

**Delft University of Technology**

**Faculty of Civil Engineering and Geosciences**

**Master of Science in Civil Engineering- Transport and Planning**

**Congestion pricing schemes controlled by the gMFD: a  
comprehensive design and appraisal to bridge the engineering  
and economic perspective**



**Michele Simoni**

**November 2013**



*Master thesis:*

**Congestion pricing schemes controlled by the gMFD: a comprehensive design and appraisal to bridge the engineering and economic perspective**

*Date:*

29<sup>th</sup> November, 2013

*Author*

Michele D. Simoni

4185390

Contact: [M.D.Simoni@student.tudelft.nl](mailto:M.D.Simoni@student.tudelft.nl)

*Thesis supervisor*

Prof. dr. ir. S.P. Hoogendoorn

Traffic Flow Theory and Simulation

Faculty of Civil Engineering and Geosciences – TU Delft

*Daily supervisor*

Dr.ir. A.J. Pel

Transport Planning

Faculty of Civil Engineering and Geosciences – TU Delft

*External supervisor*

Prof.dr. E.T. Verhoef

Spatial economics

Department of Spatial Economics – Vrije Universiteit Amsterdam

*Members thesis committee*

Ir. P.B.L. Wiggeraad

Operations & Management of Transport Systems

Faculty of Civil Engineering and Geosciences – TU Delft



## **Preface**

I would like to show my gratitude to all the people who have contributed to this thesis. Thanks to them, this research study has become an interesting and challenging opportunity of growth.

First and foremost, I owe my deepest gratitude to my daily supervisor, Adam Pel, for his constant and deeply appreciated advice and help. His comments and suggestions guided me through my research and provided me with veritable food for thought.

I would like to thank all the members of my committee, particularly Serge Hoogendoorn for his encouragement and support, and Erik Verhoef for his valuable feedback.

During my graduation I had the opportunity to visit the Institute for Transport Planning and Systems (IVT) affiliated with the Swiss Federal Institute of Technology Zurich. I am grateful to Kay Axhausen who made my stay possible and Rashid Waraich and Christoph Dobler who provided me with assistance in setting up my experiments. A sincere thank you also to all my colleagues at IVT for their helpfulness and the nice time we spent together.

In addition, a thank you to Nan Zheng for his politeness and promptness in giving me further explanations about his work that inspired this thesis study.

To my parents, thank you for your unconditional support and unwavering belief in me. This thesis would not have been possible without you.

Michele Simoni, November 2013



## Abstract

Charging drivers for the congestion they cause is a well-known concept among economists, traffic engineers and transport professionals. Many studies have been performed to explore and better address this issue during the history of transport economics and in recent years the first congestion pricing schemes have been finally implemented in some cities (Singapore, London, Stockholm). Nonetheless, several theoretical and practical problems still restrain this measure from being widely adopted. The lack of consistency with the traffic engineering perspective and limited degree of realism represent the main theoretical limitations of traditional congestion pricing models. The public resistance and fair distribution of gains and losses are the major practical barriers to the implementation of congestion pricing schemes.

In this thesis, we discuss congestion pricing models controlled by the gMFD within agent-based simulation (MATSim). With this approach we seek efficient tolling schemes from an economic and traffic perspective. Furthermore, by means of investigations of the distributional effects of the proposed schemes we look for the strategy that maximizes the social welfare and with the highest potential acceptability.

In the first phase of this study, through a review of theoretical studies and real experiences we identify the main reasons behind and factors influencing public resistance to the introduction of congestion pricing measures. Particularly, the question of equity is discussed in order to derive the necessary background for the evaluation of distributional effects. A thorough review of the major congestion pricing models proposed to date is made to construct our model. The survey goes from the traditional approaches like Marginal Cost Pricing and the bottleneck model to the most recent approaches based on macroscopic traffic models. During the review we discuss the main advantages and drawbacks of each approach. In the end, we describe the concept of Macroscopic Fundamental Diagram (MFD) and introduce important issues like the distribution of traffic and the phenomenon of hysteresis.

The experiment is set up with MATSim to study the metropolitan area of Zurich. First, we test the consistency of simulation outputs with the traffic flow theory by playing with different parameters of the traffic model in order to obtain the most reliable result. Then, we derive and explore the main macroscopic traffic relations of the network defined by the cordon. This analysis shows the influence of distribution of traffic, expressed by the spatial spread of density, on the performance of the network. Furthermore, this property can be associated to the phenomenon of hysteresis loops. In order to account for this additional complexity, a generalized macroscopic fundamental diagram (gMFD) is experimentally derived to express the relationship between accumulation, production and spread of density. Finally, three tolling schemes differently controlled by the gMFD are designed: a uniform toll (Flat Toll) that allows the system working below the critical accumulation threshold; a time-varying toll (Step Toll) that charges drivers according to the total delay inside the cordon estimated through

the gMFD; the Spread Toll that explicitly considers the question of spatial distribution of traffic inside the cordon. In particular, the Step Toll applies a charge in order to completely eliminate delay, whereas the Spread Toll penalizes users only for actual delays due to the overall increase of demand rather than clustering of congestion.

By integrating different aspects like the traffic performance improvements, travel behavior responses and economic impacts we carry out a comprehensive evaluation of the schemes. In this broad appraisal framework the typical traffic enhancement measurements (heaviness of congestion, delays etc.) are combined with economic indicators such as the variation of agents' utility. An analysis of travel behavior responses is carried out to identify the major trends associated with the tolling schemes and possible relationships with the traffic and economic impacts. Different socio-demographic groups are defined in order to evaluate the distributional impacts on inhabitants of different neighborhoods and areas, and on different trip purposes (work, home, education, shopping and leisure). Finally, considerations on the potential levels of public acceptance of the schemes are made on the basis of some indicators such as the share of winners, the average gain and loss, and the benefit-cost ratio.

The proposed schemes have shown to perform differently from each other. While the Step Toll and the Flat Toll determine comparable traffic enhancements, the first one outperforms the second one in terms of economic impacts and distributional effects. The smoother adaptation of demand determined by the Step Toll is probably its strong point. The Spread Toll, which was conceived as the fairest scheme, does not produce any considerable traffic improvement and it is even detrimental from the economic perspective. As to the question of public acceptance, although the slightly lower share of winners, the Step Toll determines higher gains than the other two schemes. Hence, from a social welfare perspective (combination of efficiency and distributional impacts), the Step Toll is the best performing scheme, followed by the Flat Toll and Spread Toll.

Besides these results, the investigations of macroscopic traffic properties of networks have shed light on important aspects of the traffic flow theory. First, the distribution of congestion expressed as spatial spread of density has considerable influence on the network performance. This property could be modeled in heterogeneous networks by introducing the spread as additional variable in the gMFD. Second, hysteresis loops in the MFD plane occur in presence of congestion and their pattern can be identified with the spread. Furthermore, the decrease of performance seems to be related to the frequency of loading-unloading cycles.

# Table of contents

Abstract.....	iii
Table of contents .....	1
Notation .....	5
1. Research background.....	9
1.1 Traditional congestion pricing methodologies .....	9
1.2 Innovative approaches consistent with the dynamic properties of congestion.....	11
1.3 Problem definition .....	13
1.4 Research objectives and main contributions.....	15
1.5 Research Questions .....	16
1.6 Research Framework .....	17
1.6.1 Development of a theoretical framework .....	17
1.6.2 Design of alternative pricing schemes .....	18
1.6.3 Comprehensive analysis and comparison of different schemes .....	19
1.6.4 Evaluation and policy implications.....	19
1.7 Structure of the report.....	20
2. Public acceptance and equity concerns of congestion pricing .....	21
2.1 Reasons for public resistance: lessons learnt from past experiences .....	22
2.1.1 Causes behind lack of public support .....	22
2.1.2 The importance of the political context.....	24
2.2 Factors influencing public acceptance.....	25
2.3 Proposed solutions to achieve public acceptance .....	27
2.4 The concept of equity .....	29
2.5 Evaluation of equity .....	30
2.5.1 Criteria.....	30
2.5.2 Measures.....	31
2.6 Previous research on distributional impacts: potential winners and losers.....	34
2.7 Conclusions and recommendations for the present study.....	36
3. Congestion Pricing Models .....	39
3.1 Standard Demand-Supply Models .....	40
3.2 First-Best Pricing and Second-Best Pricing.....	44
3.2.1 Facility-based schemes .....	44
3.2.2 Cordons .....	45

3.2.3	Zonal schemes.....	45
3.2.4	Considerations about area-based and cordon-based schemes.....	45
3.3	The Bottleneck Model.....	46
3.3.1	No-toll equilibrium.....	47
3.3.2	Optimal toll .....	48
3.3.3	Extensions .....	48
3.4	The concept of hypercongestion .....	50
3.5	Macroscopic Traffic Models.....	52
3.6	Traffic simulation models.....	55
3.7	Tolls controlled by the MFD within agent-based models .....	58
3.8	The ideal model (and this thesis).....	59
4.	The Macroscopic Fundamental Diagram .....	63
4.1	The MFD in a nutshell .....	63
4.2	Original formulation of the MFD.....	65
4.3	Further developments of the theory .....	66
4.3.1	The empirical evidence of the MFD .....	66
4.3.2	The demonstration of the existence of the MFD.....	67
4.4	Hysteresis loops and network dynamics.....	68
4.5	The influence of traffic distribution on the network performance .....	70
4.6	This research study .....	72
5.	Set-up of the experiment and model framework.....	73
5.1	Overview of MATSim: a Multi-Agent Transport Simulation .....	73
5.1.1	Development of plans.....	74
5.1.2	The utility function.....	75
5.1.3	JDEQSim: a queue-based traffic simulator .....	78
5.2	Case Study Scenario .....	79
5.3	Derivation of the macroscopic variables .....	80
5.4	Consistency of MATSim with traffic flow theory .....	81
5.5	Derivation of the MFD for the city centre of Zurich .....	83
5.6	Investigations of the macroscopic traffic relations.....	88
5.7	Identification of a Generalized Macroscopic Fundamental Diagram .....	93
5.8	Development of three alternative tolling schemes based on the (g)MFD .....	95
5.8.1	The Flat Toll.....	96
5.8.2	The Step Toll.....	96

