by nearly 5% due to the higher cost of scheduling trains that arrive very early or late. Partly to compensate, train capacity increases by about 1.7%. Equilibrium prices rise, and usage drops slightly. Welfare gain G^{nu} increases by 0.7%, and welfare gain G^{uo} increases by 6.3%. An increase in headway, h , by a given percentage amount has exactly the same effect as an equal percentage increase in β and γ . Thus, the consequences of a 10% increase in h are as shown in Table 3.

As a second experiment, parameter λ was increased by 10%. The results are shown in Table 7. In all fare regimes the number of trains drops by about 3% while train capacity increases by just over 2%. Usage drops by about 1%. Welfare gains G^{nu} and G^{uo} both increase slightly.

Similar to Tables 4 and 5, Tables 6 and 7 depict long-run effects of changes in parameter values. These effects can differ significantly from the short-run effects when capacity is given. Consider, for example, welfare gain G^{uo} . In the short run with s and m fixed, G^{uo}

Table 7: Effects of increasing parameter λ by 10%

	Fare regime		
	No-fare (n)	Optimal uniform fare (u)	Social optimum (o)
m	25.26 \rightarrow 26.03 +3.02%	24.00 \rightarrow 24.73 +3.02%	26.70 \rightarrow 27.48 +2.92%
s	1,763 \rightarrow 1,800 +2.14%	1,733 \rightarrow 1,771 +2.14%	1,710 \rightarrow 1,747 +2.15%
N	37,173 \rightarrow 36,778 -1.06%	32,600 \rightarrow 32,249 -1.08%	32,907 \rightarrow 32,550 -1.08%
<i>Gains</i>		G^{eu} : 8,714 \rightarrow 8,926 +2.4%	G^{uo} : 6,076 \rightarrow 6,161 +1.4%

is given by Eq. (49). With a 10% increase in β and γ , G^{uo} rises by a factor of $(1.1)^2$, or 21% which is more than triple the 6.3% long-run increase shown in Table 6. A 10% increase in λ causes G^{uo} to fall by a factor $(1.1)^{-1}$ or about 9%. Yet Table 7 shows that the long-run gain actually rises by 1.4%.

The large differences between the short-run and long-run effects highlight the importance of the time horizon that is adopted for planning. For example, recent empirical research has led to improved estimates of the costs of public transport crowding. A rise in the estimated unit cost of crowding (i.e., parameter λ) might dissuade a planner with a short-run perspective from implementing train-dependent fares. A planner with a long-run perspective might adopt a more favorable view. This illustrates the well-known lesson in transportation economics that pricing and capacity investment decisions are interdependent, and should be considered jointly (Lindsey, 2012).

7 Conclusion

In this article we have analyzed the time pattern of usage and crowding on a commuter rail line using a model (the *PTC* model) of individuals' trip-scheduling preferences. Users face a trade-off between riding a crowded train that arrives at a convenient time, and riding a less crowded train that arrives earlier or later than desired. We solve user equilibrium for three fare regimes: no fare, an optimal uniform fare that controls the total number of users, and an optimal train-dependent fare that also controls the distribution of users

between trains. We also solve for the optimal long-run number and capacities of trains for the three fare regimes.

We find that some properties of the *PTC* model are similar to those of the Vickrey (1969) bottleneck model of automobile traffic congestion. In particular, the optimal train-dependent fare supports the social optimum because it internalizes the crowding-cost externality on each train. If the crowding cost function is linear, the revenues generated by the higher tolls on timely trains equal the welfare gains from imposing the toll. Some other properties of the *PTC* model resemble those of the flow congestion model of automobile traffic congestion due to Henderson (1974, 1981) and Chu (1995). Under plausible assumptions, passenger loads are distributed more evenly across trains in the social optimum than in the user equilibrium. Arrivals at the destination therefore occur at a more even rate, whereas in the bottleneck model the arrival rate is constant (and equal to bottleneck capacity) throughout the arrival period. Because crowding is assumed to occur at all levels of train occupancy it is impossible to eliminate crowding costs even if fares can be varied freely. Consequently, imposing Pigouvian fares makes users worse off – at least before accounting for how the revenues are used.

Still other properties of the *PTC* model differ from both the bottleneck and flow congestion traffic models. Most striking is that if the crowding cost function is convex, the short-run welfare gain from introducing optimal train-dependent fares decreases with total ridership. The marginal social cost of accommodating an additional passenger in the system is actually higher in the social optimum than with a uniform fare even though passengers are distributed optimally across trains.

Solving for optimal transit supply in the *PTC* model is complicated by the fact that capacity has two dimensions: the number of trains and the capacity of each train. We treat a special case with linear crowding and schedule delay cost functions, and a uniform headway between trains. The ranking of optimal capacity in the no-fare and optimal uniform-fare regimes is ambiguous in general. More users take transit in the no-fare regime, but the benefit from expanding capacity is diluted by latent demand. Expanding capacity is more valuable in the social optimum than the optimal uniform-fare regime because capacity is used more efficiently. The optimal number of trains is unambiguously higher

in the social optimum because more users take additional trains. Optimal train capacity is also higher in the social optimum if the number of trains is equal, but the ranking of capacity is ambiguous when the number of trains is optimized as well.

The comparative statics effects of varying parameter values when capacity is assumed to adjust are quite different from when capacity is held fixed. This demonstrates the importance of distinguishing between short-run and long-run planning horizons.

For illustration we calibrate the model to be roughly descriptive of the Paris RER A line during morning-peak conditions. With the base-case parameter values the welfare gain from implementing efficient pricing is €0.27 per user for the optimal uniform fare, and €0.45 for the optimal train-dependent fare. While these amounts may seem modest, the system-wide gains could be large. The RER A line carries more than 300 million users per year, and on average more than 1.5 million individuals used public transport in the Île-de-France region during the morning peak (7am-9am) in 2010.³¹ Given 250 working days per year, a welfare gain of about €0.50 per trip, and doubling the number of trips to account for evening travel, the annual total welfare gain from optimal pricing amounts to nearly €400 million per year. This figure is comparable to the social saving from a road traffic cordon congestion pricing scheme. Using the bottleneck for the Île-de-France road network modeled as a monocentric city, De Lara et al. (2013) found a social saving of €606 million per year.

The analysis in this paper could be extended in various directions. One is to allow travelers to differ in their trip-timing preferences and disutility from crowding. Doing so would allow consideration of the equity implications of alternative fare regimes and service investment policies. Another extension is to combine crowding costs with queuing delay. Both forms of congestion are often manifest in real transit systems. If a distinction is also made between seated and standing passengers, as in de Palma et al. (2015), passengers can experience congestion in a number of ways: delays when accessing stations and waiting on the platform, delays when trains are too full to board, delays while boarding, discomfort while seated, greater discomfort and possibly fatigue while standing, and delays while alighting at the destination and exiting stations. The analysis of such a system is likely to

³¹See p.11 in http://www.lvmt.fr/IMG/pdf/RAPA_Chaires_Stif_2013-2014_v1.pdf

be insightful but challenging.

Appendices

Appendix A Proof of Proposition 9

As in the text, let j index trains in order of decreasing schedule delay cost so that $\delta_1 > \delta_2 > \dots > \delta_m$. (Because trains arrive early and late, the index does not correspond to the temporal sequence in which trains are run.) Since in the UE $n_j^e = g^{-1}[c^e - \delta_j]$ and $g'(\cdot) > 0$, n_j^e increases with j : $n_1^e < n_2^e < \dots < n_m^e$.

We show that $n_j^e \leq n_j^o \iff n_j^e g'(n_j^e) \leq MSC^o - c^e$. Given $n_j^o = v^{-1}[MSC^o - \delta(t_j)]$, it follows that

$$\begin{aligned}
 n_j^e &\leq n_j^o \\
 \iff g^{-1}[c^e - \delta_j] &\leq v^{-1}[MSC^o - \delta_j] \\
 \iff v \{g^{-1}[c^e - \delta_j]\} &\leq MSC^o - \delta_j \\
 \iff c^e - \delta_j + g^{-1}[c^e - \delta_j] \times g' \{g^{-1}[c^e - \delta_j]\} &\leq MSC^o - \delta_j \\
 \iff g^{-1}[c^e - \delta_j] \times g' \{g^{-1}[c^e - \delta_j]\} &\leq MSC^o - c^e \\
 \iff n_j^e g'(n_j^e) &\leq MSC^o - c^e.
 \end{aligned}$$

Variables n_j^e and n_j^o have the same ranking as $n_j^e g'(n_j^e)$, the marginal external cost of crowding in the UE, and $MSC^o - c^e$, which is constant. Because total patronage, N , is fixed, some trains are more heavily loaded in the UE, and the others are more heavily loaded in the SO. Consequently, if $ng'(n)$ is a strictly increasing function of n (i.e., $\varepsilon(n) > -1$), there exists a unique train \hat{j} such that $n_j^e < n_j^o$ when $j < \hat{j}$, $n_j^e \geq n_j^o$, and $n_j^e > n_j^o$ when $j > \hat{j}$. Conversely, if $ng'(n)$ is a strictly decreasing function of n (i.e.,

$\varepsilon(n) < -1$), there exists a unique train \hat{j} such that $n_j^e > n_j^o$ when $j < \hat{j}$, $n_j^e \leq n_j^o$, and $n_j^e < n_j^o$ when $j > \hat{j}$.

Appendix B Proof of Proposition 10

Total fare revenues from the optimal uniform fare are $R^u = \tau^u N$. Hence

$$\frac{\partial R^u}{\partial N} = \tau^u + \frac{\partial \tau^u}{\partial N} N.$$

Now

$$MSC^u = \frac{\partial TC^u}{\partial N} = \frac{\partial (c^u N)}{\partial N} = c^u + \frac{\partial c^u}{\partial N} N = c^u + \tau^u.$$

Thus

$$\frac{\partial MSC^u}{\partial N} N = \left(\frac{\partial c^u}{\partial N} + \frac{\partial \tau^u}{\partial N} \right) N = \tau^u + \frac{\partial \tau^u}{\partial N} N = \frac{\partial R^u}{\partial N}. \quad QED$$

Total fare revenues from the SO-fare are $R^o = \sum_{k=1}^m \tau_k^o n_k^o$. Hence

$$\frac{\partial R^o}{\partial N} = \sum_{k=1}^m \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o \right) \frac{\partial n_k^o}{\partial N}.$$

The marginal social cost of a trip is the same for all trains that are used:

$$MSC^o = c_k^o + \tau_k^o.$$

Hence:

$$\begin{aligned} \frac{\partial MSC^o}{\partial N} &= \left(\frac{\partial c_k^o}{\partial n_k^o} + \frac{\partial \tau_k^o}{\partial n_k^o} \right) \frac{\partial n_k^o}{\partial N}, \\ \frac{\partial MSC^o}{\partial N} n_k^o &= \left(\frac{\partial c_k^o}{\partial n_k^o} n_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o \right) \frac{\partial n_k^o}{\partial N} = \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o \right) \frac{\partial n_k^o}{\partial N}, \\ \frac{\partial MSC^o}{\partial N} N &= \sum_{k=1}^m \frac{\partial MSC^o}{\partial N} n_k^o = \sum_{k=1}^m \left(\tau_k^o + \frac{\partial \tau_k^o}{\partial n_k^o} n_k^o \right) \frac{\partial n_k^o}{\partial N} = \frac{\partial R^o}{\partial N}. \quad QED \end{aligned}$$

Appendix C Proof of Proposition 11

We prove the case for which the welfare gain G^{eo} decreases with N . The proof for the case in which G^{eo} increases follows the same steps, and is omitted. As discussed in the text, G^{eo} decreases if the marginal social cost of an additional user is higher in the SO than the UE. Thus, it suffices to show that $MSC^o > MSC^e$.

As in Appendix A, let k index trains in order of decreasing schedule delay cost so that in the no-fare equilibrium, $n_1^e < n_2^e < \dots < n_m^e$. Equilibrium cost with no fare, c^e , is determined implicitly by Eq. (20):

$$\sum_{k=1}^m g^{-1}[c^e - \delta_k] - N = 0. \quad (71)$$

This equation can be written

$$\sum_{k=1}^m f[g(n_k^e) + n_k^e g'(n_k^e)] = N, \quad (72)$$

where $f(n) \equiv v^{-1}(n)$. Since $f(v(n)) = n$,

$$f'(n) = \frac{1}{v'(n)} = \frac{1}{2g'(n) + ng''(n)}. \quad (73)$$

The marginal social cost of a trip in the no-fare equilibrium is

$$MSC^e = \frac{\partial(c^e N)}{\partial N} = c^e + N \frac{\partial c^e}{\partial N}.$$

Using Eq. (71) to derive $\frac{\partial c^e}{\partial N}$ one obtains

$$MSC^e = c^e + \frac{N}{\sum_{k=1}^m \frac{1}{g'(n_k^e)}}. \quad (74)$$

The marginal social cost of a trip in the social optimum is defined implicitly by Eq. (31):

$$\sum_{k=1}^m f[MSC^o - \delta_k] = N. \quad (75)$$

By Assumption 4, the left-hand side of Eq. (75) is a strictly increasing function of MSC^o . Suppose we substitute eqn. (74) for MSC^e in place of MSC^o in Eq. (75). If the resulting left-hand side is less than N , then $MSC^o > MSC^e$ and the proof is complete. To economize on notation, let g_k denote $g(n_k^e)$, g'_k denote $g'(n_k^e)$, and n_k denote n_k^e . After a few substitutions one can write

$$\sum_{k=1}^m f[MSC^e - \delta_k] = \sum_{k=1}^m f \left[g_k + \frac{N}{m} \frac{m}{\sum_{k=1}^m \frac{1}{g'_k}} \right].$$

Define

$$mec_k \equiv g_k + n_k g'_k, \quad (76)$$

and

$$\widetilde{mec}_k \equiv g_k + \frac{N}{m} \frac{m}{\sum_{k=1}^m \frac{1}{g'_k}}. \quad (77)$$

Given Eq. (72), we need to prove that the following expression is negative:

$$\Delta F \equiv \sum_{k=1}^m f \left[\underbrace{g_k + \frac{N}{m} \frac{m}{\sum_{k=1}^m \frac{1}{g'_k}}}_{\tilde{n}_k} \right] - \sum_{k=1}^m \underbrace{f[mec_k]}_{n_k}.$$

Given Assumption 4, $\tilde{n}_k > n_k$ for small k , and $\tilde{n}_k < n_k$ for large k . The rankings of \tilde{n}_k and n_k , and of \tilde{c}_k and c_k , are shown in Figure 15.

Function $f()$ is concave by Assumption 6a. Clearly, for all trains

$$\tilde{n}_k - n_k < (\tilde{c}_k - c_k) f'[c_k], \quad k = 1 \dots m.$$

Using (76), (77) and (73) this implies

$$\begin{aligned} \Delta F &= \sum_{k=1}^m \tilde{n}_k - \sum_{k=1}^m n_k < \sum_{k=1}^m (\tilde{c}_k - c_k) f'[c_k] \\ &= \sum_{k=1}^m \left(\frac{N}{m} \frac{m}{\sum_{k=1}^m \frac{1}{g'_k}} - n_k g'_k \right) \frac{1}{2g'_k + n_k g''_k}. \end{aligned}$$

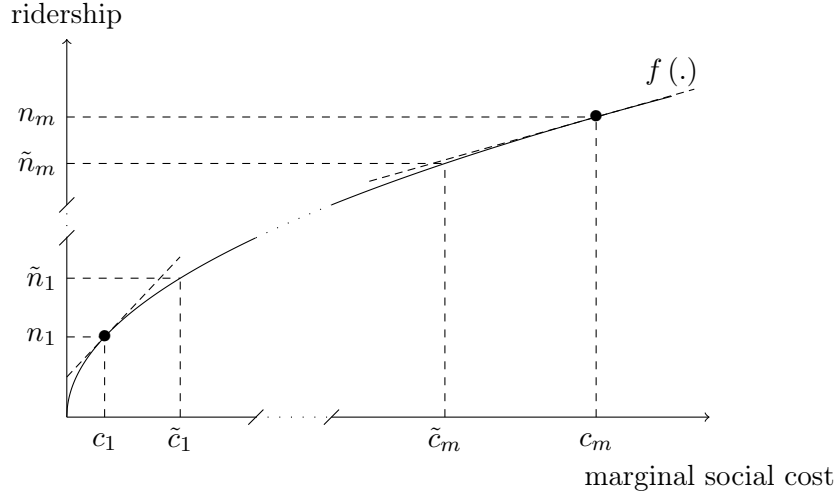


Figure 15: Ridership and marginal social cost

Now,

$$\sum_{k=1}^m \frac{1}{g'_k} = \frac{\sum_{j=1}^m \prod_{i \neq j} g'_i}{\prod_{i=1}^m g'_i}.$$

Hence

$$\begin{aligned} \Delta F &= \sum_{k=1}^m \left(N \frac{\prod_{i=1}^m g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} - n_k g'_k \right) \frac{1}{2g'_k + n_k g''_k} \\ &= \sum_{k=1}^m \left(N \frac{\prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} - n_k \right) \frac{g'_k}{2g'_k + n_k g''_k} \\ &= \sum_{k=1}^m \left(\frac{(\sum_{l \neq k} n_l) \prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} + \left(\frac{\prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} - 1 \right) n_k \right) \frac{g'_k}{2g'_k + n_k g''_k} \\ &= \sum_{k=1}^m \left(\underbrace{\left(\sum_{l \neq k} n_l \right)}_{(1)} \underbrace{\frac{\prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i}}_{(2)} - \underbrace{\frac{\sum_{j \neq k} \prod_{i \neq j} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i}}_{(3)} \underbrace{n_k}_{(4)} \right) \frac{g'_k}{2g'_k + n_k g''_k} \quad (78a) \end{aligned}$$

In the second line of eqn. (78a),

$$\begin{aligned}
& \sum_{k=1}^m \left(N \frac{\prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} - n_k \right) \\
&= N \sum_{k=1}^m \left(\frac{\prod_{i \neq k} g'_i}{\sum_{j=1}^m \prod_{i \neq j} g'_i} \right) - \sum_{k=1}^m n_k \\
&= N - N = 0.
\end{aligned}$$

Terms (1) and (2) in the last line of Eq. (78a) are decreasing functions of k . Terms (3) and (4) are increasing functions of k . Hence Eq. (78a) is negative if $\frac{g'_k}{2g'_k + n_k g''_k}$ is a non-decreasing function of k , or equivalently if $\varepsilon(n) = \frac{g''_k n_k}{g'_k}$ is a non-increasing function of k which is guaranteed by Assumption 6a. QED.