5.8.3 The Spread Toll.....	99
5.8.4 Final fares.....	101
6. Comprehensive evaluation of tolling schemes.....	103
6.1 Resulting production-accumulation relationships.....	104
6.2 Traffic performance enhancements .....	108
6.2.1 Traffic improvements for the whole day .....	108
6.2.2 Traffic improvements during the peak hours .....	110
6.2.3 Considerations of the traffic performance of tolling schemes .....	112
6.3 Economic interpretation of results.....	113
6.4 Identification of socio-demographic categories .....	115
6.5 Travel behavior changes .....	118
6.5.1 Impacts on socio-demographic groups.....	118
6.5.2 Influence of socio-demographic characteristics on travel behavior changes.....	120
6.5.3 Economic impacts of travel behavior changes .....	122
6.6 Economic efficiency and distributional impacts of the schemes.....	123
6.6.1 Comparison of overall impacts deriving from the schemes .....	123
6.6.2 Impacts of tolling schemes on different socio-demographic categories.....	123
6.6.3 Relationship between socio-demographic characteristics and economic impacts.....	127
6.7 The profile of winners and losers.....	129
6.8 Public acceptability of tolling schemes.....	131
6.8.1 Overall efficiency of tolling schemes .....	131
6.8.2 Redistribution schemes.....	133
6.8.3 Final considerations about the public acceptability issue .....	135
7. Conclusions .....	137
7.1 Main findings .....	137
7.1.1 Results from the traffic flow theory perspective.....	138
7.1.2 Research questions .....	138
7.2 Conclusions .....	141
7.2.1 Design of schemes based on the gMFD .....	141
7.2.2 Efficiency, distributional impacts and acceptability of different schemes .....	142
7.2.3 Consistency with traditional congestion pricing models .....	143
7.3 Implications for practice .....	144
7.3.1 Effectiveness of cordon-based tolls and combination with traffic management.....	144
7.3.2 Tolling schemes controlled by the gMFD for complex case studies .....	145

7.3.3 A comprehensive appraisal framework as a tool for policy-makers .....	145
7.4 Implications for research .....	146
7.4.1 Recommendations for improved research .....	146
7.4.2 Recommendations for future research directions.....	147
References .....	149
Appendix I: Distribution of densities.....	155
Appendix II: Logistic regression .....	156
Appendix III: skewness and kurtosis .....	157
Appendix IV: t-tests.....	158
Appendix V: average gains/losses for different socio-demographic groups .....	159

## Notation

### Congestion pricing models variables (Chapter 3):

$\tau$  toll

$mc$  marginal social cost

$c$  costs

$mecc$  marginal external congestion cost

$V$  volume of trips

$d(V)$  inverse demand function

$c(V)$  average variable cost of a trip

$B$  total benefits

$C$  total costs

$t^*, t_1, t_2$  preferred arrival time, trip starting time, trip ending time

$\alpha, \beta, \gamma$  marginal costs (travelling, earlier arrival, later arrival)

$N$  number of users

$R(a)$  the number of travelers who have departed before time  $a$

$\rho(a)$  departure rate

$s$  bottleneck capacity

### Macroscopic Fundamental Diagram theory (Chapter 4):

$k$  density

$q$  flow

$v$  speed

$l$  (lane-km) is the total length of the network

$d$  (km) is the length of a vehicle trip

$n$  is the number of vehicles in the network

$(g)$  exit flow

### Agent-based simulation variables (Chapter 5):

$p_j$  probability for plan  $j$  to be selected

$U_j$  utility of plan  $j$

$U_{act,i}$  utility of performing an activity  $i$

$U_{travel,i}$  (dis)utility of travelling to reach activity  $i$

$U_{wait,i}$  (dis)utility of waiting for performing activity  $i$

$U_{short,i}$  (dis)utility of performing activity  $i$  too short

$f_p$  crowd factor for shopping and leisure facilities

$\beta_{perf}$  marginal utility of performing an activity

$t_{perf,k}$  time spent performing the  $k^{th}$  activity of the same type as activity  $i$

$\beta_{wait}$  marginal utility of waiting time

$t_{wait}$  and is the waiting time  $i$

$\beta_{short}$   $\beta_{short}$  being the marginal utility of another unit of time which is missing to the minimum activity duration of 0.5h

$U_{access-egress,i,mode}$  score of access/egress to/from a vehicle

$\beta_{tt,mode}$  marginal utility of another unit of time traveling by mode

$t_{mode}$  in-vehicle travel time

$\beta_{cost,mode}$  marginal utility of another unit of time traveling by mode

$c_{i,mode}$  expenditure of money required for traveling with mode

$t_{access,mode}$  time to access mode

$t_{egress,mode}$  time to egress mode

### **Macroscopic traffic variables (Chapter 5):**

$k_j$  density of the single link  $j$

$e_j$  vehicles entering the link  $j$

$o_j$  vehicles leaving the link  $j$

$l$  length of the link

$n$  number of lanes of the link

$\Delta t$  time interval

$q_j$  outflow of link  $j$

$u_j$  average speed of link  $j$

$K$  average density of the network

$Q$  average outflow of the network

$U$  average network speed

$\gamma$  spatial spread of density

$s$  average trip distance travelled inside the cordon

$\Delta d$  delay per user

$N$  number of additional users

VOT average value of time

$\tau$  toll

### **Economic evaluation variables (Chapter 6):**

$\Delta U_j$  individual change of utility

$y_j$  average daily income

$c_{toll,i}$  individual revenue

$\Delta W$  total welfare change

### **Statistical analyses variables (Chapter 6):**

$R^2$  coefficient of determination

$\beta$  regression coefficient

p-value probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true;

t-value regression coefficient divided by its standard error;



# 1. Research background

Traffic is nowadays a critical issue in many metropolitan areas and it has several negative outcomes like congestion and negative environmental impacts. Costs determined by traffic congestion amount to about 1% of GDP in countries like France, Germany and the UK (de Palma and Lindsey, 2011). The problem is even more serious in developing countries where road capacity is far too low to accommodate the rapid and steady growth of traffic volumes. Investing in new infrastructure is an expensive solution and often the phenomenon of induced demand makes this effort fruitless (in terms of levels of congestion) in few years.

The main problem lies in the fact that delays caused by individual travelers to the others are not included in the costs and hence people do not pay the full marginal costs of their trips. A traditional way to internalize the costs deriving from congestion is to charge road usage by means of congestion pricing schemes. According to Santos and Rojey (2004): “Confront the trip maker with the true social cost of his journey to ensure that only cost-justified journeys are made and the scarce road space is allocated to those for whom the value is the highest”. The main advantage of such (transport) demand management measure is that it determines changes in travelers’ behavior such as the departure time, the route and the transport mode choice.

The original concept of applying a tax on congestion introduced by Pigou (1920) has been widely accepted by the academia and several applications and extensions have been developed in order to account for theoretical and practical related concerns.

After many years of research, finally in the last decade some congestion pricing schemes have been successfully implemented in Singapore (1998), London (2003), Stockholm (2006) and Milan<sup>1</sup> (2008). Other cities such as San Francisco and Beijing are currently developing similar schemes to be adopted in the next years.

## 1.1 Traditional congestion pricing methodologies

The extensive body of literature addressing the problem of congestion pricing can be divided in two different classes: static marginal cost pricing models that assume congestion as stationary and the bottleneck models that consider it dynamically.

The aforementioned ‘Pigouvian toll formula’ has given rise to a large category of congestion models known as Marginal-Cost Pricing (MCP) models whose first theorists are Walters (1961), Vickrey (1963) and Beckman (1965). The key concept of MCP models is to charge every link of

---

<sup>1</sup> To be precise, the “Ecopass” scheme in Milano is actually a pollution charge since it has been designed based on vehicles’ emissions

the network such as the toll equals the congestion costs imposed to all the community of users by an extra traveler in order to bring the user equilibrium toward a system optimum.

The main advantage of static MCP models is that the toll is “link-based rather than path-based” (de Palma and Lindsey, 2011) i.e. there is no need of information about origins, destinations and routes as the toll is a simple function of the link.

However, this kind of models also presents some practical and theoretical pitfalls (de Palma and Lindsey, 2011; Lindsey, 2012). First, this approach is efficient only if applied to every toll of the network (first-best tolling) even though it is hard to implement and quite impractical as urban road networks have many links. Then, in reality it is more practical to toll only a limited fraction of links of the network (second-best pricing schemes) that divides in cordon-area schemes and zonal schemes. The first ones charge a fee every time a driver enters or exits a certain area delimited by a perimeter (cordon). In the second ones a toll is paid also for driving within the tolled zone without crossing its boundary.

Another major complication of static MCP models is that the VOT that determines the trip cost function is an average number influenced by factors such as users’ heterogeneity that in turns is influenced by other issues like the level of toll, the time of day etc.

Another important drawback is that congestion is assumed as constant, while in reality it varies considerably during the day, week and year. Consequently travel times can differ significantly over the time.

In order to overcome this limit, dynamic models able to describe the evolution of travel demand (demand-side) and traffic (supply-side) over time and space have been developed.

The most popular ones introduced by Vickrey (1969) and revisited and extended by Arnott *et al.* (1993) are the bottleneck models where congestion is considered as a queue at a bottleneck and vehicles are able to move at free-flow speed in absence of queue (demand-side specification). Moreover, every user who needs to pass through a bottleneck to reach his destination incurs ‘schedule delay’ costs in case of early or late arrival. Then, during the peak hours, when demand is higher than capacity, users have to choice between suffering longer delays to arrive on time or travelling earlier or later (supply-side specification). Hence, all the travelers choose the “departure time to minimize their generalized trip cost which includes vehicle operating costs, travel time, schedule delay costs and tolls (if any)” (De Palma and Lindsey, 2011, pp 1379). The equilibrium is reached when no one can reduce his costs by changing departure time.

Different studies have extended the original model by Vickrey and analyzed various aspects such as the equilibrium order of arrivals (Daganzo, 1985), the heterogeneity of users (Newell, 1987), the elasticity of demand (Arnott *et al.*, 1990) and different toll policies such as flat tolls that do not change during the time and ‘fine toll’ that vary continuously.

The main pitfall of the traditional models presented thus far is that traffic systems are considered stationary: the performance of the network in terms of average flow and output rate is stable, while in reality traffic systems behave differently according to their levels of congestion. According to Geroliminis and Levinson (2009):

“the traditional network supply curve (desired or input demand vs. average travel cost) is not consistent with the physics of traffic. This is because for a given average flow (i.e. desired demand over a period of time) the total cost (expressed in delay terms) (i) is sensitive, during congested conditions, to small variations of flow within given period and (ii) depends on the initial state of the system (the level of congestion)” (pp 221)

Consequently the estimated toll may not be optimal and the network may remain still congested.

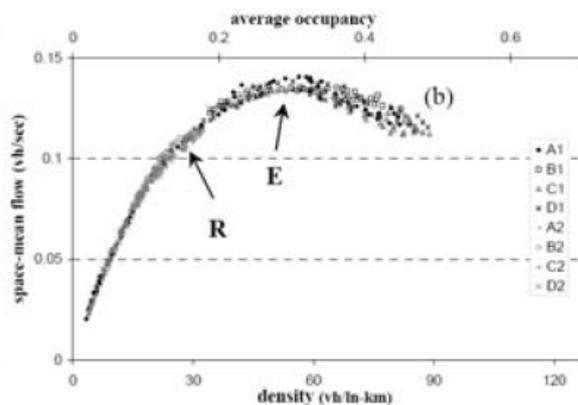
## **1.2 Innovative approaches consistent with the dynamic properties of congestion**

An important contribution to the challenge of modeling congestion pricing schemes came in recent years mainly from scholars like Daganzo (2012), Geroliminis (2009; 2012) and colleagues, who combined the bottleneck model with macroscopic traffic models. The main innovation of this paper is to model the supply of an hypothetical city centre network by means of a Macroscopic Fundamental Diagram (MFD). These models are macroscopic for both the demand-side and supply-side.