Appendix D Proof of Proposition 17

We first derive the optimal timetable for the UE, and then show that this timetable is also optimal for the SO.

Appendix D.1 Optimal timetable for user equilibrium

The optimal timetable is chosen to minimize total user costs. For the UE, total costs are given by Prop. 13:

$$TC^e = \bar{\delta}N + \frac{\lambda N^2}{ms}.$$

The timetable should therefore be chosen to minimize average schedule delay cost, $\bar{\delta}$. The timetable can be defined by the arrival time of the last train, t_m . It is clearly not optimal to set $t_m < t^*$, and have all trains arrive early, since $\bar{\delta}$ could be reduced by setting $t_m = t^*$. Similarly, it is not optimal to set $t_m > t^* + (m-1)h$, and have all trains arrive late, since $\bar{\delta}$ could be reduced by setting $t_m = t^* + (m-1)h$. Thus, one train must arrive during the interval $(t^* - h, t^*]$. Call it train \hat{k} . Train \hat{k} is the last train to arrive at or

before t^* . Average schedule delay cost is

$$\begin{aligned}
\bar{\delta} &= \frac{1}{m} \left(\sum_{k=1}^{\hat{k}} \beta (t^* - t_k) + \sum_{k=\hat{k}+1}^m \gamma (t_k - t^*) \right) \\
&= \frac{1}{m} \left(\sum_{k=1}^{\hat{k}} \beta (t^* - t_{\hat{k}} + h(\hat{k} - k)) + \sum_{k=\hat{k}+1}^m \gamma (t_{\hat{k}} - t^* + h(k - \hat{k})) \right) \\
&= \frac{1}{m} \left((t^* - t_{\hat{k}}) [(\beta + \gamma) \hat{k} - \gamma m] + (\beta + \gamma) h \frac{\hat{k}(\hat{k} - 1)}{2} + \gamma h \frac{m(m + 1 - 2\hat{k})}{2} \right) \quad (79a)
\end{aligned}$$

The first component of the right-hand side of eq. (79a), $(t^* - t_{\hat{k}})$, is the time between the arrival time of train \hat{k} and t^* . If $t_{\hat{k}} < t^*$ we can differentiate eq. (79a):

$$\frac{\partial \bar{\delta}}{\partial (t^* - t_{\hat{k}})} = \frac{(\beta + \gamma) \hat{k}}{m} - \gamma.$$

If $\hat{k} > \gamma m / (\beta + \gamma)$, then $\partial \bar{\delta} / \partial (t^* - t_{\hat{k}}) > 0$ and $\bar{\delta}$ is minimized by setting $t^* - t_{\hat{k}}$ to its minimal value, i.e $t^* - t_{\hat{k}} = 0$. Conversely, if $\hat{k} < \gamma m / (\beta + \gamma)$, then $\partial \bar{\delta} / \partial (t^* - t_{\hat{k}}) < 0$ and $\bar{\delta}$ is minimized by setting $t^* - t_{\hat{k}} = h$. Hence it is optimal to schedule one train at t^* . Call it train k^* . Replacing \hat{k} in eq. (79a) with k^* one obtains

$$\bar{\delta} = (\beta + \gamma) h \frac{k^*(k^* - 1)}{2m} + \gamma h \frac{m + 1 - 2k^*}{2}.$$

Treating k^* as a continuous variable for the moment, the first-order condition for minimizing $\bar{\delta}$ with respect to k^* is

$$k^{*o} = \frac{\gamma m}{\beta + \gamma} + \frac{1}{2}.$$

Since k^* is an integer, we have to compare $\bar{\delta}$ when $k^* = \lfloor k^{*o} \rfloor$ and when $k^* = \lfloor k^{*o} \rfloor + 1$.

We find

$$\bar{\delta}_{k^*=\lfloor k^{*o} \rfloor} - \bar{\delta}_{k^*=\lfloor k^{*o} \rfloor + 1} \leq 0 \iff \frac{\gamma m}{\beta + \gamma} \leq \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor.$$

Hence,

$$\begin{aligned} k^* &= \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor + 1 \times 1_{\frac{\gamma m}{\beta + \gamma} > \lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \rfloor} \\ t_m &= t^* + h \left(m - \left\lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \right\rfloor - 1 \times 1_{\frac{\gamma m}{\beta + \gamma} > \lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \rfloor} \right) \end{aligned}$$

In summary, if $\gamma m / (\beta + \gamma) > \lfloor \gamma m / (\beta + \gamma) + 1/2 \rfloor$, then $k^* = \lfloor \gamma m / (\beta + \gamma) + 1/2 \rfloor + 1$, and $t_m = t^* + h(m - 1 - \lfloor \gamma m / (\beta + \gamma) + 1/2 \rfloor)$. Conversely, if $\gamma m / (\beta + \gamma) < \lfloor \frac{\gamma m}{\beta + \gamma} + \frac{1}{2} \rfloor$, then $k^* = \lfloor \gamma m / (\beta + \gamma) + 1/2 \rfloor$ and $t_m = t^* + h(m - \lfloor \gamma m / (\beta + \gamma) + 1/2 \rfloor)$.

Appendix D.2 Optimal timetable for social optimum

Total costs in the social optimum are given by Prop. 15

$$TC^o = N\bar{\delta} + \frac{\lambda N^2}{ms} - \frac{s}{4\lambda} (\Delta - m\bar{\delta}^2).$$

TC^o differs from TC^e in including the third term. Recall that

$$\Delta - m\bar{\delta}^2 = \sum_{k=1}^m \delta_k^2 - \frac{1}{m} \left[\sum_{k=1}^m \delta_k \right]^2, \quad (80)$$

where

$$\delta_k = \beta [t^* - t_m + h(m - k)]^+ + \gamma [t_m - t^* - h(m - k)]^+. \quad (81)$$

As above, let \hat{k} be the last train to arrive at or before t^* . Differentiating (80) with respect to t_m , and using (81), it is possible to show after considerable algebra that

$$\frac{\partial (\Delta - m\bar{\delta}^2)}{\partial t_m} = \frac{\gamma}{\beta + \gamma} m + 1 - \hat{k}.$$

The term $\Delta - m\bar{\delta}^2$ therefore reaches an extreme point for the same \hat{k} as does $\bar{\delta}$. Hence TC^o reaches a minimum for the same timetable as TC^e .

Appendix E Derivatives of SS^e with respect to m and s

First-order conditions for a maximum of (56) are³²

$$\frac{\partial SS^e}{\partial s} = p(N) \frac{\partial N}{\partial s} - \left(-\frac{\lambda N^2}{ms^2} + \left(\bar{\delta} + \frac{2\lambda N}{ms} \right) \frac{\partial N}{\partial s} + K_s \right) = 0, \quad (82a)$$

$$\frac{\partial SS^e}{\partial m} = p(N) \frac{\partial N}{\partial m} - \left(\frac{\partial \bar{\delta}}{\partial m} N - \frac{\lambda N^2}{m^2 s} + \left(\bar{\delta} + \frac{2\lambda N}{ms} \right) \frac{\partial N}{\partial m} + K_m \right) = 0. \quad (82b)$$

The private cost of usage is given by Eq. (51) which can be written

$$p(N) - \left(\bar{\delta} + \frac{2\lambda N}{ms} \right) = \tau - \frac{\lambda N}{ms}. \quad (83)$$

The fare, τ , depends on the pricing regime. To maintain generality we assume for the moment that it can depend on N , m , and s . Substituting (83) into (82a) and (82b) yields:

$$\frac{\lambda N^2}{ms^2} + \left(\tau - \frac{\lambda N}{ms} \right) \frac{\partial N}{\partial s} - K_s = 0, \quad (84a)$$

$$\frac{\lambda N^2}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} N + \left(\tau - \frac{\lambda N}{ms} \right) \frac{\partial N}{\partial m} - K_m = 0. \quad (84b)$$

The demand derivatives are obtained by totally differentiating (51):

$$\frac{\partial N}{\partial s} = \frac{-\frac{\lambda N}{ms^2} + \frac{d\tau}{ds}}{p_N N - \frac{\lambda}{ms} - \frac{d\tau}{dN}} > 0, \quad (85a)$$

$$\frac{\partial N}{\partial m} = \frac{\frac{\partial \bar{\delta}}{\partial m} - \frac{\lambda N}{m^2 s} + \frac{d\tau}{dm}}{p_N N - \frac{\lambda}{ms} - \frac{d\tau}{dN}} > 0. \quad (85b)$$

Substituting (85a) and (85b) into (84a) and (84b), the first-order conditions become

$$\begin{aligned} \frac{\lambda N^2}{ms^2} \cdot \frac{p_N N - \tau - \frac{d\tau}{dN} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} + \frac{\left(\tau - \frac{\lambda N}{ms} \right) \frac{d\tau}{ds} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} &= K_s, \\ \left(\frac{\lambda N}{m^2 s} - \frac{\partial \bar{\delta}}{\partial m} \right) N \cdot \frac{p_N N - \tau - \frac{d\tau}{dN} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} + \frac{\left(\tau - \frac{\lambda N}{ms} \right) \frac{d\tau}{dm} N}{p_N N - \frac{\lambda N}{ms} - \frac{d\tau}{dN} N} &= K_m. \end{aligned}$$

³²Given $\bar{\delta} = \beta\gamma/(\beta + \gamma)hm/2$ as per Prop. (17), $\partial \bar{\delta}/\partial m = \beta\gamma/(\beta + \gamma)h/2$ which is a constant.

Appendix F Proof of Proposition 21

Conditional on m and N , s_*^u is given by Eq. (65). First-order condition (70a) can be rearranged to obtain an analogous equation for s_*^o . Recalling from Eq. (46) that $RV^o = \frac{s}{4\lambda} (\Delta - m\bar{\delta}^2)$, where $\Delta = \sum_{k=1}^m \delta_k^2$, the two equations for s_*^u and s_*^o can be written together as

$$s_*^j(m, N) = \sqrt{\frac{\lambda}{m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)]}} N, \quad (86)$$

where $X = 0$ for $j = u$, and $X = \frac{1}{4\lambda} > 0$ for $j = o$. Substituting (86) into the first-order conditions (67) and (70b) for m_*^u and m_*^o respectively, one obtains

$$\begin{aligned} & \nu_0 + \nu_1 \sqrt{\frac{\lambda}{m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)]}} N + \frac{\partial \bar{\delta}}{\partial m} N \\ & \quad - \frac{\lambda N}{m^2} \sqrt{\frac{m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)]}{\lambda}} \\ & - X \sqrt{\frac{\lambda}{m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)]}} N \frac{\partial (\Delta - m\bar{\delta}^2)}{\partial m} = 0. \end{aligned} \quad (87)$$

Function (87) is negative for small values of m , and over the relevant range it is increasing in m . Hence, if (87) is decreasing in X , $m_*^o > m_*^u$. Retaining only terms in (87) that depend on X , and multiplying through by $m^2 \lambda^{-\frac{1}{2}} \sqrt{m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)]} / N$, one obtains

$$\begin{aligned} & \nu_1 m^2 - m [\nu_1 m + \nu_2 - X (\Delta - m\bar{\delta}^2)] - X m^2 \frac{\partial (\Delta - m\bar{\delta}^2)}{\partial m} \\ & = -m\nu_2 - X m \left[m \frac{\partial (\Delta - m\bar{\delta}^2)}{\partial m} - (\Delta - m\bar{\delta}^2) \right]. \end{aligned} \quad (88)$$

This expression is decreasing in X if $\Delta - m\bar{\delta}^2$ is a convex function of m . Setting $k^* = \frac{\gamma}{\beta + \gamma} m$,

we find

$$\begin{aligned}\Delta - m\bar{\delta}^2 &= \frac{\beta\gamma mh^2}{12} \left[\frac{\beta\gamma m^2}{(\beta + \gamma)^2} + 2 \right]; \\ \frac{\partial (\Delta - m\bar{\delta})}{\partial m} &= \frac{\beta\gamma h^2}{12} \left[3 \frac{\beta\gamma m^2}{(\beta + \gamma)^2} + 2 \right] > 0; \\ \frac{\partial^2 (\Delta - m\bar{\delta})}{\partial m^2} &= \frac{\beta\gamma h^2}{12} \left[6 \frac{\beta\gamma m}{(\beta + \gamma)^2} \right] > 0.\end{aligned}$$

Hence $\Delta - m\bar{\delta}^2$ is a convex function of m , Eq. (88) is decreasing in X , and $m_*^o > m_*^u$. QED.

Appendix G Parameter values for numerical example

The numerical example requires base-case parameter values for β , γ , λ and h , and target values for N , m and s . The operating period was set to one hour, and target values were chosen for the optimal uniform-fare regime. This regime is intermediate in efficiency between the no-fare and SO-fare regimes, and it is arguably the most descriptive of public transit service in Paris where fares are positive and constant throughout the day.

Consider first the supply-side parameters m , h and s . According to the document “Schéma Directeur du RER A” written in June 2012 by the STIF (Syndicat des Transport d’Île-de-France), 30 trains per hour are supposed to operate during the morning peak in the East-West direction on the RER A line. However, the frequency actually achieved over the 4-year period February 2008 to February 2012 was only 24.4 trains per hour.³³ The target value for number of trains was thus set to $m = 24$, and the headway was set to $h = \frac{60}{24} = 2.5$ mins.

Two types of bi-level train sets are operated during the morning peak:³⁴

- MI2N train sets with 904 seats and standing room for 1,636 users (4 users/ m^2) for a total capacity of 2,540

³³See p.36 in http://www.stif.org/IMG/pdf/Deliberation_no2012-0163_relative_au_schema_directeur_du_RER_A.pdf.

³⁴See p.54 in http://www.stif.org/IMG/pdf/Deliberation_no2012-0163_relative_au_schema_directeur_du_RER_A.pdf.

- MI09 train sets with 948 seats and standing room for 1,683 users (4 users/ m^2) for a total capacity of 2,614 users

This suggests a value for capacity of about $s = 2,600$. However, in the model users are assumed to travel from a single origin to a single destination whereas the RER A line serves many stations. La Défense is the most popular destination, but a substantial fraction of users pass through it. Only part of train capacity is thus effectively devoted to users who exit at La Défense. After experimentation with alternative values of s , and other parameters described below, we settled on a capacity equal to two-thirds of nominal train capacity so that $s = \frac{2}{3} \cdot 2,600 = 1,733$.

Consider now the demand-side parameters. According to a January 2011 document “Étude La Défense Analyse des Trafics” prepared by the DRIEA (Direction Régionale et Interdépartementale de l’Équipement et de l’Aménagement), in 2009, 32,600 users arrived at La Défense by RER A between 8:25am and 9:25am.³⁵ This count includes users traveling in both East-West and West-East directions, but it excludes users who are passing through. Including travel in both directions results in overestimation of N , whereas excluding users who pass through La Défense results in underestimation of N . Lacking an indication as to which bias dominates and given the 2/3 coefficient applied to the initial capacity, we set $N = 32,600$.

Wardman et al. (2012) conduct a meta-analysis of estimates of β , γ and the value of travel time; call it α . They report point estimates of $\beta = 0.74 \cdot \alpha$ and $\gamma = 1.72 \cdot \alpha$ (see Table 19, p.25). For commuters in France, $\alpha = \text{€}15/\text{hr}$ (see Table 15, p.21) which is consistent with the government-recommended value. This suggests setting $\beta = 0.74 \cdot 15 = \text{€}11.1/\text{hr}$, and $\gamma = 1.72 \cdot 15 = \text{€}25.8/\text{hr}$. However, in the model it is assumed that users have the same desired arrival time, t^* . In reality, trip-timing preferences vary. The assumption of a common t^* leads to overestimation of schedule delay costs. In addition, with $\beta = \text{€}11.1/\text{hr}$ and $\gamma = \text{€}25.8/\text{hr}$., condition (37) that all trains are used was violated given plausible values for other parameters. After experimentation with alternative values of β , γ and s (noted above) we scaled down β and γ by one-third to $\beta = \text{€}7.4/\text{hr}$ and $\gamma = \text{€}17.2/\text{hr}$.

³⁵See Figure 2 on page 8 in http://cpdp.debatpublic.fr/cdpd-grandparis/site/DEBATPUBLIC_GRANDPARIS_ORG/_SCRIPT/NTSP_DOCUMENT_FILE_DOWNLOADCB59.PDF.

Empirical studies of public transit crowding often report crowding costs as time multipliers. This is consistent with evidence that disutility from crowding is proportional to amount of time spent in crowded conditions. The crowding cost parameter can then be written

$$\lambda = \alpha \cdot tt \cdot (tm - 1), \quad (89)$$

where tt is travel time and tm is the time multiplier.

According to the survey “Etude mobilité transports à la Défense - Profils, usages et modes de déplacements des salariés et habitants du quartier d’affaires” by the EPAD (Établissement Public de la Région Pour l’Aménagement de la Défense), in 2006, the average travel time incurred by public transport riders who used only one transport mode to reach La Défense was 40 mins.³⁶ This is consistent with a study by the Enquête Global Transport in 2010 which found an average travel time for commuters of 41 mins.³⁷ We thus set $tt = 40$ mins or $2/3$ hrs.

Haywood and Koning (2015) have estimated time multipliers for Paris. They obtain a linear approximation of the time multiplier (see Eq. (10), p.194) of $tm = 1 + 0.11 \cdot d$, where d is the density of passengers per square metre. Substituting the estimates of α , tt and tm into Eq. (89) one obtains $\lambda = 15 \cdot 2/3 \cdot 0.11 \cdot d$. With a density of 4 users/ m^2 for standing room on the train sets used on the RER A line (see above), this yields $\lambda = 4.4$.

Appendix H Glossary

Appendix H.1 Latin characters

c : user cost of a trip [€/user]

CS : total consumers’ surplus [€]

e : superscript for uniform-fare regime

$g(n)$: expected crowding cost function [€/user]

G^{xy} : welfare gain in shifting from pricing regime x to y

³⁶See p.11 in http://www.ladefense-seine-arche.fr/fileadmin/site_internet/user_upload/8-ENLIEN/etudes/etude-mobilite-transports.pdf.