The MFD relates space-mean flow, density and speed of an entire network of several links as Figure 1 shows. The MFD is an intrinsic property of a network itself and it exhibits a clear maximum (point E in Figure 1.1). Variations of demand do not affect either the shape or the maximum of the MFD (except for large variations in the O-D pattern determined by a single event or evacuation). Other important characteristics are the independence of the space-mean flow from the O-D patterns, the average trip length that is constant with the time and the ease of derivation from existing technologies (detectors, GPS). Studies by Daganzo (2007) and Geroliminis and Daganzo (2007, 2008) showed that urban areas where neighborhoods are uniformly congested have the same properties and that it is possible to derive their respective MFD as well. The concept of MFD can be mathematically expressed as a mass conservation

equation that governs the state of the system. It is possible to conclude that traffic systems are dynamic and in order to estimate their state it is necessary not only to know the vehicles entering the network (inflow), but also the prior state of the system.

Geroliminis and Levinson (2009) develop cordon-pricing schemes considering the network as a single bottleneck with a capacity depending on the traffic state (convex function represented by the MFD).



Two alternative tolls are set to investigate the issue of equity by exploring the effects of allowing lower or higher inflows in the system. Indeed, operating at maximum capacity allows high number of users with non-optimal performances, whereas operating at free-flow conditions allows better performances for fewer users.

Figure 1.1: MFD of the city centre of Yokohama (source: Geroliminis and Daganzo, 2008)

Gonzales and Daganzo (2012) included the presence of public transport service sharing street space with cars, but operating independently of traffic conditions (comparable to dedicated transit lanes) to develop an optimal toll. This study includes users' heterogeneity only in wished travel time and user groups (choice and captive transit riders).

A further step forward has been made by Zheng *et al.* (2012) who developed a cordon-based pricing scheme in an agent-based model (MATSim) controlled by the MFD.

Agent-based models allow accounting factors such as the users' heterogeneity and the elasticity of demand better than the traditional macroscopic demand-supply models. Indeed, in agent-based models theoretically every user can have his own utility function and value of time and being able to affect other users' decisions with his trip choices. Moreover, users can decide not only to change their departure time and mean of transport, but also not to travel (for example if the toll or the congestion is too high).

The authors, after proving that MATSim is able to reproduce a MFD for the city centre of Zurich, they create a pricing scheme controlled by the MFD by means of an 'off-line' linear feedback control process. Finally an investigation of behavioral shifts of work-related and non-work related activities is carried out.

### **1.3 Problem definition**

Benefits of congestion charge schemes in terms of net welfare gains have been widely recognized by the scientific community and policy-makers. Moreover, thus far, the experiences in cities where congestion charges have been introduced revealed successful in reducing traffic volumes (from 10% to 30% according to the Environmental Defense Fund) and other traffic externalities like pollution and traffic accidents.

However, public acceptance of such measures has been often very low and many political and public issues arose when implemented or even proposed. Public acceptance represented the main barrier for the implementation of congestion pricing projects in Manchester, Edinburgh and the Netherlands where a state-of-the-art national pricing scheme was developed. Even if such policy measure may be in principle beneficial, the population and political parties usually boycott it. Public and political acceptability is therefore a “*conditio sine qua non*” for the success of road pricing policies.

According to many scholars (Jones, 1998; Lindsey and Verhoef, 2000; Viegas, 2001) a significant reason of resistance can be identified in the negative impacts on equity of this policy measure. Indeed, even if congestion pricing can generate an overall net welfare surplus, some groups may result disadvantaged. In addition, the redistribution of welfare among different groups may be so uneven that the benefits derived from pricing schemes can be reduced and put aside (Eliasson and Mattson, 2006).

There are two major typologies of transportation equity: the horizontal equity concerning the distribution of impacts between individuals or groups in a similar situation (same needs and abilities) and the vertical equity concerning the distribution of impacts between individuals or groups with different abilities and needs (e.g. income, social class). In similar way Jones (2002) divides equity concerns in two main categories: those looking at distribution of costs and benefits across socioeconomic groups (vertical equity) and those looking at the effects on different geographical groups (horizontal equity).

Vickrey who is considered “father of Congestion Pricing” (VPTI, 2013) proposed some major guidelines to implement efficient congestion pricing schemes. Among them, he advocated the priority of equity concerns by providing benefits to all user groups and he encouraged the development of transport alternatives like public transport. Several studies (Santos and Rojey, 2004; Levinson, 2005; Ubbels and Verhoef, 2006) and experiences (such as toll roads in Norway) have shown the strong linkage between public acceptability of road pricing schemes and the use and distribution of revenues.

The welfare economics approach based on the work of Pareto recognizes the difference between efficiency and distributional aspect of policy measures. The first is concerned with the optimal production and allocation of resources, while the second is concerned with the distribution of resources throughout the society. Maximizing society's welfare intended as the combination of efficiency and equity is hard task since often the two criteria are in conflict with each other. Indeed, efforts in redistributing benefits and costs typically reduce the overall gains. The trade-off between efficiency and equity is a veritable leitmotif in economics that also applies to congestion pricing schemes. Efficiency in transportation studies corresponds to the policy's ability to maximize aggregate social welfare regardless of whether some individuals are worse off (Ecola and Light, 2009). On the other hand, as previously mentioned, congestion pricing schemes determine different effects across some population and public acceptability often depends on the distribution (and perception) of gains and losses deriving from the policy measure. Hence, when developing congestion pricing schemes it is fundamental not only to assess their efficiency, but also to investigate their impacts on different socio-demographic groups in order to guarantee the public and political acceptance necessary to implement them.

The aforementioned studies that developed models consistent with the dynamic characteristics of traffic have demonstrated to be effective from both the engineering and economic perspective as clear benefits in terms of traffic flows and travel-time savings have been shown. However, equity issues have been only marginally explored.

Geroliminis and Levinson (2009) examined the effects of optimal tolls derived from different values of maximum inflow. A trade-off between higher performances of network with lower flows and vice versa can be made by choosing the state at which the MFD should operate (free-flow or capacity regime). A traffic system working at capacity regime allows the largest number of users but with lower average speeds compared with a traffic system working at free-flow regime. This is of course a political choice between a more efficient or a more equitable traffic system. Due to the nature of the model itself that was based on the combination of two macroscopic models for both the demand and supply it was not possible to make deeper investigations of the effects of these different policies on individual travelers.

Zheng *et al.* (2012) thanks to introduction of an agent-based model in the pricing scheme have been able to analyze the impact of pricing on shifting behavior of two different categories according to the trip purpose (work and non-work related).

Since economic and engineering efficiencies are necessary, but not sufficient conditions to implement congestion pricing schemes, deeper investigations in order to compare the total

distributional effects with the degree of efficiency and welfare surplus are needed to fully assess models coordinated by the MFD.

#### **1.4 Research objectives and main contributions**

The main objective of the research project is to develop and assess congestion pricing schemes controlled by MFD within agent-based transport models that aim to maximize social welfare. Following from Levinson (2010), in this study social welfare will be considered as the combination of: “efficiency, a measure of the degree to which system outputs achieve a theoretical maximum using the same level of inputs, and equity, a measure of the distribution of outputs across some population” (p 33). The integration of the traffic flow theory with (transport) economics principles is conceived to reach traffic and economic efficiency. The agent-based framework is chosen to investigate in depth the distributional effects and identify possible weak points of the schemes in terms of public acceptability.

In order to do that, a broad analysis aimed at evaluating not only the efficiency from the traffic operations and economic perspective, but also the distributional impacts of tolling schemes on different socioeconomic and spatial groups will be carried out.

Moreover, a set of alternative pricing schemes characterized by different conceptual approaches will be developed and compared with each others in order to determine the one that maximizes total welfare. During the design phase, the engineering and economic perspective will be jointly considered in order to define optimal schemes in terms of efficiency and equity. A basic uniform toll (Flat Toll), comparable to the one implemented by Zheng *et al.* (2012) will be implemented. Then, alternative schemes characterized by a time-varying charge will be implemented. For example, a Step Toll, which changes in discrete time intervals, may be more beneficial thanks to its higher flexibility and ability to flatten the peak demand more smoothly. In order to explicitly account for the property of spatial distribution of congestion, an alternative time-varying toll defined as Spread Toll will be implemented.

At the practical level, this research project is intended to contribute to a higher awareness of congestion pricing schemes controlled by the MFD by investigating their effects from several perspectives. On the one hand their direct impacts on the traffic performance of the will be analyzed by means of macroscopic traffic indicators. On the other hand their broader economic impacts will be explored by looking at welfare impact indicators like change of utility of the agents. Furthermore, the identification of possible winners and losers among different socio-demographic categories might help policy-makers to better address the problem of public and political acceptability. This kind of approach can be relevant in European countries like the

Netherlands where distribution of welfare has a relatively high weight in the evaluation of policies.

At the theoretical level this research project aims to improve the original scheme characterized by a fixed congestion charge (Flat Toll) into more time dynamic ones (Step Toll and Spread Toll). The development of the alternative tolling schemes will be done on the basis of traffic and economic theories such as the marginal cost pricing approach. A thorough investigation of the nature of the MFD will provide a better understanding of this innovative concept. Particularly, the analyses of macroscopic traffic properties like the spatial spread of density will contribute to understand the nature of congestion at network level and lead to derivation of a generalized MFD (gMFD). Furthermore, the implementation of tolling schemes controlled by macroscopic traffic indicators within a multi-agent transport model could provide further insight into the potentialities of this planning tool and its consistency with the traffic flow theory. Finally, the investigation of short-term travel behavior changes such as the route choice, the mode choice and the departure time choice could provide additional knowledge of the impacts of cordon-based tolls.

## 1.5 Research Questions

In relation to the objectives of the research the following main research question has been formulated:

- **How do cordon-based congestion pricing schemes governed by the macroscopic fundamental diagram address social welfare?**

In order to answer the first main research question the following set of sub-questions need to be answered:

- **Do tolling schemes differently controlled by the MFD determine comparable improvements from the traffic efficiency perspective? If not, which is the most and least efficient?**
- **Are tolling schemes differently controlled by the MFD characterized by similar trends in travel behavior changes?**
- **Do tolling schemes differently controlled by the MFD determine comparable improvements from the economic welfare perspective? If not, which is the most and least efficient?**
- **Are tolling schemes differently controlled by the MFD characterized by similar distributional effects? If not, is there any socio-demographic economic category**

**that results particularly advantaged or disadvantaged by a specific tolling scheme?**

- **Do the tolling schemes differently controlled by the MFD achieve the same levels of potential acceptability?**

## **1.6 Research Framework**

The result of this graduation project will consist in the development and assessment of cordon-pricing schemes coordinated by the MFD from a social welfare perspective. In this study, the social welfare is identified as a combination of efficiency and balanced distribution of impacts. In order to arrive to this result, the following activities will need to be performed as shown in the conceptual framework of Figure 1.2.