³⁷See p.3 in http://www.driea.ile-de-france.developpement-durable.gouv.fr/IMG/pdf/Fiche_Actifs__cle0cecb9.pdf

h : time interval between successive trains [hr/train]

k : index of train

K : capacity cost function [€]

m : number of trains used [trains]

MEC : marginal external cost of a trip [€/user]

MSC : marginal social cost of a trip [€/user]

n : superscript for no-fare regime

n : number of users on a train [users]

n_k : number of users taking train k [users/train]

N : total number of users [users]

o : subscript for socially-optimal fare regime

p : private trip cost including toll [€/user]

R : total fare revenue [€]

RV : variable fare revenue from socially optimal fare schedule [€]

s : measure of train capacity [users/train]

SDC : total schedule delay costs [€]

SS : social surplus [€]

t : departure time from origin station [clock time]

t^* : desired arrival time at destination [clock time]

TC : total user costs [€]

TCC : total crowding costs [€]

u : superscript for optimal uniform-fare regime

$v(n)$: marginal social crowding cost function [€/user]

Appendix H.2 Greek characters

β : cost per minute of arriving early [€/(hr·user)]

γ : cost per minute of arriving late [€/(hr·user)]

δ : schedule delay cost function [€/user]

ε : elasticity of $g'(n)$

η : elasticity of demand

λ : crowding cost parameter [€/user]

τ : fare [€/user]

ν_0 : Capacity cost function coefficient on m [€/train]

ν_1 : Capacity cost function coefficient on $m \cdot s$ [€/user]

ν_2 : Capacity cost function coefficient on s [€·train/user]

Chapter III

Well-being in public transport: an empirical approach of the crowding effect

1 Introduction

Over the last decade, a growing body of research has focused on the qualitative attributes offered by public transportation (PT in what follows, see Litman, 2008, or OECD, 2014 for reviews). Whether the objective is to attract car drivers (to reduce pollutants' emissions in dense urban areas) or to avoid PT passengers from switching away from clean modes (because of unpleasant journey conditions), transport operators should implement policies that improve the “travel experience” of users. Among the numerous relevant qualitative attributes of PT, in-vehicle crowding - broadly understood as the density of users within carriages - is often singled out as one of the most desirable dimension (Eboli and Mazzulla, 2007; Dell'Olio et al., 2011; CRCFRI, 2012).

Economists currently exhibit a growing interest for in-vehicle crowding, especially by assessing its effects on the “generalized cost” of PT usage. As stressed theoretically (Kraus, 1991; de Palma et al., 2015, 2014) and confirmed empirically (Li and Hensher, 2011; Wardman and Whelan, 2011), time resources are dramatically more costly for individuals if con-

sumed in crowded PT. This can have strong organizational consequences because workers suffering from unpleasant PT journeys are more likely to arrive late at their job's places, to have a reduced productivity but also to plan quitting earlier their activities Mohd Mahudin et al., 2012; Tirachini et al., 2013. Also, travelers who have a car but use PT are more sensitive to perceived attributes of PT service, such as crowding (see Redman et al., 2013, for a systematic qualitative review). As for road congestion, in-vehicle crowding may influence individuals' route or mode choices, even if the PT system does not work as a "bottleneck" yet (when excessive PT usage generates objective time losses). Therefore, this feature of travels should be considered accurately when setting optimal PT supply or pricing schemes (Kraus, 1991; De Borger and Wouters, 1998; de Palma et al., 2014; Prud'homme et al., 2012; Tirachini et al., 2010; Kilani et al., 2014).

Importantly, economists rarely differentiate in-vehicle density and crowding (see for example Wardman and Whelan, 2011). According to psychologists, however, an important distinction has to be done. In a seminal work, Stokols (1972) defines the density as a "physical condition" involving spatial limitation but no psychological meaning, and the crowding as an "experimental state". As a consequence, in-vehicle density is not sufficient to describe the crowding experienced by an user during a PT journey. Other characteristics, such as personal characteristics or spatial factors, should also be considered. This distinction between the objective measure, the density, and the subjective experience, the crowding, has been modeled by Baum and Paulus (1987) and reviewed by Turner et al. (2005). Mohd Mahudin et al. (2012) propose a recent study on the psychological dimensions of rail crowding¹. As this phenomenon is per definition a perceived one, it seems relevant to use a subjective measure to address it.

The subjective well-being (SWB in what follows) is now largely accepted by economists (see Van Praag et al., 2003 or Kahneman and Krueger, 2006). It is commonly defined as "a person's cognitive and emotional evaluations of his or her life as a whole" (Diener et al.,

¹Mohd Mahudin et al. (2012) distinguish three components of the experience of passenger crowding (evaluation of psychosocial aspects of the crowded situation, emotional reactions to the crowded situation and evaluation of the ambient environment of the crowded situation) to build a valid instrument captures the relationship between crowding and the experience of stress and feelings of exhaustion. They find that the link between rail passenger crowding and the negative outcomes is mediated by emotional feelings of crowd.

2009). In the theoretical framework proposed by Van Praag et al. (2003), the SWB of a journey is one of the “Domain Satisfactions” (like health or financial satisfaction) forming the “General Satisfaction”. Since individuals are able to rate about how they are feeling during long or short periods of life (Van Praag and Ferrer-i Carbonell, 2008), Metcalfe and Dolan (2012) argue that SWB is a good measure of the underlying utility of a transport journey and that it should be the primary objective of any transport policies. Besides Ettema et al. (2010) recommend to use the SWB as a complement to decision utility derived from observed (route or mode) choices. According to them, instruments used to measure SWB, such as Likert items, are also valid in transport contexts.

Following this, a growing literature studying mobility and SWB has emerged. Delbosc (2012) provides an extensive description of the relation between travel and well-being whilst regretting the lack of studies in this field. Ettema et al. (2011) nevertheless propose a tool, the Satisfaction with Travel Scale (STS), used to address the satisfaction due to the in-vehicle activities performed during PT commutes. They find that discussion with other users has the strongest positive effect on STS, whereas activities related to entertainment and relaxation lead to lower STS. Also, Abou-Zeid and Ben-Akiva (2011) show that the overall travel satisfaction increases when the comparative happiness is higher, due to favorable comparisons to other commuters. In an experiment conducted in Switzerland on 30 car drivers switching to PT, Abou-Zeid et al. (2012) observe the dynamics of satisfaction and focus on the change in the PT perception. More recently, Abou-Zeid and Ben-Akiva (2014) review the application of SWB to transportation and they address its role in mode choices. Closer from this work, Cantwell et al. (2009) split up the satisfaction for PT into three elements: crowding, travel time reliability and monetary cost. Using an on-line survey on commuting in Dublin, Ireland, they analyze stated preference scenarios and find that the utility derived from a journey increases when the crowding decreases. This result may explain Morris and Guerra, 2014’s findings. Controlling for various factors, but not in-vehicle crowding, these authors highlight that PT travels are associated with lower users’ satisfaction compared to competing modes.

We derive from this literature a very simple theoretical framework described in Figure 16. In line with Cantwell et al. (2009), the satisfaction of a PT journey is explained by

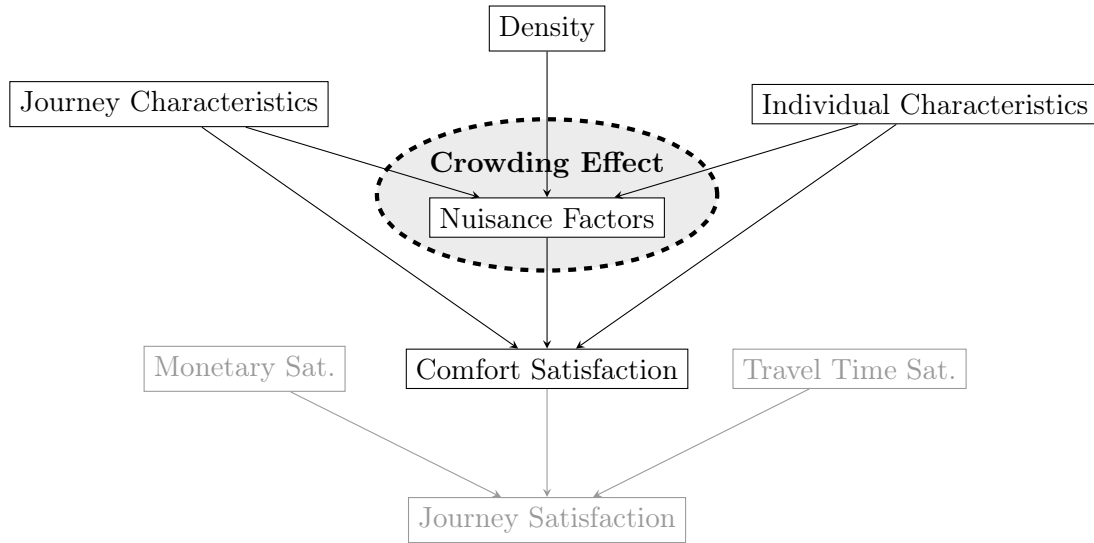


Figure 16: A simple journey satisfaction theoretical framework

the monetary satisfaction, the comfort satisfaction and the travel time satisfaction. In this paper, we focus on the comfort satisfaction and do not deal with the monetary and travel time parts. The comfort satisfaction is driven by three effects: the individual, the journey and the crowding effects. The crowding effect may be different for each user: it may depend on the travel characteristics, on the individual preferences and, of course, on the (objective) in-vehicle density. Moreover, the crowding effect is playing through various nuisance factors, here defined as *the features of a journey that are deteriorated by a high density*.² Thus high density in carriages is not costly by itself, it decreases PT users' satisfaction because it affects different aspects of the crowding experienced during a journey.

Within such a framework, we address two research questions. (i) How does the in-vehicle density relate to subjective comfort satisfaction? We identify the crowding effect by looking at the relationship between in-vehicle density and subjective comfort satisfaction stated by users. We also question the interdependency between the crowding effect and the travel time. (ii) What are the nuisance factors the crowding comes through? We test various nuisance factors (Smell, Noise, Standing...) as channels through which the crowding effect may influence the comfort satisfaction. Doing this, we describe the anatomy of the

²In this study, we consider eight nuisance factors: *Over closeness, Standing, Noise, Smell, Time loss, Waste of time, Fall and Robbery*.

crowding effect. This empirical work is done through econometric regressions of individual self-reported measures of satisfaction. Data originates from a field survey conducted late 2010 on platforms of Paris subways. As discussed by Haywood and Koning (2015), the Paris PT network constitutes a perfect case study to address in-vehicle crowding due to the recent growth in its patronage.

The main contributions of this work are the following. First, we directly assess the perception of crowding and its impact on the satisfaction of PT users, in line with the empirical literature on SWB (Kahneman and Krueger, 2006) or job satisfaction (Clark and Senik, 2010), whereas economists generally estimate “time multipliers” to integrate in-vehicle crowding to the generalized cost of PT by proposing to PT users hypothetical trade-offs between travel time and density (see Wardman and Whelan, 2011). We notably identify a crowding effect independent of the travel time, as opposed to most valuation studies. Second, we describe the underlying mechanisms playing within the crowding effect. Doing so, we also address the issue of the reliability of the self-reported satisfaction variable³ with the help of extra self-reported measures of nuisance factors related to PT crowding. This allows to decrease the bias of the comfort satisfaction measure. Third, due to the multidimensional reality of PT crowding, having a better idea of the nuisances that really affect users could also help public deciders implementing the most relevant policies. This study could thus highlight whether individuals will be better-off if they are offered additional seats, efficient cooling systems or more security in carriages.

The paper proceeds as follows. Section 2 presents the survey design as well as data used for our empirical exercises. Section 3 then focuses on the identification and on the measurement of the crowding effect. Section 4 studies the anatomy of the crowding effect, notably by looking at the nuisance factors that drive comfort satisfaction. Section 5 concludes.

³See Krueger and Schkade (2008) for a study of this reliability.

2 Data

2.1 Survey design

The data have been collected in the Parisian mass transit network late 2010. Users were interviewed directly on platforms of subway lines 1 and 4, during morning (7:30-10am) and evening (5-7:30pm) peaks.⁴ The stations where the survey has been conducted are, from East to West, *Gare de Lyon*, *Hôtel de Ville*, *Champs Elysées*, *Georges V*, *Argentine* and *Esplanade* for line 1, and, from South to North, *Denfert-Rochereau*, *Montparnasse-Bienvenüe*, *Saint Sulpice*, *Odéon* and *Les Halles* for line 4.⁵ Subway line 1 crosses Paris East-West. It is the busiest service of the subway network with 750,000 daily users in 2010. It connects most of business and touristic centers of the city, “*La Defense*” notably, i.e. the most important business district in Europe. Subway line 4 crosses Paris North-South and is the second most used service of the network, with 670,000 daily travelers in 2010. It connects users to three national trains stations: *Gare du Nord*, *Gare de Lyon* and *Gare Montparnasse*. Taken jointly, lines 1 and 4 present an important socioeconomic heterogeneity because it gives access to both wealthy and poor neighborhoods of Paris city. This heterogeneity is useful to assess the taste of individuals for crowding.

Around 1,000 PT users were surveyed whilst waiting for their train (whose service frequency may reach 1 minute 45 seconds during peaks) to arrive. Such a procedure was chosen in order to soften selection bias arisen from selective non-response. Platforms are almost the only place where subway users stop walking. As a consequence, the questionnaire had to be short and accounted only 10 questions.⁶

First, PT users were shown a density show-card (see Figure 17) and asked

Which users density do you expect to face during your immediate journey?

⁴Another part of this survey has been used to assess crowding costs in Paris subways with the use of contingent valuation techniques (see Haywood and Koning, 2015).

⁵The location of these stations in Paris subway network is provided in Appendix A.

⁶Moreover, the questionnaire includes extra questions because another purpose of the survey was to value crowding costs in Paris subways with the use of contingent experiments (see Haywood and Koning, 2015). Respondents were proposed trade-offs between randomly varying in-vehicle travel times and in-vehicle crowding levels. More precisely, 800 users were asked about their willingness to travel more against reduced crowding (“willingness to pay”) and 200 users were asked about their willingness to travel less against increased crowding (“willingness to accept”).

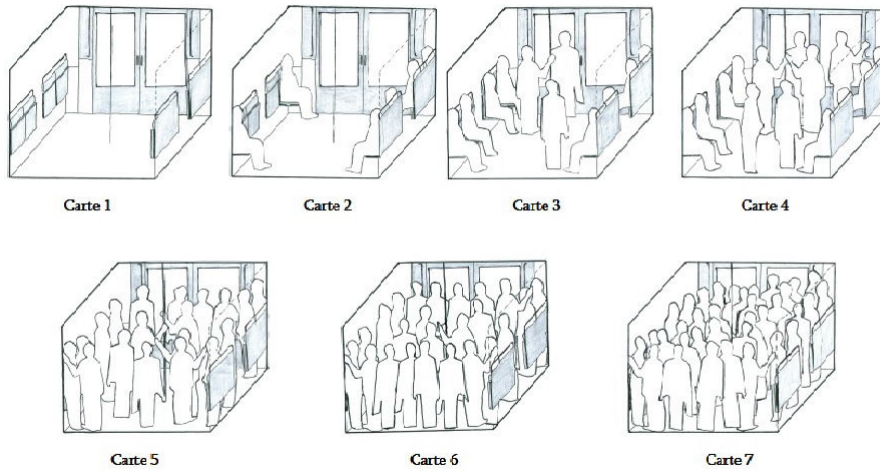


Figure 17: Density show-card used during the field survey

The crowding situations described on the show-card correspond to 0, 1, 2, 2.5, 3, 4 and 6 passengers per square meter respectively. We take this stated density as a measure of the In-Vehicle Density (*IVD*).

Then, the journey satisfaction question runs as follows:

Given this density, mark your satisfaction associated with the comfort for your immediate journey on a scale from 0 to 10. (Question A)

The mark ranges from 0 (the people are highly dissatisfied) to 10 (they are highly satisfied). We consider this mark as a measure of the in-vehicle comfort satisfaction experienced by the users during their immediate journey with the stated density, and refer to this mark as *CS* in the following.

Once the PT users had completed Question A, they were proposed either to follow-up with a more qualitative survey, accounting 10 additional questions, or to stop the survey. Clearly, there might be selection issues playing here because the people willing to answer the full questionnaire may be those who exhibit a special interest for crowding. The question on which we will focus in Section 4 was collected as follows: interviewers presented the worst crowding situation on the show-card (6 passengers per square meter) and asked to the interviewees:

On a scale from 0 to 10, mark the inconvenience associated with the following aspects when traveling in conditions similar to the ones represented on show-card 7: Over-closeness, Standing, Noise, Smell, Time loss, Waste of time, Fall and Robbery. (Questions B)

The mark ranges here from 0 (the people are not dissatisfied by the considered nuisance factor) to 10 (they are highly dissatisfied). These eight nuisance factors of crowding are defined as the features of a journey that are deteriorated by a high in-vehicle density. The reasons why these nuisance factors have been chosen are fully justified in Section 4. We refer to these dissatisfaction mark associated to the nuisance factor d as NFD_d .

2.2 Descriptive statistics

As made clear in Table 8, information is available for 999 travelers. This whole sample is divided in two sub-samples: sub-sample A is made of the 721 users that answered only to *Question A*, and sub-sample B made of the 278 other users who answered *Question A* and *Questions B*. Sub-sample B is used in Section 4 whereas in Section 3 we use the whole sample (A and B). Descriptive statistics of the peak hour⁷ subway users from the “Enquête Globale Transport” (EGT) are also displayed in the fourth column of Table 8. The EGT survey is conducted every ten years by the PT regulator in the Ile-de-France region. 18,000 households are surveyed and weighted to ensure sample representativeness at the regional scale.

Due to the survey design, a half of the interviewees travels on line 1 and a half has been asked during morning peaks. The population is almost equally distributed between men (48%) and women (52%). As expected, we observe a strong socioeconomic heterogeneity, since the income and the age are distributed in a regular way among the sample. The majority of the population is living in Paris city (53%). Only 37% of the individuals own a car, which is consistent with the important share of Parisians within the sample. The door-to-door travel time is 49 minutes, the in-vehicle travel time in lines 1 or 4 amounting only to 10 minutes. A huge majority of the sample is made of commuters (71%) who use lines 1 or 4 on a daily basis (64%).

⁷Peak hours are here defined as the 7:30-10am and 5-7:30pm periods.

We observe that 278 individuals have answered the qualitative questionnaire. As a consequence, we will be able to study the nuisance factors of in-vehicle crowding in Section 4 over around 28% of the initial sample. The main difference between both populations relates to the moments at which individuals were asked: only 36% of the sub-sample travels during morning peaks (as opposed to 50% initially)⁸. Then we notice only small differences between both populations. The sub-sample B accounts slightly less line 1 users (45%), less commuters (67%) and poorer individuals. By contrast, more Parisians (56%) and more individuals using the subway service on daily basis (67%) are found.