### **1.6.1 Development of a theoretical framework**

An extensive survey of models developed to represent congestion pricing schemes in the past is necessary to fully understand the importance of more innovative models controlled by the MFD. At this point, strong points and weak points of traditional congestion charge schemes could be identified. Moreover, a particular attention to second-best congestion pricing schemes will be paid in order to develop the necessary background to design efficient time-varying schemes.

A solid knowledge of the physics of traffic and specifically of the MFD is fundamental to comprehend and control the 'original scheme' and subsequently design alternative ones characterized by different design principles. In particular the development of flexible schemes aimed at internalizing the delays in the cordon will require a deep understanding of the macroscopic traffic properties of the network. For this reason, the main concepts and the major issues related to this topic will be explored by means of a literature review.

In order to carry out analyses and develop alternative schemes it is also necessary to understand the theoretical basis of agent based transport models like MATSim. Particularly, it is important to understand how behavioural changes such as route choice, time departure and mode choice are addressed. The previous studies accomplished with MATSim represent a starting point to acquire a good knowledge of the model.

An analysis of the unsuccessful experiences of congestion pricing in Edinburgh, Manchester and New York City will allow a deeper understanding of the question of acceptance of pricing and the connection with equity.

Then an extensive review of studies focused on the evaluation of distributional effects of congestion pricing based on simulation and real case studies will be made in order to develop the necessary background for the assessment of tolling schemes. Here, the major typologies of

inequity and winners and losers of congestion pricing schemes will be identified and different methodologies applied to evaluate the impacts of tolls on different categories of people (income class, geographic location, mode) will be investigated.

### **1.6.2 Design of alternative pricing schemes**

MATSim is a simulation program in constant evolution. For example, in the last two years (2011-2013) significant changes have been made in terms of road network, traffic models and additional features (alternative transport modes, facilities). Given these circumstances, even if the case study will be similar to the Zurich Scenario studied by Zheng *et al.* (2012), part of the research will need to be devoted to the setup of the experiments with the current version of MATSim.

In order to do that, the following necessary steps to derive a MFD for the city centre of Zurich need to be accomplished. First, the case study scenario (Zurich metropolitan area) will be set up with MATSim in order to derive a “reliable” starting no-toll scenario. Then, algorithms to estimate traffic conditions such as density, flow and speed of single links of the network need to be implemented. Results will be examined to check the consistency with the dynamics of traffic and respect the fundamental relationship of traffic flows. After that, data will be aggregated such that the macroscopic relationship of traffic flows can be identified. It is important at this stage to provide a sound proof of the ability of the model to reproduce the MFD before proceeding with the implementation of pricing schemes. This part of the research will also involve a deeper understanding of the ability of agent-based models to simulate traffic and reproduce reliable traffic outputs.

Once the setup of a reliable “scenario” is achieved, the “basic” Flat Toll scheme originally proposed by Zheng *et al.* (2012) can be finally implemented to allow the network to operate below the critical values of accumulation. Only in a second moment, it will be possible to design alternative tolling schemes. A first alternative scheme would consist of a toll that varies during discrete time intervals (Step Toll) according to the level of demand in order to eliminate delays. The main rationale of this scheme is that people should be charged for the delay they create by travelling inside the cordon, following from the traditional marginal cost pricing approach. Such kind of scheme, thanks to its higher flexibility, might determine higher benefits than the uniform toll. The optimal fare could be derived from the MFD of the network with the same linear-feedback control process used in the original scheme. A second alternative scheme would consist of a time-varying toll that also accounts for another macroscopic property of the network: the spatial spread of density. This tolling scheme (Spread Toll) would explicitly consider the fact that the decrease of performance inside the cordon might be partly determined by inner clusters of congestion.

### **1.6.3 Comprehensive analysis and comparison of different schemes**

Once the different schemes have been designed, a comprehensive analysis of their performance can be carried out to assess their efficiency, their distributional impacts and their potential levels of acceptability. On the basis of the previous literature review, some specific socio-demographic categories will be identified to investigate the impacts of tolls. First, an investigation of the effects of the tolling schemes on the traffic performance of the network will be conducted by means of different traffic indicators such as decrease of delay, reduction of queues and increase of traffic efficiency. Then, an investigation of the benefits and losses deriving from the schemes will allow a general evaluation of the economic efficiency. Gains and drawbacks can be calculated at aggregate and disaggregated level. Indeed, thanks to multi-agent micro-simulations like MATSim it is relatively straightforward to make economic evaluations deduced by individual utility changes. Furthermore, a more detailed analysis of the effects on previously identified socio-demographic groups will describe the distributional effects of the schemes and identify the most progressive/regressive ones. Changes in short-term travel behavior such route, mode and time departure choice will be examined as well to provide a more complete overview. Finally, thanks to all the previous analyses, considerations about their potential level of acceptability will be made.

### **1.6.4 Evaluation and policy implications**

In this conclusive stage of the research project, it is finally possible to answer the research questions. The conclusions will be drawn from the results of previous analyses and a comprehensive appraisal of the alternative tolling schemes will be made.

Thanks to the achieved results policy implications concerning the distributional impacts and the potential acceptability of the analyzed pricing schemes can be discussed and extended into more general theoretical and practical indications.

To conclude, some comments regarding the implementation of cordon-pricing schemes regulated by macroscopic traffic models and recommendations for further research can be provided.

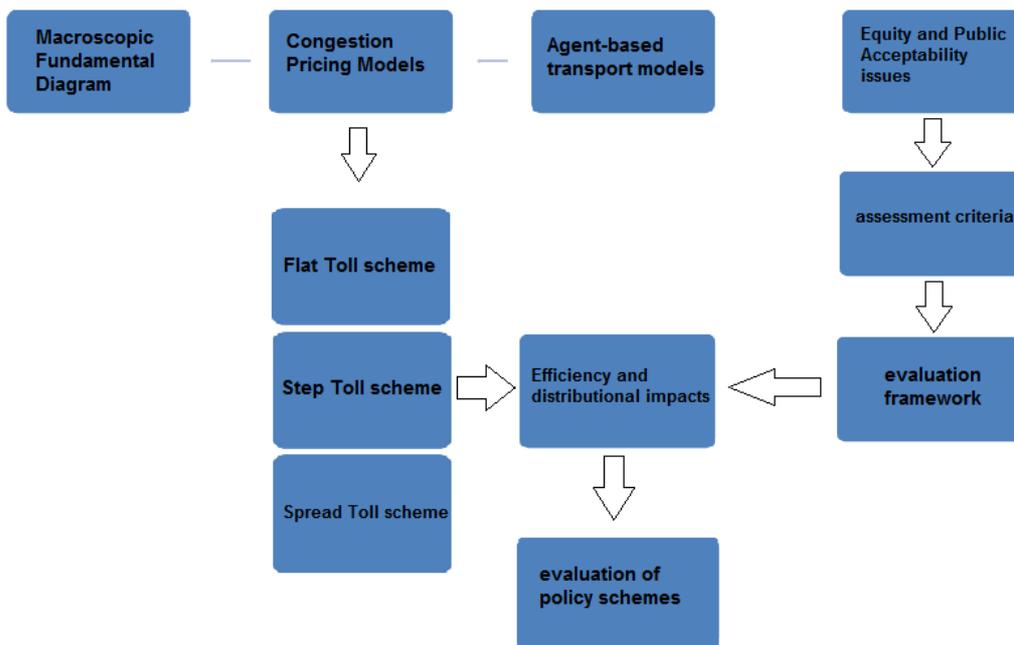


Figure 1.2: research framework flowchart

## 1.7 Structure of the report

This study is structured as it follows: the theoretical background is developed in the first part of the report, while the design and appraisal of the schemes is described in the second one. The topics of public acceptance and equity concerns are described in Chapter 2. An overview of the main congestion pricing models developed thus far is provided in Chapter 3. In Chapter 4 the concept of the MFD and the main related issues are illustrated through a literature review of the topic. Then, Chapter 5 focuses on the experiment setup and the design of tolling schemes. In Chapter 6 a series of analyses are performed in order to allow a broad evaluation of the schemes and considerations about the public acceptability of the schemes. Finally, the findings of this study and recommendations will be made in Chapter 7.

## **2. Public acceptance and equity concerns of congestion pricing**

The concept of charging road users for the congestion they create with their travels has been introduced more than fifty years ago and it has been widely accepted by the economic community. However, to date only few congestion pricing schemes have been successfully implemented worldwide: Singapore, London, Stockholm and Norway's toll rings. Lack of effectiveness and efficiency, public and political acceptability, and equity concerns represent the main barriers to the introduction of such a measure.

Public acceptance plays a crucial role in the success of those transport policies measures that strongly affect users' behavior and lifestyle such as congestion pricing schemes. Hence, according to Mayeres and Proost (2001) public and political acceptability of any pricing or taxation reform are strongly dependent on the transparency of deriving economic benefits to the majority of people. A pricing reform can be accepted only if fulfilling this condition that is the "essence of the economic approach to acceptability" (Mayeres and Proost, 2). Viegas (2001) as well stresses the importance of the perception of fairness to achieve public acceptance.

However, economic reforms usually determine benefits and losses of different extents across the population. This aspect, also known as "equity issue", plays an important role in the acceptability process at least for two reasons according to Eliasson and Mattsson (2006). First, the distribution of impacts may be so uneven that it dwarfs the overall benefits of the scheme and even questions the validity of the schemes. Second, congestion pricing may determine higher negative effects to specific groups like lower incomes and car commuters from the outskirts or simply people with poor public transport availability. Equity concerns become even more crucial for acceptability of policies in those countries where the welfare of disadvantaged has relatively high weight in the evaluation, following from the Equality Principle by Rawls (1971) according to whom policies not only should provide equal opportunities to individuals, but also should favour the worst-off when inequalities occur.

When advocating the use of charges as a measure to fight congestion, economists' focus on congestion pricing has been mainly on the objective of efficiency, for which "it is only the sum of net benefits –not the distribution- that counts" (Verhoef *et al.*, 272). However, as Levinson (2010) highlights, social welfare is given by both efficiency and equity, where efficiency is defined as "measure of the degree of which the system outputs achieve a theoretical max using the same level of inputs" and equity as "distributional effects". Eliasson and Mattsson (2006) observed that while a considerable body of literature about congestion pricing has been produced till now, the equity issue has been only marginally explored and in a very theoretical fashion: "studying the effects analytically or by numerical simulations with a well-specified but highly

simplified model of a real transport system” (p. 604). Hence, despite lack of public approval may stem from several factors (e.g. distrust in public authorities, skepticism about the effectiveness of the scheme et cetera), it would be advisable when designing congestion pricing schemes to evaluate their distributional impacts as well.

Carrying out a thorough investigation of the problem of public acceptance and identifying solutions to tackle it go beyond the scope of this research. Similarly, providing a full overview of the equity issue from all its different perspectives is not the goal of this study. Rather, the aim of this research is to approach these topics in order to introduce later public acceptance and equity considerations in the evaluation of road pricing schemes implemented with multi-agent based transport models controlled by the MFD.