Importantly, people in sub-sample B may have different perceptions of the crowding effect. As a consequence, we pay attention to the selection mechanisms in order to base our conclusions in Section 4 on a sample that is as representative as possible of the users of the Paris metro. This selection bias is classically treated with the two-steps Heckman selection model (see Appendix B).

When compared with the EGT sample, our sample is on average more manly, younger, less Parisian, poorer and more motorized. Despite this, we find that our sample is close enough to the reference sample. This moderates non-representativeness issue.

One last information critical for this study relates to the level of in-vehicle crowding experienced by PT users. As made clear in the seven last line of Table 8, few individuals have chosen on Figure 17 a situation where they can seat in carriages: only 1 interviewee has chosen the “empty subway” situation and 2.7% of the whole sample has chosen the 1 passenger per square meter situation. By contrast, more than 10% think they will face 6 passengers per square meter during their journeys. Haywood and Koning (2015) show that this proportion is higher in line 1 as compared to line 4, observation consistent with the relative patronage of services. More than 50% of the PT users think they will travel with 2.5-3 passengers per square meter around them. Moreover, we observe only small differences between samples A and B: averages for both populations are similar and amount to 3.2 passengers per square meter.

⁸This seems consistent with the existence of scheduling costs that are more important in the morning (Small and Verhoef, 2007) and that may occur if individuals would answer to the longer survey (because deviating from their preferred arrival time).

Table 8: Individual and journey characteristics for the whole sample, the two sub-samples and the Enquête Globale Transport (EGT)

Variable		Sample	Sub-s. A	Sub-s. B	EGT
Size of sample		999	721	278	2,414
Gender	Woman	51.5%	51%	52.6%	55.1%
	Man	48.5%	49%	47.4%	44.9%
Age (Years)	<20	6.3%	5.4%	8.82%	2.8% (5-14)
	20-30	34%	33.3%	36%	17.1% (15-24)
	30-40	25.9%	27.8%	21%	21.6% (25-34)
	40-50	19.8%	20.2%	18.7%	38.7% (35-54)
	50-60	10.7%	10.6%	11%	11.5% (55-64)
	>60	3.2%	2.7%	4.4%	6.3% (64-74) 2.9% (>74)
Age (continuous)	Years	35.8 (sd: 12.4)	35.9 (sd: 12)	35.5 (sd: 13.4)	
Car available	Yes	37.4%	37.8%	36.4%	33%
	No	62.6%	62.2%	63.6%	66.5%
Income (category) (euros)	<800	15.6%	14.9%	17.6%	4.4% (<800)
	800-1,500	11.7%	10.3%	15.4%	6.1% (800-1,200)
	1,500-1,800	11.8%	12.8%	9.2%	10.1% (1,200-1,600)
	1,800-2,100	14.7%	16%	11.4%	9.5% (2,000-2,400)
	2,100-2,500	15.1%	15%	15.4%	10.9% (2,400-3,000)
	2,500-3,000	13.1%	12.8%	14%	8.2% (3,000-3,500)
	3,000-4,000	10.4%	10.3%	10.7%	11.7% (3,500-4,500)
	4,000-10,000	5.6%	5.9%	4.8%	7.6% (4,500-5,500)
>10,000	1.9%	2.1%	1.5%	15.7% (>5,500)	
Income (continuous)	(euros)	2,422 (sd:2,293)	2,474 (sd: 2,352)	2,282 (sd: 2,126)	
Residence	in Paris	52.7%	48.3%	44.5%	61.6%
	outside of Paris	47.3%	51.7%	55.5%	38.4%
Peak	Morning	50%	44.8%	64%	
	Evening	50%	55.2%	36%	
Motive	Work	70.2%	71.8%	66%	56%
	Other	29.8%	28.2%	34%	44%
Line	Line 1	50.1%	48.3%	55.1%	
	Line 4	49.9%	51.7%	44.9%	
Travel time (minutes)	Total	48.1 (sd: 36.7)	48.6 (sd: 37.2)	46.9 (sd: 35.4)	41.5
	In-vehicle surveyed	9.7 (sd: 6.5)	9.7 (sd: 6.6)	9.6 (sd: 6.25)	
Daily	Yes	63.3%	61.9%	66.9%	
	No	36.7%	38.1%	33.1%	
IVD (pass/m ²)	0	0.1%	0.1%	0%	
	1	2.8%	3.4%	1.1%	
	2	16.2%	17.3%	13.2%	
	2.5	26.4%	25.6%	28.7%	
	3	24.2%	23.9%	25%	
	4	20%	19.8%	20.6%	
	6	10.2%	9.8%	11.4%	

Notes. The continuous age variable is derived from the categorical age variable by allocating to a user a value equal to the center of the age category to which he belongs. For example, users in the 30-40 category are assumed to be 35 years old. Users in the youngest and oldest categories are respectively assumed to be 18 and 70 years old. The continuous income variable is derived from the categorical income variable by allocating to a user a value equal to the center of the income category to which he belongs. For example, users in the 1,800-2,100 category are assumed to have a net monthly income of 1,950 euros. Users in the poorest and wealthiest categories are respectively assumed to have a net monthly income of 4,000 and 15,000 euros.

3 The crowding effect

3.1 Estimation strategy

The crowding effect has already been analyzed through contingent valuation experiments (Wardman and Whelan, 2011; Haywood and Koning, 2015; Kroes et al., 2014), theoretical approach (de Palma, Lindsey, and Monchambert, 2014; de Palma, Kilani, and Proost, 2015) or stated preferences scenarios (Cantwell et al., 2009). Here, we assess it with self-reported comfort satisfaction scores.

In line with the simple theoretical framework presented in Figure 16, the comfort satisfaction (CS) is assumed to depend on the individual preferences, on the travel characteristics and on IVD . We do not observe the exact comfort satisfaction because our dependent variable, CS , is measured on an 11-point discrete scale. As a consequence, the satisfaction is estimated by means of ordered probit model.⁹ In order to use answers to stated satisfaction questions, interpersonal comparability at an ordinal level has to be assumed: a user with a CS of 6 is strictly more satisfied than one with a CS of 5.¹⁰

We first describe the latent variable, CS^* , as:

$$CS^* = \alpha IVD + \sum_{k \in K} \gamma_k x_k + \varepsilon \quad (90)$$

where IVD indicates the in-vehicle density, x is a set of K control variables: line where the user is surveyed (dummy), total journey travel time (hours), in-vehicle travel time in the surveyed line (hours), morning or evening peak (dummy), daily usage of the line (dummy), gender (dummy), car availability (dummy), individual net monthly income (euros, expressed in a logarithmic form), age (centuries) and residence in Paris (dummy). ε captures the unobservables.

In this specification, α measures the crowding effect. x controls for the journey characteristics and for the users preferences which do not depend on IVD but still influence the comfort satisfaction. In Figure 16, these effects are represented by the two arrows

⁹The choice of a logit or a probit model is often at the author's discretion. For example, Van Praag et al. (2003) use a probit model because "it is more flexible than the ordered logit".

¹⁰Cardinality of CS is not assumed here. *A priori*, the difference between a CS of 10 and a CS of 8 may be different from the difference between a CS of 6 and a CS of 4.

respectively from “Journey Characteristics” and “Individual Characteristics” to “Comfort Satisfaction”. This specification assumes that the crowding effect is only influenced by *IVD*, i.e. for a given level of in-vehicle density, all users perceived the same crowding inconvenience, regardless of individual or journey characteristics. In order to take into account the subjective perception of crowding into the *CS*, equations including interaction terms are also estimated. Interaction terms reveal whether or not the effect of variable changes in different contexts. In this case, we want to test if the crowding effect changes when individual or travel characteristics vary:

$$CS^* = IVD(\delta + \lambda_i x_i) + \sum_{k \in K} \kappa_k x_k + \varepsilon, \forall i \in K. \quad (91)$$

Eq. (91) differs from Eq. (90) due to the inclusion of interaction terms between *IVD* and another explanatory variable x_i . The interpretation of these specifications is as follows: if the coefficient of the interaction term (λ_i) is statistically significant, this users’ or journey characteristic influences *CS* via the crowding effect. If the coefficient of the non-interacted variable is significant, this variable drives the level of requirement of users regarding comfort in PT, independently of *IVD*.

3.2 Descriptive statistics

As explained in previous Section, the whole surveyed travelers (N=999) had to select on the show-card (Figure 17) the *IVD* they expected to face once the interview finished. Then, they had to give a score, ranging from 0 to 10 and describing the *CS* associated with such travel conditions.

According to Table 9, *CS* seems positively related to the level of passenger density experienced by PT users: the higher the passenger density, the lower the self-reported *CS*. By contrast, we do not observe any clear relationship between the *CS* score and the in-vehicle travel duration. This is not surprising, the effect of travel time on users’ satisfaction is unclear. In a study on Germany, Stutzer and Frey (2008) show that people with longer commuting time report systematically lower SWB. However, a growing body of research has stressed that traveling time could be positively valued by individuals and

Table 9: Distribution of CS

CS	Distrib. (%)	Ave. IVD	Ave. IVTT
0	6.3	5.1	9.8
1	4.3	4.3	9.7
2	9.5	4.0	10.2
3	11.4	3.4	10.1
4	12.6	3.0	10.1
5	21.9	2.8	9.8
6	17.5	2.7	9.3
7	10.4	2.6	8.8
8	4.3	2.5	9.3
9	1.2	2.0	11.1
10	0.6	2.6	7.9

Notes. This table reports descriptive statistics for sub-samples clustered by *CS*. Column 2 reports the part of each sub-sample into the whole sample. Columns 3 and 4 respectively report the average in-vehicle density, in users per square meter, and the average in-vehicle travel time, in minutes, in each sub-sample.

do not necessarily represent a cost (Huang et al., 2005; Jain and Lyons, 2008). Besides, we see that the distribution’s queue for “worst” *CS* scores is larger than the one for “good” experiences: the three worst marks (0-1-2) represent nearly 20% of answers whereas the three better ones (8-9-10) account only for 6%.

3.3 Estimates results

Eqs. (90) and (91) are estimated on the whole sample with the use of ordered probit. Table 10 reports estimations of Eq. (90) and Eq. (91) when x_i in Eq. (91) is the individual (net) monthly income expressed in a logarithmic form. All results from specifications allowed by Eq. (91) are reported in Appendix C. The chi statistics show that all models are statistically significant, as compared to the null models without any predictor. The pseudo- R^2 are small but they are consistent with the belief that only a share of the SWB measures depends on objective variables. This share seems to vary between 8 and 20% (Kahneman et al., 1999).

The estimate of Eq. (90) confirms that *IVD* influences negatively the *CS*, due to the sign and the significance of the *IVD* coefficient. Moreover, the estimated cut-offs for the latent variable CS^* are distributed in a remarkably regular order. Thus the mean

Table 10: Effects of the density and the income on the *CS*

	(1)		(2)	
	CS*		CS*	
	Coef.	Std. err.	Coef.	Std. err.
Crowding effects:				
IVD (users/m ²)	-0.550 ***	0.044	0.699 **	0.334
IVD × ln(Individual net monthly income)			-0.166 ***	0.043
Journey controls:				
Line (1=line 1/0=line 4)	0.074	0.077	0.085	0.078
Door to door travel time (hours)	0.070	0.052	0.065	0.051
In-vehicle travel time (hours)	-0.086	0.310	-0.085	0.315
Peak hour (1=morning/0=evening)	0.179 ***	0.067	0.184 ***	0.067
Daily usage of the line (1=Y/0=N)	-0.125*	0.070	-0.127*	0.070
Individual controls:				
Gender (1=male/0=female)	0.126*	0.066	0.117*	0.066
Car available (1=Y/0=N)	-0.060	0.071	-0.085	0.071
ln(Individual net monthly income (euros))	-0.094*	0.049	0.406 ***	0.132
Age (centuries)	0.321	0.316	0.238	0.320
cut1	-4.139	0.335	-0.459	0.979
cut2	-3.754	0.329	-0.065	0.978
cut3	-3.189	0.320	0.511	0.976
cut4	-2.721	0.314	0.985	0.974
cut5	-2.306	0.312	1.405	0.974
cut6	-1.655	0.310	2.060	0.973
cut7	-1.036	0.307	2.680	0.973
cut8	-0.414	0.304	3.303	0.971
cut9	0.181	0.309	3.902	0.973
cut10	0.598	0.324	4.323	0.969
Number of observations	999		999	
Likelihood function	-1953.041		-1943.405	
Pseudo R ²	0.086		0.090	
Prob > chi2	0.000		0.000	
Akaike IC	3.950		3.933	
Number of iterations	4		4	

Notes. This table reports results from ordered probit estimations of Eq. (90) and Eq. (91) when x_i in eq. (90) is ln(Individual net monthly income (euros)). *significant at 10%; **significant at 5%; ***significant at 1%.

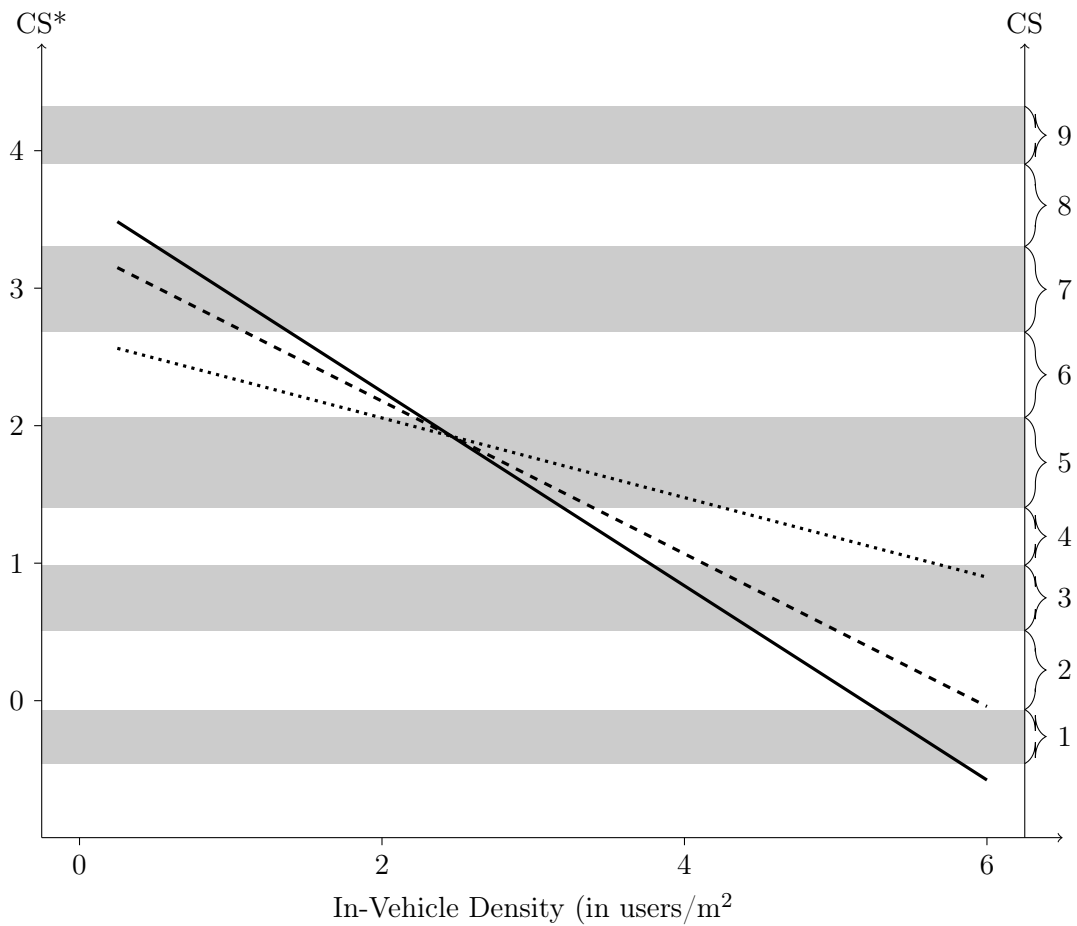
distance between two successive cuts is 0.524 and the variance of these distances is 0.011. Considering that one additional user per square meter decreases the predicted latent CS^* by 0.55, this means that, on average, one additional user per square meter decreases by one the predicted CS measure.

Analyzing all interaction terms in Appendix C (Table 15), we find that only the individual income variable influences the dissatisfaction assigned to IVD . The Akaike Information Criterion (AIC) confirms that Eq. (91) estimated with the the individual (net) monthly income expressed in a logarithmic form produces the best goodness of fit.¹¹ Interacted income coefficient equals -0.166: the wealthier the users, the more the density decreases their CS . Interestingly, the non-interacted income coefficient is significant and negative in the model without interaction (column (1)) whereas it is positive and still significant in the model with interaction (column (2)). This has to be interpreted with the interaction terms between \ln (Individual net monthly income (euros)) and IVD . Indeed, the first model does not distinguish the constant effect of the income on the CS from the influence of the income on the crowding effect, whereas the second does. To illustrate this result, the expected CS^* and the expected CS for the representative user with various level of income (400, 2 000 and 5 000 euros net monthly income) are drawn as a function of IVD in Figure 18.¹² Other things being equal, wealthier users have a lower CS^* when vehicles are very crowded (6 users per square meter). Nevertheless, their CS^* increases more quickly when IVD increases.

Turning to travel variables, we find that the in-vehicle travel time coefficient is not significant. As a consequence, the CS does not seem to be driven by the amount of time spent into the vehicles. This is not in line with much of the economic literature which estimates multipliers of the value of in-vehicle travel time as measure of the crowding cost (Wardman and Whelan, 2011; Li and Hensher, 2011; Haywood and Koning, 2015; Wardman and Murphy, 2015). However, other studies on the Paris Region by Kroes et al. (2014) and de Lapparent and Koning (2015) display the same result: a fixed crowding penalty better fits the data than a time multiplier specification. A possible justification

¹¹The Akaike Information Criterion measures the quality of a model relative to the number of variables used.

¹²Hereafter, the representative user is someone who shares all characteristics with sample average.



- Expected CS* of the representative user with a net monthly income of 400 euros.
- - - - Expected CS* of the representative user with a net monthly income of 2 000 euros.
- Expected CS* of the representative user with a net monthly income of 5 000 euros.

Figure 18: Representative user expected CS* and expected CS as a function of the in-vehicle density and of the income.

of this result is that the in-vehicle travel time is short for most of our sample - about 10 minutes in average -, which does not allow the time effect to become visible.