Hence, in this chapter the questions of public acceptance and equity will be addressed as it follows. In the first section (Section 2.1), the main reasons behind public resistance to congestion pricing schemes will be determined with the help of ex-post evaluation studies of the (unsuccessful) experiences of Edinburgh, Manchester and New York City. Then, after an overview of factors influencing public support (Section 2.2), some of the solutions proposed by major theorists of congestion pricing will be presented (Section 2.3). In Section 2.4 the concept of equity and its meaning in this research will be introduced. Criteria and quantifying measures of equity will be shortly described in Section 2.5. The potential winners and losers deriving from congestion pricing schemes will be identified through previous research studies (Section 2.6). Finally, some final considerations and recommendations will be made in Section 2.7.

## **2.1 Reasons for public resistance: lessons learnt from past experiences**

The recent failures of tolls with proven economic and environmental benefits in cities like Edinburgh, Manchester and NYC have demonstrated that public acceptance represents a major barrier to the introduction of congestion pricing schemes. The recognition that public acceptability plays a key role in the implementation of congestion charges raises several issues about the reasons behind people’s resistance, the main influencing factors and possible solutions to tackle the problem.

### **2.1.1 Causes behind lack of public support**

One of the main causes of lack of public support of congestion charges consists in the little trust about the commitment of local authorities and institutions in providing benefits to people affected by the toll. In several cases people distrusted promises of reinvestment of revenues in projects or initiatives potentially beneficial to them (e.g. public transport improvements). Instead, tolls were often perceived as a trick of the government to use road users as “cash cows”. This is the case of Manchester where officials’ pledge to spend revenues in road and transit

improvements (some of them already promised but never realized before) was seen very skeptically. Similarly, a survey of Edinburgh residents after the pricing referendum revealed that many voters doubted about the commitment of local government in investing toll revenues in public transport improvements. Opponents in New York City argued that rebates would have not been permanent and that the federal government would have reduced financial support to local authorities as a consequence of the improved budget (Schaller, 2010). This issue, also defined as “credible commitment problem” (Manville and King, 2012), is clearly an intrinsic feature of this policy measure since congestion charge requires individuals to believe the collectors and if “this disbelief cannot be ameliorated, the agreement won’t be reached, and any gains from agreement will be lost” (p. 230).

Another important cause of public resistance lies in a wrong perception of congestion pricing. As Pridmore and Miola (2011) suggest in a recent report, acceptance of transport policy measures can be significantly improved when “the public is aware of the negative impacts [of car travels] and they understand the need for measures to address these impacts” (p. 8). However, in the case of congestion charges, as Jones (1998) observes, it is difficult to convince the drivers to accept the concept of paying for the external costs (congestion) associated with their trips. Indeed, roads are considered as a public good that should be available free of charge. Furthermore, congestion charge is generally not perceived as an effective measure to tackle congestion, especially where public transport does not represent a valid alternative to car and the elasticity to tolls is estimated to be low (Jones, 1998). One of the main arguments of the opponents (especially car users) of Edinburgh congestion charge was the disbelief of benefits deriving from such a measure (Gaunt et al., 2007). In New York City as well, opponents argued that a charge would have produced little improvement of congestion in Manhattan that was rather determined by taxis and trucks (Schaller, 2010).

Several studies have demonstrated that inadequate awareness of the effectiveness of congestion charge was often determined by a lack of understanding. This is the reason why high complexity has a negative impact on public acceptance. For example, residents of Edinburgh opposed the proposed pricing scheme also because its operating principles were not clear enough or at least not well explained (Gaunt et al., 2007). On the contrary, London Congestion Charge consisting of a single cordon-based flat toll during the entire day, even if “not economically optimal (in academic terms)” (Pridmore and Miola, 2011) was very understandable. This view is in contrast with Verhoef (2006) according to whom the complexity of the measure does not affect levels of acceptance. Then, policy makers might adopt more complex pricing schemes in order to achieve better performances (i.e. differentiated pricing instead of flat tolls) without losing public support.

Finally, congestion charge is considered unfair. The impacts of urban road pricing may vary significantly across different socio-economic and geographic groups. Consequently categories that feel disadvantaged might oppose it actively. This is the case of car-users in Edinburgh and New York City who firmly opposed the introduction of the schemes since they regarded potential gains in terms of travel time savings as questionable and not worthy of the fee. Furthermore, a charge that affects especially commuters from the outer boroughs is likely to be seen as unfair. Indeed, in Edinburgh 3 out of 4 local authorities of the outer boroughs fiercely opposed the introduction of the proposed pricing schemes because of its apparent unfairness in terms of location and functioning (Laird et al., 2007). In NYC, commuters from the periphery represented the main opponents of the proposed toll schemes. It is worth mentioning that even relatively small groups of the population (for example only 5% of the commuters to NYC would have been affected by the toll) are able to prevent the implementation of these schemes. As Table 2.1 shows, the main factors behind public resistance in the cities of Edinburgh, Manchester and NYC can be identified as lack of understanding and distrust in the effectiveness and fairness of the scheme (towards the entire community or regarding specific groups). However, also specific political situations, as it will be briefly explained in the following section, need to be considered when evaluating the success (or failure) of congestion pricing schemes in terms of public acceptance.

	Edinburgh	Manchester	NYC
<b>Distrust in revenue reinvestment</b>		✓	✓
<b>Distrust in effectiveness of the scheme</b>			✓
<b>Lack of understanding</b>	✓		
<b>Perception of unfairness</b>	✓	?	✓

**Table 2.1: major causes behind public resistance in Manchester, Edinburgh and New York City**

### 2.1.2 The importance of the political context

In addition to the aforementioned causes of failures of congestion pricing in Manchester, Edinburgh and New York City, it is worth mentioning that the political circumstances and the approval process played an important role as well.

First, factors like the timing of the congestion charge proposals and referendums were able to boost the introduction of the scheme. It is the case of London where the Mayoral election (with congestion charging as part of Ken Livingstone’s program) was held before the development of comprehensive details of the scheme. In Stockholm, the referendum was held after a six-month trial scheme that helped to increase the levels of acceptability.

The different legislative procedures and the involvement of citizens in the decision process should also be considered. In London, the strong political authority of the Mayor allowed him to implement London Congestion Charge(LCC) without any legislative approvals (Schaller, 2010).

In Stockholm, although no political leadership was present, the support from the Green Party to the government coalition depended on the implementation of congestion pricing (Eliasson, 2010).

It is also interesting to see that, while in Manchester and Edinburgh, cordon pricing was voted down by large electoral majorities, in New York City a relatively small fraction of people (about 5% of employed in NYC residents commuting to Manhattan) managed to block the proposal at the Assembly in spite of the approval in the City Council and State Senate.

The involvement and political influence of residents from the suburbs proved to be fundamental in the approval process as well. In effect, in cities like Edinburgh and Manchester people from the outer boroughs (in the majority opposed to the toll) had the possibility to participate in the referendum. In New York, the State Assembly that represented also the interests of residents outside Manhattan succeeded in stopping the proposal. In Stockholm instead, where the referendum was only consultative for municipalities surrounding Stockholm, where the majority voted against the congestion tax.

All these aspects stress the importance of the specific political context where congestion pricing schemes are proposed. In particular, the political power of people living in the outskirts and surrounding municipalities (often car commuters toward the tolled area) seems to be a crucial factor for the success or failure of the proposal.

## **2.2 Factors influencing public acceptance**

Extensive research has been carried out to identify conditions influencing acceptability of congestion charge schemes. Other than the attitudinal factors related to the understanding, the persuasion of the effectiveness of the scheme and the infringement of freedom, there is a series of socio-economic aspects that seem to affect levels of acceptability across individuals. For example, features like education and income, according to classical economic theories might play a crucial role in acceptability, as high incomes will likely be more willing to pay for improved travel conditions because of their higher value of time (VOT). This hypothesis has been confirmed by Verhoef *et al.* (1996; 1997) who found that higher educated, high income and in general people with high willingness-to-pay for time gains find pricing schemes more acceptable than others. Another interesting and in some way expectable finding by Jaensirisak *et al.* (2005) is that car-ownership can affect the levels of acceptance as well. Normally, people relying on public transit are more favorable to charges as they would not be negatively affected.

Other studies have investigated the relationship between design of the scheme and acceptance. For example, Jaensirisak *et al.* (2005) found that the spatial and time dimensions of the toll are an important feature to be considered to achieve public acceptance of the scheme. Indeed,

charges limited to the central area and to peak hour periods seem to increase the level of acceptability. Finally, lower levels of charge, as expected, seem to have higher public support.

A key issue of public acceptability of congestion charge schemes is represented by the usage of revenues. The allocation of revenues strongly affects the public opinion on road pricing as it was demonstrated by a survey by Verhoef *et al.* (1996) where morning peak road users in the Randstad region of the Netherlands stated that their opinion about road pricing. It emerged from that survey that the approval depended on the allocation of revenues of the majority of people (83%). As Viegas (2001) observes, devoting a significant part of the revenues to improvements of the mobility system is fundamental for the credibility and acceptability of the whole scheme, particularly in a period when there is very low confidence in the politicians' management of public funding. This need for revenue hypothecation and its use to benefit the majority of people has been confirmed by Jones (1991) who demonstrated that acceptability could almost double if the use of revenues were explicitly stated. Similarly, Harrington *et al.* (2001) by means of a binary probit model applied to residents of Southern California to forecast their support of congestion pricing (of all roads in the region) found that the level of acceptance decreases by increasing the fare, but it rises with increasing tax rebates.

Revenues can be rebated to users in several ways. The most typical options consist of improvements of public transport alternatives, investments in transport infrastructure (this was the case of the Norwegian toll-rings), and decrease of transport taxes such road or fuel taxes. However, revenues can be used also outside the transport sector for broader objectives such as reductions of income taxes. Inconsistencies emerged from the several studies about the preferred revenue allocation. According to some studies (Jones, 2002) spending on PT improvements seems to be best alternative, while according to other ones (Verhoef *et al.*, 1996; Lex, 2002) objectives in direct interest of road users' such reduction of car or fuel taxes and road investments would be the most supported. A recent study by Ubbels and Verhoef (2006) where acceptance of transport policy measures in general was related to the use of revenues, confirmed that respondents (Dutch commuters) were more in favor of compensations to car drivers by means of a car tax cut to a lesser extent to lower fuel taxes and new roads. An interesting aspect revealed by this study is that features such as income, car-ownership and weight of vehicles strongly influence the acceptance of certain allocation categories. For example, low-income groups generally have stronger preferences for lower income taxes, while high-income groups prefer the construction of new roads. The explanation for that could be that the former usually drive less and the latter have typically higher marginal utility. Overall, these findings suggest that policy-makers should consider also socio-economic characteristics during the design process of schemes if they want to achieve higher levels of acceptance.

Finally it is worth mentioning that factors such as the severity of congestion and the trip length also influence public opinion about congestion pricing (Verhoef *et al.*, 2007). People who suffer from severe congestion and travel longer distances have generally more positive attitude both in terms of willingness-to-pay for time savings and general positive opinion toward the necessity and effectiveness of the measure.