Also, we observe in Table 10 that traveling in the morning brings more comfort satisfaction than traveling in the evening. One potential explanation is that the destination is often home when traveling during the evening peak, and that users are in general more impatient to arrive at home than at other destinations. Moreover, we can assume that individuals are more tired after a full workday and, consequently, suffer more from crowded PT travels in the evening.

Daily users seem to have different *CS* as compared to temporary users. There is no "habituation"-effect playing here: daily users are not used to PT travel, they tend to be less satisfied by their journey comfort than temporary users. This is in line with Baum and Greenberg (1975) who found that expectations do not reduce people's perception of general level of discomfort.

Individual characteristics are the last set of explanatory variables. Other things being equal, men have a lower level of expectations than women. This effect is present in models (1) and (2) in Table 10. We can hypothesize that men suffer less than women from the various nuisances linked to PT crowding, such as lack of safety. Later issue became recently important in France. French Ministry of Social Affairs, Health and Women's Rights, Marisol Touraine, thus commended a survey (Bousquet et al., 2015) showing that all interviewed women reported sexual harassment while riding the subway. Moreover, six out of ten women fear an aggression or a theft in the Ile-de-France PT, against three out of ten men (Bon et al., 2011). Clearly, such effects may be magnified by one high *IVD* and thus lead to lower stated *CS* for women, as questioned in the next Section.

4 Anatomy of the crowding effect

After having identified the crowding effect through the stated satisfaction of PT users, we now address its anatomy. We first discuss the eight nuisance factors retained as channels of the crowding effect, i.e. the features of the journey that are deteriorated by a high in-vehicle density. Then we propose an empirical analysis showing how do these channels

impact comfort satisfaction.

4.1 The nuisance factors

Literature Review

It is largely accepted that crowding has, in general, negative impacts on individuals. Langrehr (1991) showed that “In crowded conditions, people performed complex tasks more poorly and became more frustrated”. In order to assess the perception of crowding across users, however, several dimensions of PT crowding should be distinguished. We have identified eight aspects, called nuisance factors, which may be affected by a high *IVD*. They may be gathered into five categories: psychological, physical, sensory, temporal and risky. One major constraint is that users have to be able to mark these dimensions. Potential abstract dimensions such as the “lack of control” are hardly quantifiable for users, despite it is a consequence of high density which is largely admitted by psychologists (see Cox et al., 2006, or Mohd Mahudin et al., 2012).¹³

The psychological category is made of one dimension, the *Over closeness*. When *IVD* is high, users are closer to each other. Here, the over-closeness is a synonym of stress and lack of control, which are caused by crowding in general (Epstein, 1981) and in PT services in particular (Epstein et al., 1981; see also Vine, 1982, for a critical review of the link between density and stress). For women, the effect of over-closeness may be magnified by the risk of sexual harassment.

The physical category is made of the *Standing* position. When the density is high in a train, a part of the users has to keep standing. This leads to physical tiredness. Bousenna et al. (1982) show that long standing position causes pain and discomfort. As a matter of fact, time multipliers used to value crowding costs are often distinguished between seating and standing situations, the later being clearly higher (see Wardman and Whelan, 2011).

The sensory dimensions are the *Noise* and the *Smell*. A high level of noise may cause discomfort and mental health problems. This effect has been identified by Bhattacharya et al. (1995) in the Calcutta subway. They find that the crowd increases the level of noise.

¹³The interested reader is referred to Mohd Mahudin et al. (2012) for a description and an analysis of psychological aspects of crowding.

Bad smells may also be a consequence of a high density of users in train, because when *IVD* increases, the temperature also does. This effect is accentuated if the ventilation is not adequate (Gershon et al., 2005).

We describe temporal dimensions as the *Time Loss* and the *Waste of Time*. A high *IVD* causes increased dwell times at stations due to boarding and alighting. At a macro level, a high patronage in PT increases the vulnerability of the networks because of potential incidents. This makes the travel time very unreliable. These two aspects show that high *IVD* is positively correlated with longer travel time, which decreases the appeal of transport (McFadden, 1974; Noland et al., 1998). The wasted time dimension represents the fact that when the density is high, users are not able to perform tasks they would like during their PT journeys (Langrehr, 1991), such as read a newspaper or a work.

The risky dimensions of crowding are characterized by the risk of *Fall* and the risk of *Robbery*. Cartledge (2003) shows that if standing users feel vulnerable to the risk of fall, seated users “will find themselves providing a cushion for those standing, with a concomitant increase in the risk of head to head clashes”. The risk of robbery is also more important in crowded contexts, as shown by Uzzell and Brown (2007).

Descriptive statistics

Table 11 presents statistics for the users’ dissatisfaction scores (NFD_d) related to every d nuisance factor of in-vehicle crowding. Recall that only 27% of the whole sample took part to the qualitative survey. Moreover, travelers had to respond with respect to the worst comfort situation described on Figure 17, i.e. 6 passengers per square meter. To describe the anatomy of the crowding effect, NFD_d has to be interpreted with respect to the stated dissatisfaction of other nuisance factors: measuring the bad smells as a 1 does not have the same meaning if all the other dimensions are measured as a 1 or if they are measured as a 10. Therefore we analyze simultaneously the various NFD_d and their ranks.¹⁴ For any dimension, we display the first quartile, the median, the third quartile, the mean and the standard deviation of NFD_d and of its rank.

¹⁴The rank was obtained by ordering all the dissatisfaction measures for one user. If the two highest dissatisfaction measures are equal, their rank is 1 and the rank of the following nuisance factor is 3.

Table 11: Rank and score statistics for the 8 underlying nuisance factors of crowding

Category	Dimension	Rank					NFD				
		p25	p50	p75	mean	sd	p25	p50	p75	mean	sd
Psychological	Over-closeness	1	1	3	2.0	1.537	6	8	10	7.7	2.525
Physical	Standing	1	3	5	3.3	2.316	5	7	9	6.3	3.208
Sensory	Noise	3	4	6	4.2	2.106	3	5	8	5.2	2.924
	Smell	1	3	5	3.1	1.906	5	7	9	6.6	2.827
Temporal	Time loss	2	5	6	4.3	2.262	3	5	7	5.1	2.874
	Wasted time	2	4	6	3.9	2.294	3	5	8	5.5	3.116
Risky	Fall	4	6	7	5.2	2.297	0	4	6	3.9	3.286
	Robbery	1.5	4	5	3.8	2.119	3	5	8	5.5	3.198

Notes. This table reports descriptive statistics for each of the nuisance factors of crowding defined in sub-subsection 4.1. Column 1 (**category**) reports the category of the nuisance factor. Columns 3, 4 and 5 respectively report the 25th percentile, the median and the 75th percentile of the rank of the interest nuisance factor dissatisfaction measure relative to dissatisfaction of other nuisance factors. Column 6 reports the mean value of the rank. Column 7 reports the standard deviation of the rank. Columns 7 to 11 reports the same statistics than columns 3 to 7 for the self-reported nuisance factor dissatisfaction.

The *Over-closeness* experienced during a crowded journey is the most dissatisfying nuisance factor. More than the half of the users rank this feature as the most unpleasant, and the mean value of NFD_d is the highest (7.7). This is not surprising since on the one hand, the over-closeness in PT implies an intrusion in the user' individual space. On the other hand, it may strengthen the unpleasant effect of other dimensions such as bad smells, noise or robbery. Said differently, the over-closeness is inherent to the in-vehicle crowding.

Then come *Smell* and *Standing*. Their median rank is third and their median NFD_d are 6.6 and 6.3 respectively. The dissatisfaction due to the bad smells is high when the train is crowded because users are very close to each other. Moreover, a high *IVD* may increase the temperature in the train and may also deteriorate the journey conditions. The standing position is directly tied to the crowding: the more users in the train, the lower the probability to find a sit.

The *Robbery*, the *Wasted Time*, the *Noise* and the *Time Loss* factors are moderately rated. The average ranks of these four nuisance factors are about 4, the averages NFD_d are about 5.5-5.1 and their medians are 5. Crowded PT journeys are seen as particularly vulnerable to thefts. Users are close to each other, physical contacts are frequent and facilitate theft. Note that the wasted time is somewhat put forward by users because they

are not able to do what they would like (read a book or a newspaper, work...), whereas according to Ettema et al. (2012) activities performed during travel influence positively the PT users' SWB. The noise is a characteristic of the crowd and causes tiredness and stress to users. Finally, crowding may induce time loss because loading and unloading times are longer. In latter case, the objective travel time is impacted by an excessive PT patronage (as for road congestion). But this phenomenon does not seem problematic for Paris subway users.¹⁵

Lastly, risks of *Fall* due to high density are viewed as negligible by subway users. Such a nuisance factor has clearly the lowest average NFD_d (3.9) and average rank (5.2). This is probably linked to the facts that older people are more sensible to this dimension and that the average age of our sample is low (in Sub-sample B, only 4.4% of users are older than 60). Moreover, the proximity of other travelers within carriages may avoid someone falling directly on the ground. Instead, a PT user will fall on someone's body when trains driving style is not smooth.

4.2 Effects of the nuisance factors on the comfort satisfaction

Estimation strategy

First, we want to test if the set of nuisance factors taken as a whole forms a scale that informs about the comfort satisfaction. A user reporting higher dissatisfaction scores is assumed to have a lower CS , through two possible channels: higher NFD_d may indicate a higher level of requirement towards PT, independently of IVD , or it may indicate a lower tolerance to crowding, through the various nuisances influenced by crowding. The level of requirement may here be considered as an individual fixed effect.

To test these hypothesis, two extra assumptions are required. The NFD_d are assumed to be cardinal measures of the dissatisfaction. It means that the nuisance factors are of equal importance and that the difference in dissatisfaction between a NFD_d of 10 and a NFD_d of 8 equals the difference between a NFD_h of 6 and a NFD_h of 4. We also need to assume that the differences in NFD_d across users are stable whatever the in-vehicle

¹⁵Despite a growing patronage in Paris PT networks over the last decade, regularity indicators (describing the share of travelers having to wait on platforms more than 3 minutes during peaks) were hold constant.

density, IVD .

CS^* is now modeled as:

$$CS^* = IVD (\alpha_1 + \alpha_2 \overline{NFD}) + \beta \overline{NFD} + \sum_{k \in K} \gamma_k x_k + \theta \hat{\lambda} + \varepsilon. \quad (92)$$

where IVD indicates the in-vehicle density, $\overline{NFD} = \sum_d NFD_d$ is the sum of the nuisance factor scores given by the asked individual, x is a set of K control variables.¹⁶ The inverse Mills ratio $\hat{\lambda}$ controls for the selection bias described in Appendix B.

According to this specification, the coefficients of interest are α_2 and β . If α_2 is negative, the scale formed by the sum of NFD_d measures the intensity of the crowding effect. If β is negative, users who report to be more dissatisfied by the various nuisance factors are less satisfied by the comfort in PT, independently of IVD .

Second, we test nuisance factors as channel through which IVD may decrease CS^* :

$$CS^* = IVD \left(\sum_d \alpha_d NFD_d \right) + \beta \overline{NFD} + \sum_k \gamma_k x_k + \theta \hat{\lambda} + \varepsilon. \quad (93)$$

The coefficients of interest are the α_d . One negative α_d parameter means that a user who is more dissatisfied by the nuisance factor d is more dissatisfied by IVD . Moreover, if α_i is lower than α_d , the contribution of the nuisance factor i to the whole crowding effect is higher than the contribution of the nuisance factor d .

Results

Eqs. (92) and (93) are estimated on sub-sample B (278 individuals), still with the use of ordered probits. Table 12 reports results¹⁷. Since the main individual and journey effects have been discussed in the previous section, the discussion hereafter focuses on the crowding effect coefficients. Once again, the pseudo-R2 for all three regressions are rather small and fall between 0.096 to 0.109.

¹⁶These control variables are : line where the user is surveyed (dummy), total surveyed journey travel time (hours), immediate journey travel time (hours), morning or evening peak (dummy), daily usage of the line (dummy), gender (dummy), car availability (dummy), ln (Individual net monthly income (euros)), age (centuries) and residence in Paris (dummy)

¹⁷Complete regression results are displayed in Appendix D, see Tables 16, 17 and 18.

Table 12: Effects of the crowding on the CS

	(1)		(2)		(3)	
	CS*		CS*		CS*	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Crowding effect:						
IVD (users/m ²)	-0.495 ***	0.093	-0.401*	0.222		
IVD X \overline{NFD} (0-80)			-0.002	0.005		
IVD X Standing NF					-0.019 **	0.008
IVD X Over-closeness NF					-0.015	0.010
IVD X Noise NF					-0.015*	0.008
IVD X Robbery NF					0.001	0.009
IVD X Fall NF					0.007	0.009
IVD X Smell NF					-0.007	0.009
IVD X Time Loss NF					-0.013	0.009
IVD X Wasted Time NF					-0.020 **	0.008
\overline{NFD}	-0.015 ***	0.004	-0.008	0.014	0.015 **	0.007
Journey controls:	Y		Y		Y	
Individual controls:	Y		Y		Y	
Selection control:						
$\hat{\lambda}$	1.480	0.951	1.481	0.949	1.823*	0.966
Number of observations	278		278		278	
Likelihood function	-541.648					
		-541.469		-533.784		
Pseudo R ²	0.096		0.096		0.109	
Prob > chi2	0.000		0.000		0.000	
Akaike IC	4.055		4.061		4.049	
Number of iterations	4		4		4	

Notes. This table reports results from ordered probit estimations of Eq. (92) in columns (1) and (2), and of eq. (93) in column (3). *significant at 10%; **significant at 5%; ***significant at 1%.

First, the nuisance factors scale formed by the sum of the NFD_d is negatively correlated with the CS (coefficient of \overline{NFD} equal to -0.015 in column (1)). This user fixed effect shows that users reporting a higher nuisance factors scale are less satisfied by the comfort in PT, whatever the level of IVD . However, the model displayed in column (2) of Table 12 shows that the nuisance factors scale does not interact with the crowding effect to explain CS . The nuisance factors scale informs of the individual level of requirement, but it does not inform of the individual crowding perception. This may be due to the fact that users do not attach the same importance to the nuisance factors (see Table 11) while the nuisance factors scale assume the nuisance factors are homogeneous.

Column (3) displays results of the estimation of Eq. (93). The significantly positive coefficient for the inverse Mills ratio, $\hat{\lambda}$, in column (3), indicates a simple ordered probit model without including $\hat{\lambda}$ would suffer from selection bias.

The estimated coefficients of the interaction between IVD and the nuisance factors scores are negative and significant for *Standing*, *Noise* and *Wasted Time*. Users who are more dissatisfied by one of these three nuisance factors perceive a higher utility cost of crowding. Therefore these three nuisance factors may be considered as channels of the crowding effect. A higher IVD decreases CS^* because it increases the nuisances due to *Standing*, *Noise* and *Wasted Time*. Due to the value of the coefficients and the standard errors, we have to be cautious when comparing the strengths of the nuisance factors. *Standing* and *Wasted Time* seem to have the highest impact on CS , followed by *Noise*. When the in-vehicle density is high, users incur a disutility because they have to stand, because they are not able to spend their time as they would like and because the general level of noise increases.

4.3 Heterogeneity of users

From a public policy perspective, our results suggest that the comfort satisfaction of Paris subway users may be increased by addressing one (or more) of the four channels identified previously. Obviously, interventions trying to reduce the *Wasted Time* nuisance are hard to define, especially since former factor relates to stress and lack of control caused by overcrowding. Concerning the *Standing* and *Noise* features, potential strategies nevertheless

Table 13: Main effects of socio-economic variables on Nuisance Factors Dissatisfaction scores

	Standing	Wasted time	Noise
Gender (1=male)	(-)	(-)	(-)
Income		(+)	(+)
Age		(-)	
Car available (1=Y)		(-)	
Live in Paris (1=Y)			(-)

The signs displayed in this table are the signs of significant coefficients obtained through ordered Probit estimations of Nuisance Dissatisfaction Scores. These estimations are available in Appendix E.

Reading: (+) means that a policy addressing this nuisance factor would increase more the comfort satisfaction of users with these characteristics. (-) means that a policy addressing this nuisance factor would increase less the comfort satisfaction of users with these characteristics.

do exist. Thus investing in rolling-stocks that propose a better sound insulation may be relevant. Advertisements aimed at making subway users' behaviors more respectful to their co-travelers may also be of interest. Lastly, increasing the service frequency and/or the seat capacity of vehicles could increase the likelihood that the PT users find an available seat. Later policies could also soften the nuisances caused by the *Wasted Time* dimension.

It is worth noting, however, that PT users probably do not value in the same manner these various nuisance factors. As a consequence, addressing one specific nuisance factor may favor one type of users. In order to investigate this issue more in depth, we have estimated ordered probit looking at the determinants of NFD_d for the nuisance factors *Standing*, *Wasted Time* and *Noise* (see Appendix E). Signs of the significant coefficients are displayed in Table 13.

It appears that being a woman has a constant (positive) effect on every NFD_d . Put differently, women are more dissatisfied than men by the three main nuisance factors related to in-vehicle crowding. Then we observe that wealthy individuals are more likely to be affected by *Wasted Time* and *Noise*. Somehow interestingly, car-owner users seem to compare the crowding conditions in PT with the individual car travel conditions. As a consequence, they find the *Wasted Time* less penalizing than other users do, maybe because they know that they can occupy their travel time in a better way than if they had to focus on the road traffic, whatever the level of density. Moreover, old people tend to be less affected by this nuisance factor. Lastly, we find that Parisians are less affected by

Noise.

5 Conclusion

A growing body of research currently focus on the relationships between subjective well-being and transportation. This paper has used an original survey collected on Paris subway platforms to investigate the crowding effect. The empirical analysis has taken the responses to questions about dissatisfaction related to various features of comfort when the in-vehicle density is very high as measures for user perception of crowding, and the responses to a travel satisfaction question as a measure for user subjective well-being.

The main conclusions can be summarized as follows: (1) on average, an extra-user per square meter decreases by one the 0 to 10 scale measure of comfort satisfaction related to PT travel; (2) contrary to previous studies, we do not find any empirical evidence supporting that the travel time accentuates the crowding effect; (3) when trains are full, wealthier users are more dissatisfied than poorer ones, however their SWB increases more quickly when the in-vehicle density decreases; (4) we identify three channel through which the in-vehicle density decreases the comfort satisfaction: an higher probability to stand for all or part of the journey, a poorer use of the time during the journey, and noisier travel conditions; (5) women and wealthy individuals are more likely to benefit from any policy addressing of or more of these four channels.