### **2.3 Proposed solutions to achieve public acceptance**

Several solutions to the problem of acceptance of congestion pricing schemes have been proposed in different directions.

In order to tackle the problem of public distrust in the usage of revenues, Anas and Lindsey (2011) advocate the practice of earmarking the revenues for local public transport service, and infrastructure improvements for cyclists and pedestrians. Earmarking revenues for investment in transport infrastructure is the preferred option because it is considered as durable and irreversible, in contrast with reductions of fees or taxes that are viewed as easily retractable. Such a strategy was implemented with success in London and Stockholm.

Furthermore, improvements of public service before the implementation of the pricing scheme may provide an additional boost to public support (Anas and Lindsey, 2011). Positive effects can be even greater if such improvements are presented together with the pricing scheme in a single “policy package”. It is the case of Singapore, where several “push” measures like charge ownership taxes, fuel taxes, and parking fees, and “pull” measures like improvements of public transport service and infrastructure improvements were introduced together with the ERP. Similarly, in Stockholm during the trial scheme local public transit service was enhanced by new bus lines, park-and-ride facilities and increases of capacity of the underground and commuter lines (Eliasson, 2011).

Regarding the issue of understandability of the pricing schemes, Viegas (2001) suggests to sacrifice an optimal economic efficiency for clearer schemes, at least in the first stage. Pricing levels can be adjusted progressively in the following phase in order to achieve higher performances, giving at the same time the opportunity to all the agents to adapt smoothly to the new conditions and adopt gradually new strategies: in the short term different route and travel time choice; in the medium term mode and travel choice; and in the long run working and living location choices. In contrast with this vision, Eliasson (2010) argues that an oversimplification of the scheme may dwarf its performance and deriving benefits.

As far as the issue of revenue allocation is concerned, several economic models based on marginal cost theory have been developed in consideration of the issue of acceptability. In order

to minimize “any distortionary impacts and perception of gauging” (Levinson, 2010) some scholars (Dial (1999) from Levinson (2010)) suggested to adopt a minimal-revenue congestion pricing where the total raised revenue would equal zero. Alder and Cetin (2001) theorized a sort of social optimal assignment theory according to which revenues collected on the most used routes should be rebated to drivers on less desirable ones. Levinson and Rafferty (2004) proposed a similar approach where drivers who start the queue, also called “delayers”, compensate those at the back of the queue in order to spread the peak. Furthermore, credit-based pricing schemes based on credits and fees according to the average usage of facilities can be implemented (Kockelman and Kolmange 2005). Mayeres and Proost (2002) extend the concept of Pareto-frontier to the trade-off between equity and efficiency of road pricing and consider “changes to financing acceptable when they are Pareto-improving” (Levinson, 2010). More generally, Goodwin (1989) and Small (1992) proposed a “Rule of Three” according to which revenues should be equally split and reinvested in: reductions of taxes such as fuel or car taxes; investments in public transport services; investments in new roads or subsidies to travelers. Using this approach, nearly everyone affected would have some offsetting benefits so that the majority would be favorable to the introduction of the scheme. However, as Anas and Lindsey (2011) observe, the allocation of revenues should not aim only at improving acceptability, but also at allowing sufficient levels of efficiency. Indeed, acceptability concerns should not compromise broader efficiency objectives.

Scheme	Against/Negative	For/Positive
Stockholm Congestion Charging (2006, 2007 <sup>2</sup> )	2001: 51% 2006: 46%	2001: 38% 2006: 52%
London Congestion Charge (2006)	2002: 40% 2003: 24/28%	2002: 40% 2003: 48/57%
Urban road pricing in Norway		
Bergen (1986)	Before: 81% After: 42%	Before: 19% After: 58%
Oslo (1990)	Before: 70% After: 59%	Before: 30% After: 41%
Trondheim (1991)	Before: 91% After: 53%	Before: 9% After: 47%

**Table 2.2: changes of acceptability of congestion charging (source: Pridmore and Miola, 2011)**

According to several studies (Pridmore and Miola, 2011; Levinson, 2010; Eliasson, 2010) increase of public support after the introduction is physiological in pricing schemes, when properly designed. As Table 2.2 shows, positive attitude towards congestion charge increased significantly in the short term. Some of the reasons behind this phenomenon can be found in higher benefits (and lower disadvantages) than expected and in a sort of psychological resignation to something which has already been introduced (Eliasson, 2010). This is one of the

<sup>2</sup>A trial scheme operated for 7 months in 2006 before the congestion charge was implemented on permanent basis in 2007

reasons why trial projects might help people to understand the benefits deriving from the pricing measure. Moreover, such a phenomenon may allow transport policy-makers to make more optimistic forecasts during the planning phase.

## **2.4 The concept of equity**

The definition of equity represents a crucial issue in the assessment of equity itself (Ramjerdi, 2006). Equity can be considered from several perspectives such as opportunities, rights, resources, wealth, income, utility and so on. According to some scholars (Ecola and Light, 2009), achieving equality in one dimension may even determine inequality in other ones. Indeed, adopting a certain perspective implies different approaches in the analyses and ultimately different evaluations. This is the reason why there is no standard or best way to consider transportation equity and it would be a good practice to consider different perspectives in the same study. It is not the purpose of this research to present all the ways equity has been defined and to compare them. It is however important to highlight that evaluating transport equity is not a standard procedure and it is a very subjective process that depends on the ways people are categorized, the impacts considered and the different ways to measure them (Van Wee and Geurs, 2011).

The transport planning literature focuses on the distributional effects on individual characteristics such as gender, age and physical imparity, while the economic literature characterizes impacts in terms of variations of costs and benefits of different individuals or groups. The planning literature mainly analyses the effects of congestion pricing systems on the ability of “transportation-disadvantaged individuals or groups to participate in life activities, such as jobs, medical care, education and shopping” (Ecola and Light, 9). Economists usually refer to equity in regard to distributional impacts on different income categories and working and/or living locations.

Hence, following from Ecola and Light (2009), in this report the *equity* issue will be defined as the “fair or reasonable” distribution of costs and benefits (monetary and non-monetary) among the members of society. In the congestion pricing literature equity evaluations traditionally concern the outcomes of a charge compared with the original non-pricing conditions. However, as it will be pointed out in the following sections, this survey and research will focus on the estimation of the distribution of measurable effects; broader moral considerations about the fairness of the schemes will not be explored in depth.

Efficiency in transportation studies corresponds to the “policy’s ability to maximize aggregate social welfare regardless of whether some individuals are made worse off” (Ecola and Light, 10).

Welfare is considered here as the overall well-being of people determined by benefits and costs deriving from policy strategies measured in terms of money or utils. Since congestion pricing determines changes in travel conditions (variations of travel time) and travel behavior (modal shift, changes in departure time), all these aspects need to be monetized. A policy measure is defined regressive when it imposes greater burden (costs) on lower incomes. Uniform taxes are typically regressive as everyone pays the same amount regardless of his economic conditions.

Following from this conception of equity and efficiency, a dualism between these two objectives often arises when congestion pricing is implemented as schemes are typically designed to maximize overall welfare of people rather than providing an equal distribution of costs and benefits.

Finally, the concept of equity can be applied in several different ways. Three major approaches to equity in transportations studies can be identified: horizontal, vertical and spatial. Horizontal equity concerns the distribution of impacts between individuals or groups considered equal. Then, an horizontally equitable policy equally benefits or worsens people belonging to the same group. Vertical equity is concerned with the distribution of costs and benefits across groups with different abilities and needs. Typically, differentiation of groups is made on income or wealth basis. Spatial equity combines both horizontal and vertical equity to evaluate impacts on people living or working in different locations. Part of the literature considers spatial equity a particular typology of vertical equity with a strong link with the concept of accessibility as it assumes that everyone has the same right to basic level of access to activities (van Wee and Geurs, 2011).

## **2.5 Evaluation of equity**

When analyzing equity, the choice of the evaluative space is crucial. There are several ways to measure the distribution of benefits and costs and people's preferences for different outcomes (Ecola and Light, 2009). Every approach reflects a different perspective in the evaluation and attention towards particular groups and issues.

### **2.5.1 Criteria**

Traditionally, transportation economists adopt welfare-based measures based on microeconomic theories to identify the equity impacts of new policies. In this way, benefits and costs affecting individuals before and after the introduction of congestion pricing system can be compared. The main benefits deriving from congestion pricing schemes are reduced travel times, reduced emissions and decreased traffic accidents. Costs typically associated with congestion pricing schemes are the money paid by drivers who decide to drive anyway in the

tolled area/facilities, changes in departure time, shift to other modes and reduced amount of travel.

In welfare-based studies of equity impacts of congestion charges, the economic burden for different groups can be calculated by simply accounting for the incidence of toll payments, or by considering travel behavioral changes when travelling. The most sophisticated studies can include also the impacts deriving from different revenue allocation strategies. Theoretically, welfare-based approach allows investigations of both horizontal and vertical equity concerns and it can be suitable for different considerations of equity, but it also presents some drawbacks (Ecola and Light, 2009). First, most of the analyses have been based on hypothetical situations where tolls were incorporated in sophisticated models rather than real data obtained from actual congestion pricing implications. As a result, it is difficult to verify the accuracy of most of the models used to evaluate congestion pricing. Furthermore, in order to keep models from becoming intractable, assumptions are usually made on the costs of behavioral changes (e.g. value of travel time savings) or on the rebate mechanism (not so straightforward as often considered).

### **2.5.2 Measures**

Evaluating distributional effects of congestion charges is a thorny problem. As a matter of fact, there are not only a large variety of impacts deriving from it, but there are also many different ways to measure and assess them. Moreover, as Ecola and Light (2009) observe, the evaluation of equity for congestion pricing policies is complicated by the following aspects:

1. *Differences in context.* The location where congestion pricing is applied plays a crucial role in terms of outcomes. People's residential and working location, the presence of valuable alternative modes to car significantly affects the outcomes. For this reason, it is not possible to compare equity concerns of different cities.
2. *Dependence on model.* Studies of congestion pricing often rely on highly sophisticated computer models that, however, are mainly designed to forecast traffic conditions deriving from transportation policies or investment rather than investigating distributional impacts. It is almost impossible to account for all the cross-correlations between variables like income, car-ownership, gender, household size etc. Then, a trade-off between "good geographic resolution and good representation of the simultaneous distribution of socioeconomic variables" (Eliasson and Mattson, 2006) arises.
3. *Assumptions about the value of monetary benefits and costs.* Giving a monetary value to benefits and costs, especially to travel time changes is a complex issue that requires a series of assumptions. Indeed, the value of travel time varies not only across different

incomes, but it also depends on the trip purpose (e.g. commuter or leisure trips) (Small et al., 2005).

4. *Dependence between forms of transportation finance.* Equity impacts of congestion pricing are partly dependent on the operating system of financing transportation, i.e. investments and taxes for road and public transport services. Hence, equity concerns need to be evaluated from a broader perspective.

Different measures and even reference units used to analyze distributional impacts imply different assumptions and perspectives in the evaluation of equity (Litman, 2003). The most common reference units used to compare transport policy impacts are: per-capita, per-trip, per-passenger-mile, or per-dollar. When using per capita as index, analysis typically assumes that benefits/costs should be equally distributed to every person. Using per-trip indirectly infers that people travelling more have right to more resources. Cost recovery ratio between the costs imposed by an user and what he gives back through taxes and user fees implicitly implies that public resources should be distributed among people proportionally to the contribution given.

Even the scale chosen to evaluate the impacts may lead to significantly different outcomes. For example, congestion pricing may be overall beneficial for the municipality, but disadvantageous for the surrounding region and vice versa.

These considerations are in line with Ramjerdi (2006) according to whom different measures of inequality reflect different perceptions of the issue because of their both normative and descriptive nature. Indeed, these measures not only “describe differences in a population relative to a given variable such as income, but they can also represent the manner in which these differences should be measured” (p. 68).

According to Ramjerdi (2006) inequality measurements can be divided into three typologies: statistical, welfare and axiomatic. Statistical measures examine the distribution of a characteristic such as income in the population. Major statistical measures are range, variance, measure of variation, log variance, Gini measure, and Theil’s entropy measure. Welfare measures (e.g. Kolm and Atkinson measurements), which derive from economics, include equity concerns in a social welfare function. Axiomatic ones identify how well certain required properties (axioms) are addressed.

Table 2.3 illustrates some pros and cons related to the most common inequality statistical measures. Gini Coefficient is probably the most popular inequality index for its simplicity and relative ease to be derived measures the distribution of income in a population. The Theil Coefficient derived from the generalized entropy index is also a very suitable statistic to measure inequality, especially because of it is additive across different spatial groups. The

Coefficient of Variation is a more general normalized measure of dispersion of a probability distribution.

Ramjerdi (2006) used statistical and welfare measures to investigate equity impacts of different fares and reinvestment strategies of the Oslo toll ring (Oslo Package) and demonstrated that policy packages ranked differently depending on the chosen indicator. This study confirms that it is difficult to make an evaluation on the basis of a single measure and that it is necessary to consider a broader set of measures.

Measure	Definition	Pros	Cons
Range <sup>3</sup>	$R = Y_{max} - Y_{min}$	Understandable; computationally easy	Rough; does not weight observations; issue of outliers
Variation and Coefficient of variation <sup>4</sup>	$c = \sqrt{V}/\bar{Y}$	Fairly understandable; if weighted immune to outliers; account for all observations	Require comprehensive individual level data; no standard for an acceptable level of inequality
Gini <sup>5</sup>	$G = \frac{1}{2n^2\bar{Y}} \sum_{i=1}^n \sum_{j=1}^n  Y_i - Y_j $	Good economic index; incorporate all data; allows direct comparison between units with different size populations; intuitive interpretation	Requires comprehensive individual data; more sophisticated computations;
Theil's entropy	$T = \frac{1}{N} \sum_{i=1}^N \left( \frac{Y_i}{\bar{Y}} \right) \log \left( \frac{Y_i}{\bar{Y}} \right)$	Can effectively use group data; allows the researcher to parse inequality into within groups and between group components	not very intuitive; cannot directly compare populations with different size or group structures

Table 2.3: measures (adapted from Ramjerdi (2006))

To sum up, it is important to understand the assumptions and perspectives related with the choice of the reference units to evaluate equity concerns. Especially, because of the limited possibility of translating and generalizing criteria and measures, the adoption of different criteria and measures needs to be in line with the goal of the study.

<sup>3</sup> Y is a measure of welfare

<sup>4</sup>  $\bar{Y}$  is the mean level of welfare

<sup>5</sup> N is the number of observations on welfare;

## **2.6 Previous research on distributional impacts: potential winners and losers**

Many scholars argued (Levinson, 2010; Anas and Lindsey, 2011) that congestion pricing schemes are intrinsically inequitable as the distribution of gains (time savings) and losses (toll) will never be “perfect” among individuals living and working in different locations or having different value of time. For this reason, in order to make adjustments and develop policy strategies to achieve the fairest possible schemes, it is first necessary to identify potential winners and losers during the design stage. Several characteristics may affect the outcomes of a congestion charge and different categories of people can be identified when investigating equity impacts. These factors are once again influenced by the different conception of equity. From a “transport policy-maker perspective” potential losers may be people with disability, disabled, children and elderly and immigrants. From the “transport economists perspective” factors that may determine potential losers are income, car ownership, living and working location and availability of public transport services.

In this context, the focus of this review is on the economic studies of distributional effects rather than on the evaluation of equity. This approach is based on objective measures and it does not include any moral judgment leading to conclusions about the “fairness” of the policy measures.

According to several economists, congestion charges will be likely regressive as people with higher incomes have higher VOT and they will probably benefit more from travel time savings. Furthermore, people with lower income have typically less flexibility of working times; hence they would not be able to reschedule their trips to avoid tolls (Arnott *et al.*, 1994). This view has been mainly endorsed by Small (1983) and Arnott *et al.* (1994), Santos and Rojey (2004) who investigated the distributional impacts of hypothetical single facilities and area-based tolls. These studies differentiate from each other in the evaluation of gains and losses. For example, Small (1983) estimated welfare effects at aggregated level by simply considering modal change and cost of travelling, while Arnott *et al.* (1994) include in the evaluation the costs of early or late arrival as well. Despite different approaches, all these studies demonstrated that tolls are typically regressive before they are rebated as drivers with higher VOT (such as high-income groups) have higher benefits. However, in case revenues are redistributed by means of a lump-sum the system might turn progressive.

In contrast with this view, other scholars believe congestion pricing schemes may benefit more low-income groups when they mainly rely on public transport. Indeed, they would not be affected by the toll and they would have the largest benefits (especially if revenues are reinvested in public transport improvements as Evans (2002) noted). This assumption has been confirmed by Eliasson and Mattsson (2006) who developed a method for a quantitative

assessment of distributional effects applied to the congestion-charging scheme in Stockholm. In this study, the authors quantify the benefits and losses of congestion pricing as the net effects of four components: higher travel costs, changed travel behavior, shorter travel times and revenue generation; in order to investigate different revenue allocation schemes. The main finding was that the scheme was generally progressive as high-income groups live mostly in the city centre and consequently they are more affected by the toll (they pay more and they reduce their trip more). Reinvestments in public transport improvements seemed to be the most progressive solution.

Car usage and public transport availability might also play a crucial role in the distributional impacts of toll. As it was already mentioned before, public transport users will likely have higher benefits than car users (unless public transport services become overcrowded because of the modal shift), since they would gain from better traffic conditions without any extra cost. However, the introduction of a toll may determine strongly negative impacts for those people who drive car to commute and they have no valid public transport alternative.

The geographical distribution of costs and benefits is also an important aspect that needs to be considered as the effects of tolls might vary significantly for people living and working inside or outside the tolled area. Furthermore, travel behavioral changes (mode, route, timing) to avoid fee for people who have to cover long distances to reach their workplace, may also determine overall negative offset. In this direction a study about the efficiency and equity effects of different cordon and area pricing schemes was carried out by Maruyama and Sumalee (2007). The authors, by means of trip-chain equilibrium model applied to the city of Utsunomiya in Japan, demonstrated that area schemes are more efficient, but also higher levels of spatial inequity (measured by Gini coefficient). Furthermore, investigations about the effects of different coverage area highlighted that wider perimeters implied increases of welfare and equity impacts.

As it is possible to deduce from the previous considerations, it is not possible to come to clear-out conclusions about winners and losers of congestion charges. Indeed, all the aforementioned aspects need to be considered simultaneously to achieve a comprehensive evaluation of the outcomes for every single case study. For example, car user characteristics, public transport availability, and living and working locations in European cities are completely different from those in American cities. Levinson (2010) tried in a very schematic way to identify some possible categories of winners and losers according to different reactions to the introduction of the toll, represented in Table 2.4.

Category	Winners	Losers
Unchanged, fee charged (tolled)	Travelers valuing the time savings higher than the fee	Travelers valuing the time savings below the fee, but having only unattractive travel alternatives
Changed to toll facility	Persons now finding it profitable to undertake trip (or change trip timing, route, or mode choice), even with a fee because travel time will be reduced	
Changed from toll facility	Travelers who switch from driving to bus or HOV services which are now better because of lower congestion	Persons abstaining from travel or changing to less attractive travel times, routes or modes to avoid fee
Unchanged, fee not changed	Public transport and HOV users experiencing time savings	Persons experiencing congestion on road or public transport caused by persons who have changed travel behavior to avoid fee

Table 2.4: winners and losers (source: Levinson, 2010)

## 2.7 Conclusions and recommendations for the present study

In the previous sections, by means of ex-post evaluations of real experiences and simulations of hypothetical congestion pricing schemes, the issues of public acceptance and equity were presented. It clearly emerged that public acceptance is a veritable *conditio sine qua non* for the implementation of congestion pricing schemes, as even cutting-edge schemes with proven potential benefits have failed when public support was missing. Equity concerns seem to be one of the main reasons behind resistance together with wrong perception and comprehension of the problem and distrust in public authorities' commitment of rebates. When a policy measure is considered unfair towards certain groups, these ones will likely protest and block if they have enough influence in the decision process.

An appropriate design of congestion charges might increase potential levels of equity and ultimately of acceptability. Hence, quantitative analyses of distributional effects that allow transport policy-makers to make objective assessments of the schemes and determine potential winners and losers need to be carried out. Indeed, identifying large disparities in the outcomes during the design stage would allow adjustments in the models that eliminate or reduce inequity and make the measure more tolerable. Thus far, as Jaensirisak *et al.* (2005) observed, little attention has been paid to impacts deriving from different design characteristics such as the location, timing, level of charge and complexity of the charging regime. Furthermore characteristics such as transportation mode and level of income that certainly affect travelers' disposition to accept pricing measures, have been only marginally explored and seldom considered in the evaluations.

The aim of this research is to include considerations about the distributional effects of cordon tolls controlled by macroscopic traffic models implemented in agent-based transport models in

order to allow further speculation of the issue of public acceptance. Thanks to its high level of disaggregation (individuals or households), this typology of models allows very detailed analyses.

Although equity concerns are only a part of the reasons behind public resistance to pricing measures, an objective evaluation of distributional effects of alternative schemes may represent an initial step toward a better understanding of their effects and the formulation of effective solutions. Indeed, analyses of socio-economic and geographical characteristics of the system are necessary in order to develop progressive schemes and where not implement adequate redistribution strategies.