Obviously, this research requires further studies in order to get a better understanding of PT users' satisfaction. First, we have only considered the crowding effect of individuals well-being. An comprehensive analysis should also look at travel time and monetary determinants of overall transport satisfaction. Moreover, our survey asked only current PT users. In order to identify the factors that would ease modal shift policies from cars towards PT, motorists' preferences cannot be ignored. Despite these limitations, we believe that our study contributes, in an original way, to the empirical assessment of over-crowding in PT systems. Valuation studies based on contingent experiments could add qualitative questions, in order to precise the source of nuisances caused by an excessive patronage in PT.

Appendices

Appendix A Paris subway map

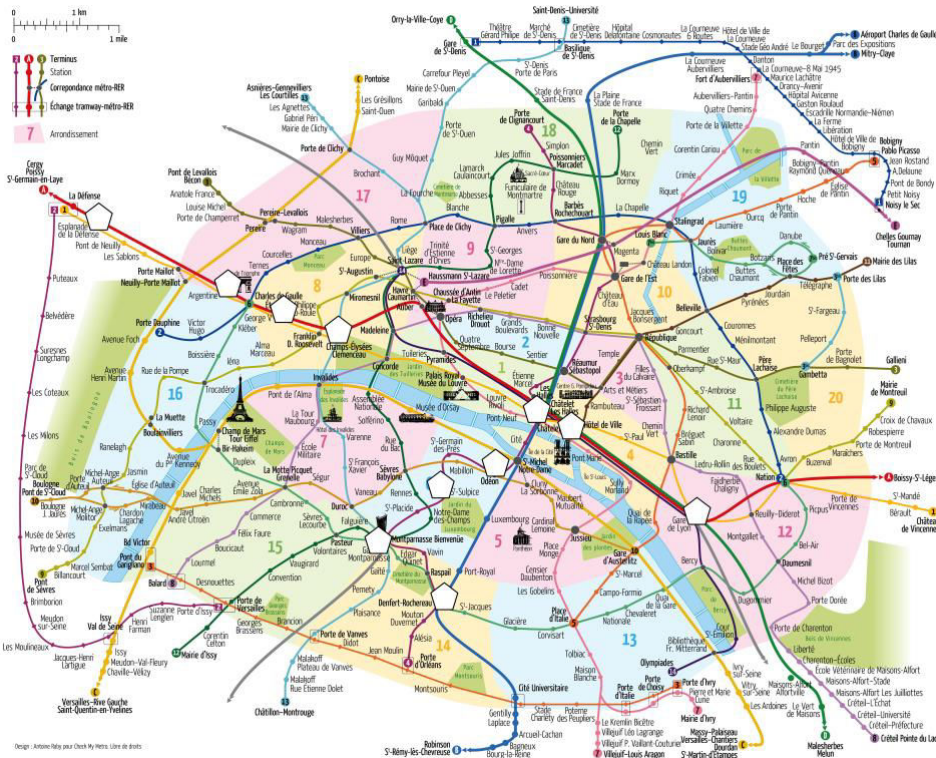


Figure 19: Map of the Parisian underground PT network. Stations in which the survey has been collected are highlighted by white pentagons.

Appendix B Treatment of the selection bias

We use a standard Heckman¹⁸ selection equation to control for the fact that we have information concerning the nuisance factors of crowding only for users who decided to answer the complete survey.

Formally, this selection issue can be expressed as follows:

$$\begin{aligned} NFD_d^* &= \begin{cases} X_d\beta_d + \varepsilon_d & \text{si } y > 0 \\ . & \text{si } y \leq 0 \end{cases}, \text{ for } d \in [1; 8] \\ y &= W\alpha + \gamma \end{aligned} \quad (94a)$$

$$R = \begin{cases} 1 & \text{si } y > 0 \\ 0 & \text{si } y \leq 0 \end{cases} \quad (94b)$$

where NFD_d^* is the latent variable associated with the dissatisfaction mark given to the nuisance factor of crowding d , NFD_d , y is a latent variable associated with the probability to reply to *Questions B* and W is a vector of variables which influence the probability to reply to *Questions B*. γ captures the unobservables.

We only observe R which equals one if the respondent has answered the questions about the nuisance factors of comfort, and 0 otherwise. The estimator of β_d is not biased if $E[\varepsilon_d|\gamma] = 0$. This is, *a priori*, not the case in our study as users who took the time to reply may grant more importance to the inconvenience due to crowding than other users.

We assume that the error terms are bivariate normally distributed, so that:

$$\begin{pmatrix} \varepsilon_d \\ \gamma \end{pmatrix} \rightsquigarrow N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (\sigma_d)^2 & \rho\sigma_d \\ \rho\sigma_d & 1 \end{pmatrix} \right], \text{ for } d \in [1; 8]$$

This assumption is consistent with the use of probit estimations. It implies:

$$E[NFD_d^*|X, y > 0] = X^k\beta_d + E[\varepsilon - d|\gamma > -W\alpha] = X_d\beta_d + \rho\sigma_d \frac{\varphi(W\alpha)}{\Phi(W\alpha)}, \text{ for } d \in [1; 8]$$

where φ is the standard normal density function and Φ is the standard normal cumulative

¹⁸See Heckman (1979).

distribution function.

We obtain the regression equation:

$$NFD_d^* = X_d\beta_d + \rho\sigma_d\lambda + e_d, \text{ for } d \in [1; 8]$$

where e_d verifies:

$$\begin{aligned} E[e_d|y > 0] &= 0; \\ V[e_d|y > 0] &= (\sigma_d)^2(1 - \rho^2\delta); \end{aligned}$$

with $\delta = \lambda(\lambda + W\alpha)$ and $\lambda = \frac{\varphi(W\alpha)}{\Phi(W\alpha)}$.

Therefore, the inverse Mills Ratio is given by:

$$\hat{\lambda} = \frac{\varphi(W\hat{\alpha})}{\Phi(W\hat{\alpha})},$$

$\hat{\alpha}$ being estimated with a probit model. The inverse Mills Ratio $\hat{\lambda}$ is used as an instrument that incorporates the characteristics of users who did not reply to the whole survey.

To estimate correctly $\hat{\lambda}$, we need instruments which are not correlated with the mark but correlated with the probability of answering the whole survey. We rely on three instruments here: the time the user answered the survey (during the morning peak or during the evening peak), a dummy equal to 1 if the user answers a question related to his "willingness to pay" during the first part of the survey, and equal to 0 if he answers a question related to his "willingness to accept", and a dummy equal to one if the gender of the interviewed user is different from the one of the interviewer.

Common sense suggests that the two instruments should be relevant. Indeed, no one of these variables should affect the marks a user gives to the different features of crowding, especially since that mark was awarded with respect to the 6 passengers per square meter situation.

A user may have more time to spend in the evening because he can more easily deviate from his desired arrival time. However, it is not supposed to influence the drivers of the crowding perception. Indeed, on lines 1 and 4, "bunching" phenomenon is not observed:

an overcrowded train does not imply that the following one will be less crowded. No user wait for the following train, except if the train is completely full.

The second instrument, willingness to pay, was chosen because some surveyed travelers were randomly picked to answer a "willingness to accept" question instead of a "willingness to pay" during the first part of the survey. When asked on their "willingness to accept", some of these users selected the 6 passengers per square meter situation as reference point. Because it was impossible, in that case, to propose them an hypothetical worsening of comfort, the interviewer skipped this question. Consequently, these users had more time to complete the qualitative survey. This particular selection effect is due to the random assignment of the survey design.

The interviewer gender effect during surveys is known in the literature (see Kane and Macaulay, 1993; Catania et al., 1996; Huddy et al., 1997). This effect may influence the decision to take the survey, and, in some specific cases such as sexual behavior (Catania et al., 1996), gender inequality (Kane and Macaulay, 1993) or feminists and political activism (Huddy et al., 1997), it may also influence the answers. In our survey, questions are not related to such gender controversy. There is nor reason for answers to be influenced by the interviewer gender. However, the gender may influence the decision to take Question B of the survey.

The binary probit model describing the likelihood of answering the qualitative survey is globally significant. Results are displayed in Table 14. As expected, users were more prone to answer the entire survey when they were interviewed during the evening peak rather than during the morning peak. This is consistent with descriptive statistics in Table 8 and with the existence of larger scheduling costs at that time of the day. The other instruments (*Reply to willingness to pay question* and *Gender match*) are not statistically significant. By contrast, we observe that the sub-sample accounts less line 1 users than it should, perhaps because they spend more time in vehicles (see Haywood and Koning, 2015). Daily users are also more disposed to take Question B.

Table 14: Probit model for the selection (Output=1 if the user answers the whole survey)

	y	
	Coef.	Std. err.
Instruments:		
Peak hour (1=morning/0=evening)	-0.465 * **	0.088
Reply to willingness to pay question (1=Y/0=N)	0.118	0.106
Gender match (1=Y/0=N)	0.134	0.087
Journey effects:		
Line (1=line 1/0=line 4)	-0.169*	0.089
Door to door travel time (hours)	0.037	0.074
Daily usage of the line (1=Y/0=N)	0.152*	0.092
Individual effects:		
Gender (1=male/0=female)	-0.022	0.087
ln Individual net monthly income (euros)	-0.056	0.064
Age (Centuries)	0.169	0.404
Constant	-0.206	0.431
Number of observations	999	
Likelihood function	-569.752	
Pseudo R ²	0.036	
Prob > chi2	0.000	
Number of iterations	3	

Notes. This table reports coefficients from a probit estimation of y , the latent variable associated with the dummy equal to 1 if the interviewee replied to *Questions B, R*, defined in eq. (94b). This estimation is made on the vector of variables W defined in eq. (94a). *significant at 10%; **significant at 5%; ***significant at 1%.

Appendix C *CS* estimatesTable 15: Effects of the crowding on the *CS*

	<i>CS</i> *			
	Coef.	Std. err.	Pseudo R ²	<i>AIC</i>
(1) x_i = Line (1=line 1/0=line 4)				
IVD	-0.465***	0.077		
IVD $\times x_i$	-0.130	0.082		
x_i	0.464*	0.238	0.087	3.948
(2) x_i = Door to door travel time (hours)				
IVD	-0.531***	0.076		
IVD $\times x_i$	-0.023	0.069		
x_i	0.144	0.213	0.086	3.952
(3) x_i = In-vehicule travel time (hours)				
IVD	-0.533***	0.069		
IVD $\times x_i$	-0.104	0.277		
x_i	0.259	0.941	0.086	3.952
(4) x_i = Peak hour (1=morning/0=evening)				
IVD	-0.512***	0.058		
IVD $\times x_i$	-0.079	0.066		
x_i	0.426**	0.202	0.086	3.950
(5) x_i = Daily usage of the line (1=Y/0=N)				
IVD	-0.517***	0.062		
IVD $\times x_i$	-0.052	0.070		
x_i	0.037	0.211	0.086	3.951
(6) x_i = Gender (1=male/0=female)				
IVD	-0.555***	0.057		
IVD $\times x_i$	0.013	0.064		
x_i	0.085	0.198	0.086	3.952
(7) x_i = Car available (1=Y/0=N)				
IVD	-0.568***	0.056		
IVD $\times x_i$	0.040	0.067		
x_i	-0.187	0.213	0.086	3.952
(8) x_i = ln(Individual net monthly income (euros))				
IVD	0.699**	0.334		
IVD $\times x_i$	-0.166***	0.043		
x_i	0.406***	0.132	0.090	3.933
(9) x_i = Age (centuries)				
IVD	-0.521***	0.063		
IVD $\times x_i$	-0.043	0.054		
x_i	0.408	0.707	0.087	3.950

Notes. This table reports results from 9 ordered probit estimations of *CS*^{*}, the latent variables associated with *CS*, on x_i , the variable of interest defined in column 1, *IVD*, the in-vehicule density measured in users/m², and the interaction term between x_i and *IVD*, and x , the set of k explanatory variables defined in subsection 3.1. All estimations are based on 999 observations. *significant at 10%; **significant at 5%; ***significant at 1%.

Appendix D Including *NFD* to explain the comfort satisfaction

Table 16: *CS*, first specification

	(1)		(2)	
	Coef.	Std. err.	Coef.	Std. err.
Crowding effect:				
IVD (users/m ²)	-0.491 ***	0.094	-0.495 ***	0.093
Journey controls:				
Line (1=line 1/0=line 4)	-0.066	0.168	-0.248	0.206
Door to door travel time (hours)	-0.043	0.127	-0.024	0.129
In-vehicule travel time (hours)	-0.241	0.561	-0.275	0.567
Peak hour (1=morning/0=evening)	0.249*	0.133	-0.266	0.368
Daily usage of the line (1=Y/0=N)	-0.037	0.129	0.120	0.166
Individual controls:				
Gender (1=male/0=female)	0.101	0.134	0.073	0.133
Car available (1=Y/0=N)	-0.245*	0.127	-0.258 **	0.129
ln(Individual net monthly income (euros))	-0.094	0.100	-0.147	0.100
Age (years)	-0.411	0.624	-0.271	0.627
$\sum NFD$	-0.014 ***	0.004	-0.015 ***	0.004
Selection control:				
$\hat{\lambda}$			1.480	0.951
cut1	-5.093 ***	0.737	-3.927 ***	1.156
cut2	-4.666 ***	0.728	-3.502 ***	1.154
cut3	-3.984 ***	0.698	-2.817 **	1.135
cut4	-3.558 ***	0.684	-2.388 **	1.125
cut5	-3.099 ***	0.679	-1.927*	1.123
cut6	-2.475 ***	0.674	-1.301	1.119
cut7	-1.819 ***	0.666	-0.641	1.110
cut8	-1.273 **	0.646	-0.088	1.095
cut9	-0.786	0.644	0.403	1.088
cut10	-0.503	0.648	0.689	1.101
Number of observations	278		278	
Likelihood function	-542.784		-541.648	
Pseudo R ²	0.094		0.096	
Prob > chi2	0.000		0.000	
Akaike IC	4.056		4.055	
Number of iterations	4		4	

Notes. This table reports results from ordered probit estimations of Eq. (92) with $\alpha_2 = 0$. Regression in column (2) includes the inverse Mills ratio to control for selection. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 17: *CS*, second specification

	(1) CS*		(2) CS*	
	Coef.	Std. err.	Coef.	Std. err.
Crowding effect:				
IVD (users/m ²)	-0.398*	0.214	-0.401*	0.222
IVD (users/m ²) X $\sum NFD$	-0.002	0.004	-0.002	0.005
Journey controls:				
Line (1=line 1/0=line 4)	-0.061	0.168	-0.244	0.206
Door to door travel time (hours)	-0.040	0.127	-0.022	0.129
In-vehicule travel time (hours)	-0.224	0.561	-0.258	0.567
Peak hour (1=morning/0=evening)	0.253*	0.135	-0.262	0.368
Daily usage of the line (1=Y/0=N)	-0.035	0.131	0.122	0.168
Individual controls:				
Gender (1=male/0=female)	0.095	0.134	0.067	0.134
Car available (1=Y/0=N)	-0.253 **	0.124	-0.267 **	0.126
ln(Individual net monthly income (euros))	-0.091	0.101	-0.144	0.100
Age (years)	-0.412	0.624	-0.271	0.626
$\sum NFD$	-0.008	0.014	-0.008	0.014
Selection control:				
$\hat{\lambda}$			1.481	0.949
cut1	-4.772 ***	0.986	-3.603 ***	1.345
cut2	-4.344 ***	0.981	-3.177 **	1.345
cut3	-3.661 ***	0.965	-2.492*	1.332
cut4	-3.233 ***	0.956	-2.061	1.325
cut5	-2.774 ***	0.954	-1.600	1.325
cut6	-2.149 **	0.951	-0.972	1.321
cut7	-1.495	0.956	-0.314	1.322
cut8	-0.950	0.935	0.237	1.305
cut9	-0.465	0.940	0.727	1.303
cut10	-0.183	0.945	1.011	1.316
Number of observations	278		278	
Likelihood function	-542.607		-541.469	
Pseudo R ²	0.094		0.096	
Prob > chi2	0.000		0.000	
Akaike IC	4.062		4.061	
Number of iterations	4		4	

Notes. This table reports results from ordered probit estimations of Eq. (92). Regression in column (2) includes the inverse Mills ratio to control for selection. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 18: CS, third specification

	(1) CS*		(2) CS*	
	Coef.	Std. err.	Coef.	Std. err.
Crowding effect:				
IVD (users/m ²) X Standing NF	-0.019 **	0.008	-0.019 **	0.008
IVD (users/m ²) X Over-closeness NF	-0.016	0.010	-0.015	0.010
IVD (users/m ²) X Noise NF	-0.014*	0.008	-0.015*	0.008
IVD (users/m ²) X Robbery NF	0.001	0.009	0.001	0.009
IVD (users/m ²) X Fall NF	0.006	0.009	0.007	0.009
IVD (users/m ²) X Smell NF	-0.007	0.009	-0.007	0.009
IVD (users/m ²) X Time Loss NF	-0.011	0.009	-0.013	0.009
IVD (users/m ²) X Wasted Time NF	-0.020 **	0.008	-0.020 **	0.008
Journey controls:				
Line (1=line 1/0=line 4)	-0.150	0.162	-0.373*	0.198
Door to door travel time (hours)	-0.026	0.124	-0.002	0.125
In-vehicule travel time (hours)	-0.121	0.618	-0.169	0.629
Peak hour (1=morning/0=evening)	0.269 **	0.135	-0.364	0.371
Daily usage of the line (1=Y/0=N)	-0.024	0.133	0.167	0.168
Individual controls:				
Gender (1=male/0=female)	0.035	0.137	0.004	0.137
Car available (1=Y/0=N)	-0.346 ***	0.130	-0.368 **	0.131
ln(Individual net monthly income (euros))	-0.044	0.099	-0.110	0.101
Age (years)	-0.289	0.639	-0.109	0.642
$\sum NFD$	0.015 **	0.007	0.015 **	0.007
Selection control:				
$\hat{\lambda}$			1.823*	0.966
cut1	-3.469 ***	0.672	-2.014*	1.073
cut2	-3.012 ***	0.668	-1.560	1.076
cut3	-2.302 ***	0.649	-0.845	1.064
cut4	-1.868 ***	0.641	-0.408	1.058
cut5	-1.397 **	0.639	0.067	1.057
cut6	-0.759	0.639	0.708	1.057
cut7	-0.106	0.645	1.366	1.059
cut8	0.446	0.634	1.926*	1.053
cut9	0.962	0.651	2.450 **	1.058
cut10	1.281*	0.679	2.773 **	1.090
Number of observations	278		278	
Likelihood function	-535.458		-533.784	
Pseudo R ²	0.106		0.109	
Prob > chi2	0.000		0.000	
Akaike IC	4.054		4.049	
Number of iterations	4		4	

Notes. This table reports results from ordered probit estimations of Eq. (93). Regression in column (2) includes the inverse Mills ratio to control for selection. *significant at 10%; **significant at 5%; ***significant at 1%.

Appendix E Users preferences for the nuisance factors

We test some socioeconomic variables that may drive the dissatisfaction of the three nuisance factors by which the crowding effect decreases the comfort satisfaction: *Standing*, *Wasted Time* and *Noise*.

Appendix E.1 Estimation strategy

In *Questions B*, users mark their level of dissatisfaction about the nuisance factors of crowding assuming that the in-vehicle density is the highest, i.e 6 users per square meter. We wish to test if some socioeconomic variables drive these self-reported marks. Due to the nature of data, ordered probit models are used to explain the marking of features. We therefore estimate the following equation:

$$NFD_d^* = \beta_{1d}X + \beta_{2d}Z + \beta_{3d}\hat{\lambda} + \varepsilon_d \quad (95)$$

where NFD_d^* is the latent variable associated with the dissatisfaction mark given to the nuisance factor d . X is a set of individual characteristics: gender (dummy), car availability (dummy), $\ln(\text{Individual net monthly income (euros)})$, age (centuries) and live in Paris (dummy). It is conceivable that answers to *Questions B* are affected by the current journey of users. To control for these effects, we also include a characteristics of the journey, Z : line where the user is surveyed (dummy) and the immediate journey travel time (hours). $\hat{\lambda}$ is the inverse Mills ratio estimated in Appendix B. It allows to control for the selection bias. Finally, we want to focus on the drivers of the relative contribution of each feature to the crowding dissatisfaction. Dissatisfaction is assumed to be ordinally comparable across users.

Appendix E.2 Regression results

To provide more details on the users preferences, we now turn to econometric analysis. Tables 19 reports the results of estimating (95) through ordered probit procedures.

First, there is a clear gender effect: men are a lot less dissatisfied than women by the three nuisance factors. This is in line with results from Meyers-Levy and Maheswaran

Table 19: Ordered probit estimates for NFD_{Standing} (1), $NFD_{\text{Wasted Time}}$ (2), and NFD_{Noise} (3)

	(1)		(2)		(3)	
	Standing NF Coef.	Std. err.	Wasted Time NF Coef.	Std. err.	Noise NF Coef.	Std. err.
Individual effects:						
Gender (1=male/0=female)	-0.327 **	0.129	-0.236*	0.127	-0.373 ***	0.127
Car available (1=Y/0=N)	-0.037	0.140	-0.323 **	0.139	0.096	0.138
ln(Individual income)	0.041	0.099	0.225 **	0.098	0.165*	0.098
Age (years)	-0.196	0.575	-1.116 **	0.569	-0.173	0.562
Live in Paris (1=Y/0=N)	-0.170	0.134	0.031	0.133	-0.220*	0.132
Journey controls:						
Line (1=line 1/0=line 4)	-0.216	0.144	-0.159	0.142	-0.020	0.141
In-vehicule travel time (hours)	1.281 **	0.628	-0.363	0.620	0.240	0.620
Selection control:						
$\hat{\lambda}$	0.202	0.345	0.281	0.340	-0.240	0.339
cut1	-0.996	0.668	0.087	0.653	-0.589	0.650
cut2	-0.934	0.668	0.156	0.653	-0.464	0.650
cut3	-0.678	0.666	0.435	0.654	-0.176	0.651
cut4	-0.473	0.665	0.607	0.654	0.111	0.651
cut5	-0.365	0.665	0.740	0.655	0.226	0.652
cut6	0.066	0.665	1.287	0.655	0.798	0.654
cut7	0.177	0.665	1.491	0.655	1.057	0.655
cut8	0.361	0.665	1.654	0.655	1.344	0.654
cut9	0.889	0.665	2.217	0.658	1.823	0.655
cut10	1.120	0.666	2.641	0.662	2.177	0.657
Number of observations	278		278		278	
Likelihood function	-592.841		-609.196		-621.578	
Pseudo R ²	0.011		0.015		0.012	
Prob > chi2	0.124		0.021		0.049	
Number of iterations	3		3		3	

Notes. This table reports coefficients from ordered probit estimations of NFD_{Noise}^* and $NFD_{\text{Wasted Time}}^*$, the latent variables associated with the reported NFD_{Noise} and $NFD_{\text{Wasted Time}}$. *significant at 10%; **significant at 5%; ***significant at 1%.

(1991) and Meyers-Levy and Sternthal (1991) who found that women have been found to process information in more detail, resulting in a greater sensitivity to environmental factors.

A large income effect is also noticeable. Wealthier users are more bothered by the *Wasted Time* and *Noise* than poorer ones. This effect is not surprising and corresponds to results found in Section 3. It is consistent with their higher value of time.

The car ownership influences the perception of crowding nuisances. Car-owner users seem to compare the crowding conditions in PT with the individual car travel conditions.

As a consequence, they find the *Wasted Time* less penalizing than other users do, maybe because they know that they can occupy their travel time in a better way than if they had to focus on the road traffic, whatever the level of density.

An age effect is perceptible as for *Wasted Time*, and a “Parisian” effect plays in the *Noise* perception.

Conclusion

This dissertation addressed the foundations of users behavior with respect to the congestion in public transport. Chapter I investigated the two-way relation between public transport patronage and level of punctuality. It showed that in usual cases reliability and public transport patronage in equilibrium are lower than they would be in optimum. Moreover, it showed that an increase in the travel cost of the alternative mode increases the reliability of the public transport mode via a demand effect. This is reminiscent of the well-known Mohring effect. However, there are some limitations. For example, public transport punctuality is taken into account whereas alternative mode reliability is not. If the alternative mode circulates on roads, as it is the case with the taxi mode, the unreliability due to traffic jams may be a crucial issue and should be incorporated in the study. To the best of my knowledge, the model developed in this chapter is the first to analyze the punctuality of public transport together with a double heterogeneity - in preferred arrival time and in location - of users. It is also of primary interest for microeconomic modeling.

Chapter II addressed the user trade-off between schedule delay cost and congestion cost. The displayed model is to the public transport what the bottleneck model is to the private transport. I showed that the congestion in public transport has properties that are different from congestion on private transport. In particular, a time-varying optimal pricing does not entirely remove the congestion. Optimal supply is investigated. The model is applied to a Paris RER A segment serving La Défense station. The implementation of an optimal train dependent pricing is found to improve the social welfare of €0.45 per individual journey. The specification of the model could be improve by including others aspects. Heterogeneity of users is one of them. It is essential to study who loose and who

win when crowding pricing is implemented, especially if users differ in wealth, location, scheduling possibilities or crowding disutility.

Chapter III provided an econometric analysis of the crowding effect. It showed that the disutility due to crowding increases with the density of users and with the wealth of users. Moreover, three causes of crowding disutility are identified: a higher probability to stand for all or part of the journey, a poorer use of the time during the journey, and noisier travel conditions. As in most empirical studies, the quality and the precision of the analysis is constrained by the data available. Thus a larger sample would be appreciable in such an analysis. In the same vein, many empirical studies on subjective well-being use time fixed effects tools to control for the reference points of individuals. Such keenness is absent from this chapter. Nevertheless, the results seem robust and interesting for public policy makers.

Through this dissertation, I studied the micro-foundations of the behaviors of public transport users. I focused on their perception of crowding and on the way they get used to unreliability and crowd. The comprehension of these behaviors is of first interest to improve the efficiency of policies aiming at increasing the patronage for public transport. Indeed, the misreading of these mechanisms may be an important obstacle to any measure targeting a shift in the modal split. In brief, the service quality, a crucial aspect of the public transport appeal, varies with the patronage. Of course, there is still a lot to be done in the public transport economics. This field is expected to remain topical in the coming years, in particular considering the foretold increase in size of cities in most areas of the world.

As I am concluding my thesis, I am left with some unanswered questions. Part of these questions fall into microeconomic theory, others are more related to econometrics, but all of them regard the ways public transport could be improved. One question I would like to investigate is the impact of users heterogeneity on optimal public transport supply and pricing. Which type of users would take advantage of time dependent pricing? I would like to investigate how the heterogeneity in value of time or in the scheduling possibility influence the trade-off between crowding and scheduling. This study could take place in a bi-modal framework characterized by a public transport mode subject

to crowding, and a private mode subject to traffic jams and increased travel times.

Another potential extension of Chapter II is to consider more complex public transport networks. This can be done by introducing more than two stations in the line and by allowing users to depart and arrive at any of these stations. The limit case of such extension would be to model complete networks with several lines and several modes. Coding and simulations would be naturally needed. Such a model would allow to produce accurate estimations for cost-benefit analysis.

A project I plan to carry out is to investigate the optimal public transport network size and the optimal level of subsidies. These questions are of main interest. Indeed, public transport benefits are sometimes fantasized by politicians who are appealed by a showcase effect. For example, the building of a new tramway facility gives the image of a city dynamic and concerned about all environmental issues. However, what are the real consequences of such a project? To attempt to answer these questions, I intend to conduct an econometric study of the changes in patronage for public transport and in modal split following an increase in road or public transport supply or in service quality, at a city level. This could help to better understand impact and efficiency of public transport policies.

Bibliography

- Abou-Zeid, M., Ben-Akiva, M., 2011. The effect of social comparisons on commute well-being. *Transportation Research Part A: Policy and Practice* 45 (4), 345–361.
- Abou-Zeid, M., Ben-Akiva, M., 2014. Satisfaction and travel choices. In: *Handbook of Sustainable Travel*. Springer, pp. 53–65.
- Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V., Ben-Akiva, M., 2012. Happiness and travel mode switching: findings from a swiss public transportation experiment. *Transport Policy* 19 (1), 93–104.
- Allen, J., Levinson, H., 2014. Accommodation of long-term growth on north america’s commuter railroads. *Transportation Research Record* (2419), 40–49.
- Alonso, W., 1964. *Location and land use: toward a general theory of land rent*. Harvard University Press.
- Anderson, M. L., 2014. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *The American Economic Review* 104 (9), 2763–96.
- Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. *Journal of Urban Economics* 27 (1), 111–130.
- Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review* 83 (1), 161–179.
- Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P., 2004. *The demand for public transport: a practical guide*.

- Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review* 37 (2), 191–229.
- Baum, A., Greenberg, C. I., 1975. Waiting for a crowd: The behavioral and perceptual effects of anticipated crowding. *Journal of Personality and Social Psychology* 32 (4), 671–679.
- Baum, A., Paulus, P. B., 1987. Crowding. *Handbook of environmental psychology* 1, 533–570.
- Becker, G. S., 1965. A theory of the allocation of time. *The Economic Journal* 75 (299), 493–517.
- Beirao, G., Cabral, J. S., 2007. Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy* 14 (6), 478–489.
- Bhattacharya, S. K., Bandyopadhyay, P., Kashyap, S. K., 1995. Calcutta metro: is it safe from noise pollution hazards? *Industrial Health* 34 (1), 45–50.
- Bon, D., Scherer, S., Zauberman, R., Robert, P., Heurtel, H., 2011. L'insécurité en île-de-france. rapport final: enquête "victimation et sentiment d'insécurité" de 2009.
- Bousquet, D., Moiron-Braud, E., Ronai, E., Ressot, C., 2015. Avis sur le harcèlement sexiste et les violences sexuelles dans les transports en commun. Tech. rep., Haut conseil à l'égalité entre les femmes et les hommes.
- Boussenna, M., Corlett, E., Pheasant, S., 1982. The relation between discomfort and postural loading at the joints. *Ergonomics* 25 (4), 315–322.
- Bowman, L. A., Turnquist, M. A., 1981. Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General* 15 (6), 465–471.
- Cantwell, M., Caulfield, B., O'Mahony, M., 2009. Examining the factors that impact public transport commuting satisfaction. *Journal of Public Transportation* 12 (2), 1–21.

- Cartledge, J., 2003. Overcrowding on the railways, a health and safety issue? In: Presentation to HSC Railway Industry Advisory Committee, Birmingham.
- Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., Coates, T. J., 1996. Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly* 60 (3), 345–375.
- CEBR - INRIX, 2014. Annual cost of gridlock in europe and the us will increase 50 percent on average to \$293 billion by 2030. Tech. rep.
URL <http://inrix.com/press/annual-cost-of-gridlock-in-europe-and-the-us-will-increase-50-percent-on-average-to-293-billion-by-2030/>
- Chakraborty, P., September 7th 2011. Overcrowding in city buses a real menace. *The Times of India*. Last accessed 16 August 2015.
URL <http://timesofindia.indiatimes.com/city/nagpur/Overcrowding-in-city-buses-a-real-menace/articleshow/9892390.cms>
- Chu, X., 1995. Endogenous trip scheduling: the henderson approach reformulated and compared with the vickrey approach. *Journal of Urban Economics* 37 (3), 324–343.
- Ciccone, A., 2002. Agglomeration effects in europe. *European Economic Review* 46 (2), 213–227.
- Clark, A. E., Senik, C., 2010. Who compares to whom? the anatomy of income comparisons in europe. *The Economic Journal* 120 (544), 573–594.
- Coulombel, N., de Palma, A., 2014. Variability of travel time, congestion, and the cost of travel. *Mathematical Population Studies* 21 (4), 220–242.
- Cox, T., Houdmont, J., Griffiths, A., 2006. Rail passenger crowding, stress, health and safety in britain. *Transportation Research Part A: Policy and Practice* 40 (3), 244–258.
- CRCFRI, 2012. A literature review of passenger crowding on metropolitan railways: psychological, cultural human and economic factors. Project R2.104.

- De Borger, B., Van Dender, K., 2006. Prices, capacities and service levels in a congestible bertrand duopoly. *Journal of Urban Economics* 60 (2), 264–283.
- De Borger, B., 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 Lapparent, M., Koning, M., 2015. Analyzing time sensitivity to discomfort in the paris subway: an interval data model approach. *Transportation*, 1–21.
- De Lara, M., de Palma, A., Kilani, M., Piperno, S., 2013. Congestion pricing and long term urban form: Application to paris region. *Regional Science and Urban Economics* 43 (2), 282–295.
- de Palma, A., Kilani, M., Proost, S., 2015. Discomfort in mass transit and its implication for scheduling and pricing. *Transportation Research Part B: Methodological* 71 (0), 1–18.
- de Palma, A., Lindsey, R., 2001. Optimal timetables for public transportation. *Transportation Research Part B: Methodological* 35 (8), 789–813.
- de Palma, A., Lindsey, R., Monchambert, G., 2014. Economics of crowding in public transport. Working Paper presented at the ITEA Conference, Toulouse 2014.
- Delbosc, A., 2012. The role of well-being in transport policy. *Transport Policy* 23, 25–33.
- Dell’Olio, L., Ibeas, A., Cecin, P., 2011. The quality of service desired by public transport users. *Transport Policy* 18 (1), 217–227.
- DeSerpa, A. C., 1971. A theory of the economics of time. *The Economic Journal* 81 (324), 828–846.
- Diener, E., Oishi, S., Lucas, R. E., 2009. 17 subjective well-being: The science of happiness and life satisfaction. *Oxford handbook of positive psychology*, 187.
- Eboli, L., Mazzulla, G., 2007. Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation* 10 (3), 21–34.

- ECMT, 1999. *Traffic Congestion in Europe*. OECD Publishing.
- Epstein, Y. M., 1981. Crowding stress and human behavior. *Journal of Social Issues* 37 (1), 126–144.
- Epstein, Y. M., Teitelbaum, R., Karlin, R. A., Katz, S., Aiello, J. R., 1981. An assessment of the effectiveness of two tactics to reduce arousal in crowded mass transit settings. *Journal of Applied Social Psychology* 11 (3), 259–267.
- Ettema, D., Friman, M., Gärling, T., Olsson, L. E., Fujii, S., 2012. How in-vehicle activities affect work commuters' satisfaction with public transport. *Journal of Transport Geography* 24, 215–222.
- Ettema, D., Gärling, T., Eriksson, L., Friman, M., Olsson, L. E., Fujii, S., 2011. Satisfaction with travel and subjective well-being: Development and test of a measurement tool. *Transportation Research Part F: Traffic Psychology and Behaviour* 14 (3), 167–175.
- Ettema, D., Gärling, T., Olsson, L. E., Friman, M., 2010. Out-of-home activities, daily travel, and subjective well-being. *Transportation Research Part A: Policy and Practice* 44 (9), 723–732.
- European Commission, 2011. *WHITE PAPER Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system*.
- Fosgerau, M., Karlström, A., 2010. The value of reliability. *Transportation Research Part B: Methodological* 44 (1), 38–49.
- Fosgerau, M., Small, K. A., 2013. Hypercongestion in downtown metropolis. *Journal of Urban Economics* 76 (0), 122–134.
- Gershon, R. R., Qureshi, K., Barrera, M., Erwin, M., Goldsmith, F., 2005. Health and safety hazards associated with subways: a review. *Journal of Urban Health* 82 (1), 10–20.
- Haywood, L., Koning, M., 2015. The distribution of crowding costs in public transport:

- New evidence from Paris. *Transportation Research Part A: Policy and Practice* 77, 182–201.
- Heckman, J. J., 1979. Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society* 47 (1), 153–161.
- Henderson, J. V., 1974. Road congestion: a reconsideration of pricing theory. *Journal of Urban Economics* 1 (3), 346–365.
- Henderson, J. V., 1981. The economics of staggered work hours. *Journal of Urban Economics* 9 (3), 349–364.
- Henderson, J. V., 1997. Externalities and industrial development. *Journal of Urban Economics* 42 (3), 449–470.
- Hensher, D. A., 1998. The imbalance between car and public transport use in urban australia: why does it exist? *Transport Policy* 5 (4), 193–204.
- Hensher, D. A., Stopher, P., Bullock, P., 2003. Service quality - developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A: Policy and Practice* 37 (6), 499–517.
- Huang, H.-J., Tian, Q., Gao, Z.-Y., 2005. An equilibrium model in urban transit riding and fare polices. In: *Algorithmic Applications in Management*. Vol. 3521. Springer Berlin Heidelberg, pp. 112–121.
- Huang, H.-J., Tian, Q., Yang, H., Gao, Z.-Y., 2007. Modal split and commuting pattern on a bottleneck-constrained highway. *Transportation Research Part E: Logistics and Transportation Review* 43 (5), 578–590.
- Huddy, L., Billig, J., Braccioldieta, J., Hoeffler, L., Moynihan, P., Pugliani, P., 1997. The effect of interviewer gender on the survey response. *Political Behavior* 19 (3), 197–220.
- Jain, J., Lyons, G., 2008. The gift of travel time. *Journal of Transport Geography* 16 (2), 81–89.

- Jensen, M., 1999. Passion and heart in transport - a sociological analysis on transport behaviour. *Transport Policy* 6 (1), 19–33.
- Joly, I., Masson, S., Petiot, R., 2006. Les déterminants de la demande en transports collectifs urbains: comparaison internationale et analyse économétrique. *Les Cahiers Scientifiques du Transport* (50), 91–120.
- Kahneman, D., Diener, E., Schwarz, N., 1999. *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.
- Kahneman, D., Krueger, A. B., 2006. Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives* 20 (1), 3–24.
- Kane, E. W., Macaulay, L. J., 1993. Interviewer gender and gender attitudes. *Public Opinion Quarterly* 57 (1), 1–28.
- Kelton, S., August 9th 2012. Overcrowding, cost and unreliability causes commuters to abandon public transport. *The Advertiser, Adelaide*. Last accessed 16 August 2015.
URL <http://www.news.com.au/national/south-australia/adelaide-public-transport-usage-has-slumped-dramatically-new-figures-reveal/story-fndo4dzn-1226446709644>
- Kilani, M., Proost, S., van der Loo, S., 2014. Road pricing and public transport pricing reform in paris: complements or substitutes? *Economics of Transportation*, 175–187.
- King, D., Srikuenthiran, S., Shalaby, A., 2014. Using simulation to analyze crowd congestion and mitigation at canadian subway interchanges. *Transportation Research Record: Journal of the Transportation Research Board* 2417 (1), 27–36.
- Knockaert, J., Tseng, Y.-Y., Verhoef, E. T., Rouwendal, J., 2012. The spitsmijden experiment: A reward to battle congestion. *Transport Policy* 24, 260–272.
- Koning, M., 2011. *Essais sur la congestion dans les transports à paris*. Phd thesis, Université Panthéon-Sorbonne - Paris I.
URL <https://tel.archives-ouvertes.fr/tel-00717656/document>

- Kraus, M., 1991. Discomfort externalities and marginal cost transit fares. *Journal of Urban Economics* 29 (2), 249–259.
- Kraus, M., 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.
- Kroes, E., Kouwenhoven, M., Debrincat, L., Pauget, N., 2014. Value of crowding on public transport in île-de-france, france. *Transportation Research Record: Journal of the Transportation Research Board* 2417 (1), 37–45.
- Krueger, A. B., Schkade, D. A., 2008. The reliability of subjective well-being measures. *Journal of Public Economics* 92 (8), 1833–1845.
- Langrehr, F. W., 1991. Retail shopping mall semiotics and hedonic consumption. *Advances in Consumer Research* 18 (1), 428–433.
- Li, Z., 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. M., Wilson, N. H., 1992. Dwell time relationships for light rail systems. *Transportation Research Record* (1361).
- Lindsey, R., 2012. Road pricing and investment. *Economics of transportation* 1 (1), 49–63.
- Litman, T., 2008. Valuing transit service quality improvements. *Journal of Public Transportation* 11 (2), 43–63.
- Litman, T., 2013. Understanding transport demands and elasticities.how prices and other factors affect travel behavior. Retrieved from <http://www.vtpi.org/elasticities.pdf>.
- McFadden, D., 1974. The measurement of urban travel demand. *Journal of Public Economics* 3 (4), 303–328.
- Metcalfe, R., Dolan, P., 2012. Behavioural economics and its implications for transport. *Journal of Transport Geography* 24, 503–511.
- Meyers-Levy, J., Maheswaran, D., 1991. Exploring differences in males' and females' processing strategies. *Journal of Consumer Research*, 63–70.

- Meyers-Levy, J., Sternthal, B., 1991. Gender differences in the use of message cues and judgments. *Journal of Marketing Research* 28 (1), 84–96.
- Mills, E. S., 1967. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review* 57 (2), 197–210.
- Mohd Mahudin, N. D., Cox, T., Griffiths, A., 2012. Measuring rail passenger crowding: Scale development and psychometric properties. *Transportation Research Part F: Traffic Psychology and Behaviour* 15 (1), 38–51.
- Mohring, H., 1972. Optimization and scale economies in urban bus transportation. *The American Economic Review* 62 (4), 591–604.
- Monchambert, G., de Palma, A., 2014. Public transport reliability and commuter strategy. *Journal of Urban Economics* 81, 14–29.
- Morris, E. A., Guerra, E., 2014. Mood and mode: does how we travel affect how we feel? *Transportation* 42 (1), 1–19.
- Muth, R. F., 1969. *City and Housing*. Chicago: University of Chicago Press.
- Noland, R. B., Small, K. A., Koskenoja, P. M., Chu, X., 1998. Simulating travel reliability. *Regional Science and Urban Economics* 28 (5), 535–564.
- Oakland, W. H., 1972. Congestion, public goods and welfare. *Journal of Public Economics* 1, 339–357.
- OECD, 2007. *Managing Urban Traffic Congestion*. OECD Publishing.
- OECD, 2014. *Valuing convenience in public transport*. Tech. rep., ITF Round Tables.
- Oldfield, R., Bly, P., 1988. An analytic investigation of optimal bus size. *Transportation Research Part B: Methodological* 22 (5), 319–337.
- OMNIL, 2011. *Les transports en commun en chiffres - édition 2011*. Tech. rep., last accessed 16 August 2015.
URL http://www.omnil.fr/IMG/pdf/transports_en_commun_en_chiffres_edition_2000-2009.pdf

- Oum, T. H., Waters II, W. G., Fu, X., 2008. Transport demand elasticities. In: Hensher, D., Button, K. (Eds.), *Handbook of Transport Modelling*, 2nd edition. Oxford: Elsevier.
- Parry, I. W., Small, K. A., 2009. Should urban transit subsidies be reduced? *The American Economic Review* 99 (3), 700–724.
- Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., White, P., 2006. The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy* 13 (4), 295–306.
- Peer, S., Verhoef, E., Knockaert, J., Koster, P., Tseng, Y.-Y., 2015. Long-run versus short-run perspectives on consumer scheduling: evidence from a revealed-preference experiment among peak-hour road commuters. *International Economic Review* 56 (1), 303–323.
- Pepper, J., Spitz, G., Adler, T., 2003. Customers perspectives on using multilevel coaches to increase rail system capacity. In: *Transportation Research Boards 82nd Annual Meeting*, Washington, DC.
- Ponti, M., 2011. *A Handbook of Transport Economics*. Edward Elgar, Ch. 28 Competition, regulation and public service obligations, pp. 661–683.
- Pownall, C., Prior, M., Segal, J., 2008. What rail passengers will do to get a seat. In: *European Transport Conference 2008; Proceedings*.
- Proost, S., Van Dender, K., 2001. The welfare impacts of alternative policies to address atmospheric pollution in urban road transport. *Regional Science and Urban Economics* 31 (4), 383–411.
- Proost, S., Van Dender, K., Courcelle, C., De Borger, B., Peirson, J., Sharp, D., Vickerman, R., Gibbons, E., O'Mahony, M., Heaney, Q., Van den Bergh, V., Verhoef, E., 2002. How large is the gap between present and efficient transport prices in europe? *Transport Policy* 9 (1), 41–57.
- Prud'homme, R., Koning, M., Lenormand, L., Fehr, A., 2012. Public transport congestion costs: The case of the Paris subway. *Transport Policy* 21, 101–109.

- Raveau, S., Muñoz, J. C., De Grange, L., 2011. A topological route choice model for metro. *Transportation Research Part A: Policy and Practice* 45 (2), 138–147.
- Redman, L., Friman, M., Gärling, T., Hartig, T., 2013. Quality attributes of public transport that attract car users: A research review. *Transport Policy* 25 (0), 119–127.
- Rothschild, M., Stiglitz, J. E., 1970. Increasing risk: I. a definition. *Journal of Economic Theory* 2 (3), 225–243.
- Small, K. A., 1982. The scheduling of consumer activities: Work trips. *The American Economic Review* 72 (3), 467–479.
- Small, K. A., 2008. Urban transportation. *The Concise Encyclopedia of Economics*. Last accessed 16 August 2015.
URL <http://www.econlib.org/library/Enc/UrbanTransportation.html>
- Small, K. A., Verhoef, E. T., 2007. *The economics of urban transportation*. Routledge.
- Steg, L., 2005. Car use: lust and must. instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice* 39, 147–162.
- Steinmann, L., 2013. Horaires de travail décalés : l'idée de la sncf approuvée par les usagers. *Les Echos*, 2 April 2013. Last accessed 16 August 2015.
URL http://www.lesechos.fr/02/04/2013/LesEchos/21408-070-ECH_horaires-de-travail-decales---l-idee-de-la-sncf-approuvee-par-les-usagers.htm#RDMb4p2ErRAM033w.99
- STIF, 2013. Le financement du fonctionnement des transports publics. Tech. rep., last accessed 7 June 2013.
URL <http://www.stif.org/organisation-missions/volet-economique/financement-transports-publics/financement-transports-franciliens-442.html>
- STIF, 2014. La qualité de service en chiffres - bulletin d'information trimestriel sur la qualité de service des transports en Île-de-france. Tech. rep., last accessed 16 August 2015.
URL <http://www.stif.org/IMG/pdf/bulletin-qualite-service-15.pdf>

- Stokols, D., 1972. On the distinction between density and crowding: Some implications for future research. *Psychological review* 79 (3), 275.
- Stutzer, A., Frey, B. S., 2008. Stress that doesn't pay: The commuting paradox. *The Scandinavian Journal of Economics* 110 (2), 339–366.
- Sutton, J., 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. The MIT press.
- Tabuchi, T., 1993. Bottleneck congestion and modal split. *Journal of Urban Economics* 34 (3), 414–431.
- The Economist, Oct 19th 2003. Transport in london - underground, overground. *The Economist*, 19 October 2003.. Last accessed 16 August 2015.
URL <http://www.economist.com/news/britain/21588072-london-has-built-about-good-transport-network-it-could-given-its-constraints-time>
- The Economist, April 28th 2011. America's transport infrastructure: Life in the slow lane. *The Economist*, 28 April 2011. Last accessed 16 August 2015.
URL <http://www.economist.com/node/18620944>
- Tian, Q., Huang, H.-J., Lam, W., 2009a. How do transit commuters make trade-offs between schedule delay penalty and congestion cost? Empirical study in Beijing. *Transportation Research Record* (2134), 164–170.
- Tian, Q., Huang, H.-J., Yang, H., 2007. Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system. *Transportation Research Part B: Methodological* 41 (6), 616–631.
- Tian, Q., Yang, H., Huang, H., Mou, H., 2009b. Dynamic congestion pricing in urban transit system. In: *International Conference on Transportation Engineering 2009*. ASCE, pp. 1542–1547.
- Tirachini, A., Hensher, D. A., Jara-Díaz, S. R., 2010. 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), 231–242.

- Tirachini, A., Hensher, D. A., Rose, J. M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice* 53, 36–52.
- Tseng, Y.-Y., Verhoef, E. T., 2008. Value of time by time of day: A stated-preference study. *Transportation Research Part B: Methodological* 42 (7), 607–618.
- Turner, S., Corbett, E., O'Hara, R., White, J., 2005. Health and safety effects of crowding: Hazard identification (t307). London: Health & Safety Laboratory.
- UITP, 2014a. Key statistics. Tech. rep., last accessed 16 August 2015.
URL <http://www.uitp.org/key-eu-statistics>
- UITP, 2014b. Local public transport trends in the european union. Tech. rep., last accessed 16 August 2015.
URL <http://www.uitp.org/statistics-brief-local-public-transport-trends-european-union>
- UITP, 2014c. World metro figures. Tech. rep., last accessed 16 August 2015.
URL <http://www.uitp.org/world-metro-figures-2013>
- Uzzell, D., Brown, J., 2007. Conceptual progress in understanding fear of crime in railway stations. *Psicologia* 21 (2), 119–137.
- Van Praag, B. M., Ferrer-i Carbonell, A., 2008. Happiness quantified: A satisfaction calculus approach. Oxford University Press.
- Van Praag, B. M., Frijters, P., Ferrer-i Carbonell, A., 2003. The anatomy of subjective well-being. *Journal of Economic Behavior & Organization* 51 (1), 29–49.
- Veitch, T., Partridge, J., Walker, L., 2013. Estimating the costs of over-crowding on melbourne's rail system. In: Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia.
- Vickrey, W. S., 1969. Congestion theory and transport investment. *The American Economic Review* 59 (2), 251–260.

- Vine, I., 1982. Crowding and stress: 2. a personal space approach. *Current Psychological Reviews* 2 (1), 1–18.
- Vives, X., 1990. Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics* 19 (3), 305–321.
- Vuchic, V. R., 2005. *Urban transit: operations, planning, and economics*. Wiley, New York.
- Wardman, M., 2004. Public transport values of time. *Transport Policy* 11 (4), 363–377.
- Wardman, M., Chintakayala, P., de Jong, G., Ferrer, D., 2012. European wide meta-analysis of values of travel time. Tech. rep., Significance.
- Wardman, M., Murphy, P., 2015. Passengers' valuations of train seating layout, position and occupancy. *Transportation Research Part A: Policy and Practice* 74, 222–238.
- Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: evidence and lessons from british experience. *Transport Reviews* 31 (3), 379–398.
- Wei, X., Jul 26th 2011. Officials urge to ease subway overcrowding. *China Daily*. Last accessed 16 August 2015.
URL http://usa.chinadaily.com.cn/china/2011-07/26/content_12983921.htm
- Willsher, K., May 21th 2014. French railway operator sncf orders hundreds of new trains that are too big. *The Guardian*. Last accessed 16 August 2015.
URL <http://www.theguardian.com/world/2014/may/21/french-railway-operator-sncf-orders-trains-too-big>
- Winston, C., Maheshri, V., 2007. On the social desirability of urban rail transit systems. *Journal of Urban Economics* 62 (2), 362–382.
- Xie, C., Fukuda, D., 2014. Scheduling preference modeling of rail passengers in the tokyo metropolitan area and evaluation of time-varying fare policy for a congested urban railway line. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)* 70 (5), 535–548.

Three essays on the economics of congestion in public transport

Abstract: This dissertation addresses the foundations of user's behavior with respect to the congestion in public transport. It is made of three distinct essays. The two first essays investigate how users get used to lack of punctuality and crowding. The third essay presents an empirical analysis of the crowding effect. In the first chapter, I consider the modeling of a bi-modal competitive network involving a public transport mode, which may be unreliable, and an alternative mode. The public transport reliability set by the public transport firm at the competitive equilibrium increases with the alternative mode fare, via a demand effect. This is reminiscent of the Mohring Effect. The study of the optimal service quality shows that often, public transport reliability and thereby patronage are lower at equilibrium compared to first-best social optimum. In the second chapter, to study the behavioral implications and costs of crowding, I develop a structural model in which public transport users face a choice between traveling in a crowded train and arriving when they want, and traveling earlier or later to avoid crowding but arriving at an inconvenient time. I derive the user equilibrium and socially optimal distribution of passengers across trains, show how the optimum can be decentralized using train-specific fares, and characterize the welfare gains from optimal pricing. Properties of the model are compared with those obtained from the bottleneck and flow congestion models of road traffic. In the third chapter, I investigate the influence of in-vehicle crowding on the comfort satisfaction experienced during a public transport journey. Moreover, I describe the anatomy of the crowding effect by testing various nuisance factors (Smell, Noise, Standing...) as channels through which crowding may decrease the comfort satisfaction. I find a clear crowding effect: on average, an extra-user per square meter decreases by one the expected 0 to 10 scale individual well-being. I do not find any empirical evidence of this effect being intensified by the travel time. However, the crowding effect increases with the income of users. I find three causes of crowding disutility: a higher probability to stand for all or part of the journey, a poorer use of the time during the journey, and noisier travel conditions. These features of discomfort matter more for women and wealthy individuals.

Keywords: Congestion; Crowding; Reliability; Public transport; Microeconomics; Microeconometrics.

Trois essais sur l'économie de la congestion dans les transports publics

Résumé : Cette thèse s'intéresse aux fondements du comportement des usagers face à la congestion dans les transports publics. Elle se compose de trois essais distincts. Les deux premiers essais examinent l'adaptation des usagers au manque de ponctualité et à l'inconfort. Le troisième essai offre une analyse empirique de l'effet «foule ». Dans le premier chapitre, je présente un modèle de compétition bimodale entre une ligne de transport public sujette aux retards, et un mode alternatif. A l'équilibre du marché, le niveau de fiabilité choisi par la firme en charge des transports publics augmente avec le prix du mode alternatif via un effet «demande », évoquant l'effet Mohring. L'étude de la qualité de service optimale montre que souvent, la fiabilité des transports publics et donc leur fréquentation à l'équilibre sont inférieurs à ce qu'ils seraient à l'optimum social. Dans le second chapitre, afin d'étudier le coût de l'inconfort et ses implications quant au choix de l'heure de déplacement, je développe un modèle structurel dans lequel les usagers des transports public choisissent entre voyager dans un véhicule bondé et arriver à destination à l'heure désirée ou alors voyager plus tard ou plus tôt pour éviter la foule. J'établis la distribution des usagers parmi les véhicules à l'équilibre et à l'optimum social, et montre que l'optimum social peut être décentralisé par une tarification fine par véhicule. Les propriétés du modèle sont comparées avec celles du modèle du «bottleneck »et des modèles de congestion routière. Dans le troisième chapitre, j'analyse l'influence de la densité d'usagers sur la satisfaction liée au confort durant un déplacement en transport public. De plus, je décris l'anatomie de l'effet «foule »en testant différents aspects (odeur, bruit, position debout...) comme des causes d'inconfort lorsque la densité d'usagers augmente. J'identifie un net effet «foule »: en moyenne, un usager supplémentaire par mètre carré diminue de 1 la satisfaction liée au confort qui est mesurée sur une échelle de 0 à 10. Je ne trouve pas de corrélation entre temps de transport et l'effet «foule ». Cependant, cet effet augmente avec le revenu des usagers. Trois causes de désutilité liée à la foule sont identifiées : une plus grande probabilité de devoir voyager debout, un usage limité du temps, et des conditions de déplacement plus bruyantes. Ces désagréments sont plus importants chez les femmes et les usagers les plus aisés.

Mots clés : Congestion ; Confort ; Fiabilité ; Transports publics ; Microéconomie ; Microéconométrie.