### 3. Congestion Pricing Models

Congestion pricing represents a veritable meeting point between traffic engineering and transport economics research. Since the first study by Pigou (1920) an extensive body of knowledge about congestion pricing models has constantly grown to develop efficient and effective schemes. Traditionally, several criteria have been applied for the design of road pricing schemes according to different policy objectives. For example, as Chakirov and Erath (2012) noted in a recent report: “Maximizing social welfare by optimizing the overall network performance in terms of total travel time is among economists the most accepted approach” (p.2). Other approaches addressed the engineering aspects of congestion to improve traffic performances of the system such as optimal speeds, whereas other ones steered the research toward the maximization of profits and/or revenues. Consequently, such a wide variety of studies have lead to different and occasionally even conflicting outcomes. Despite that, it is possible to identify some common trends the evolution of road pricing models in order to confront with several important theoretical and practical related issues. Figure 3.11 represents a schematic overview of major papers about congestion pricing models with a particular focus on the dynamic representation of congestion.

One of the main topics of research consisted in the heterogeneity of users, particularly concerning the aspect of value-of-time (and utility functions) that determines different reactions to charges in terms of trip timing and mode choice. Important contributions from this point of view came from Newell (1987) who elaborated a model incorporating different cost functions for users relative to their respective work starting time and Verhoef (2002) who considered differentiation of willingness to pay according to the typology of user.

Another fundamental goal that have characterized the evolution of congestion pricing models has been modelling the elasticity of demand (sensitivity to variations of tolls) to better represent travel behavioural changes such as departure time, mode, route, and travel choice determined by tolls. Significant progresses in this direction have been made by scholars like Arnott *et al.*(1993), Yang and Meng (1998), and Verhoef (2002) that improved traditional Marginal Cost Pricing and Bottleneck Models. In more recent years de Palma *et al.* (2004) and Zheng *et al.* (2012), by means dynamic traffic simulation models or agent-based transport models like METROPOLIS and MATSim, they have been able to reproduce congestion pricing schemes with high levels of users’ heterogeneity and demand elasticity, and consistent with traffic conditions.

Further complexity has been introduced in the models to develop the initial models of a single facility or two alternative roads into complex ones modeling large transport networks with

thousands agents. It is worth mentioning, however, that these achievements have been made possible also thanks to important technological progresses that have significantly accelerated the computational process.

In the following literature review instead, more attention is given to the evolution from static to dynamic models because a full understanding of the dynamic characteristics of congestion is a key requirement for the implementation of successful control strategies. De Palma and Fosgerau (2010) defined “static” those models treating congestion as “constant over some given time period” and where “time dimension is not explicitly involved” (p.2). Indeed, given certain characteristics of the demand (travel patterns) and supply (characteristics of the network) the travel time can be directly derived. Furthermore, traffic conditions are treated as constant as all the characteristics are considered as averages of a certain time period. Despite static congestion pricing models represent an useful benchmark for exploratory evaluations of pricing strategies, they clearly appear as an oversimplification of the phenomenon of congestion that is instead inherently dynamic as it (usually) occurs in specific periods of the day (peak hours), it evolves over time and it is “triggered” according to the existing traffic conditions. Therefore, during the history of congestion pricing several attempts have been made in order to develop models consistent with the dynamic characteristics of traffic.

Following this line of reasoning, this literature review starts with a description of static marginal-cost pricing (MPC) models that treat congestion as a static traffic condition (Section 3.1). After a short digression into the issue of first-best and second-best pricing models (Section 3.2), the bottleneck models that explicitly consider variations of the demand during peak-hours are described in Section 3.3 In Section 3.4 and Section 3.5 the issue of hypercongestion is shortly described and macroscopic traffic models are introduced. Finally, an outlook of more recent fully dynamic models based on dynamic equilibrium simulator (METROPOLIS) and on agent-based transport models (MATSim) are described in Section 3.6 In the final part of chapter, a brief presentation of the approach used in this research study and a discussion of the ideal congestion pricing models are presented respectively in Section 3.7 and 3.8.

### **3.1 Standard Demand-Supply Models**

Static Marginal cost pricing (MCP) models originate from the main concept theorized by Pigou (1920) that congestion is the result of a market distortion, consisting in a discrepancy between the travel cost perceived and borne by single road users and the real cost of their travel imposed on the society. Hence, a user charge should be introduced in order to correct it and allow an efficient usage of the road system. This section briefly presents an overview of this

theory, following on from Small and Verhoef (2002) who in turn summarized the main concepts of economists such as Pigou (1920), Knight (1924), Walters (1961) and Vickrey (1963, 1969).

Given the assumption that congestion is the only externality and distortion in the economy, the delay and additional costs generated by a user to all the other travelers are defined as marginal external congestion cost (*mecc*). The *mecc* is given by the difference between the total costs on society called marginal social costs (*mc*), and the average costs already borne by the traveller himself (*c*).

As Small and Verhoef (2007) pinpoint: “An efficient level of road use is obtained when each trip that is made provides benefits as least as great as its social cost”. Consequently every traveler should pay for the “full” costs his trip generates. In order to do achieve this situation, an “optimal toll” equal to the *mecc* should be set so that the external costs generated by each traveler are internalized:  $\tau = mc - c = mecc$ .

From an engineering perspective an optimal toll determines a shift from what Wardrop (1952) called user equilibrium to a system optimum, characterized by minimum total travel time. From an economical perspective, the equivalent result is reached by maximizing net welfare intended as the difference between aggregate consumer benefit and total cost. However, there is a slight difference between these two concepts as optimal demand doesn't not necessarily corresponds to the system optimum. The latter concerns more the traffic assignment (route choice), rather the first concerns the overall demand (origins-destinations).

The derivation of the optimal toll applied to a single road and single time period is briefly presented below. Additional assumptions include considering users as a continuum (rather than discrete entities) and apply the same value of time for all the users.

$d(V)$  is the inverse demand function and  $c(V)$  the average variable cost of a trip where  $V$  corresponds to the volume of trips made per unit time. In a situation of equilibrium, users' willingness-to-pay  $d(V)$  to the price  $p$  equals the sum of the average cost  $c(v)$  and the toll  $\tau$ :

$$d(V) = p \equiv c(V) + \tau \quad (3.1)$$

The total benefit  $B$  of using the road is given by the area under the demand curve up to the equilibrium volume of trips:

$$B = \int_0^V d(v) dv \quad (3.2)$$

The total cost ( $C$ ) of  $V$  trips correspond to:

$$C = V \cdot c(V) \quad (3.3)$$

Other costs like capital expenditures can be ignored since they do not affect the solution. The aggregate welfare can be identified as social surplus  $W$  that is given by the difference between total benefits and total costs ( $W=B-C$ ). Hence, maximizing  $W$  yields to:

$$\frac{dW}{dV} = d(V) - c(V) - V \cdot \frac{dc}{dV} = 0 \rightarrow d(V) = mc(V) \quad (3.4)$$

Then, it follows from (2.1) that the optimal price is:

$$p = mc(V) = c(V) + V \cdot \frac{dc}{dV} \quad (3.5)$$

That matches with the initial rule of setting the toll equal to the marginal external congestion cost:

$$\tau = mc(V) - c(V) = V \cdot \frac{dc}{dV} \quad (3.6)$$

Hence, the optimal toll, also known as Pigouvian toll, corresponds to the marginal external cost of each trip multiplied by the number of users affected.

Figure 3.1 provides a classic representation of the concept of MCP models. The “no-toll equilibrium” occurs at point E, while the “toll equilibrium” occurs at point O. The social surplus (shaped area EOS) deriving from the introduction of the optimal toll  $\tau$  is given by the difference between social cost saved (area under the curve  $mc$ ) and the loss of demand (area under the curve  $d$ ).

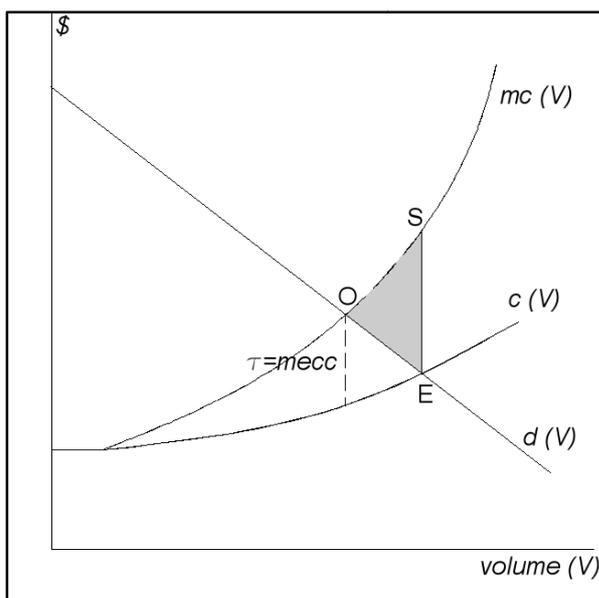


Figure 3.1: Marginal Cost Pricing models

Lindsey (2012) identifies several issues emerging from this approach including: the estimation of demand and cost curves that in turn determines the original no-toll equilibrium; the consideration of the toll collection costs that in case of relatively inelastic demand (hence, small efficiency gains) might exceed the benefits; the fact that the introduction of the toll does not entirely eliminate travel delays is problematic as queuing time is a complete loss for society whereas toll payments are not; the eventuality of public acceptance concerns due to the increased users' private costs determined by the toll.

The Pigouvian toll formula has given rise to a large number of studies that extended the original model in order to apply the model to road networks or include additional features of complexity such as the heterogeneity of users.

Wardrop (1952) introduced the concept of optimal congestion pricing on a network in order to bring it from *user equilibrium* to a *system optimum*. The first condition corresponds to a situation of unpriced network where every user chooses the route with the lowest cost, whereas the second one corresponds to a situation where the route choice allows the most efficient use of the system and minimizes the total travel time thanks to implementation of marginal cost pricing.

Other important contributions to MCP models came from Arnott *et al.* (1994), Small and Yan (2001) and Verhoef (2002) where additional heterogeneity of users was introduced by means of different values of time in order to find an optimal toll on a limited amount of links.

The Pigouvian toll formula represents a fundamental breakthrough in the theory of congestion pricing and it exhibits some important positive features. First, its main rationale and its derivation are rather straightforward. Second, as de Palma and Lindsey (2011) noted in a recent study, the toll depends exclusively on the traffic volumes on the links and the information needed to set the charge is limited to the travel conditions on the single segments. This is a quite practical advantage, as neither O-D matrix nor route choice information is needed to estimate the toll.

On the other hand, MCP models present some practical and theoretical drawbacks (de Palma and Lindsey, 2011; Lindsey, 2012). First, the efficiency is guaranteed only if the model is applied to every link of the network (first-best tolling), even though it is hard to implement and quite impractical as urban road networks have many links. Therefore, in order to overcome these practical issues a series of schemes applied to a limited portion of links called second-best pricing schemes has been developed (see Section 3.2). Second, static MCP models based on a constant demand-supply relationship omit important dynamic features of congestion like the

timing aspect of travel demand. Since traffic conditions are considered as constant over the link, the model cannot represent variations of demand occurring during the day such as peak-hour increases. This aspect of the models constitutes an important theoretical gap as in the reality travel times during peak and off-peak periods differ significantly from each other. For this reason a new category of congestion pricing models explicitly accounting for the time dimension in the demand arise (see Section 3.3).

Static MCP models	
advantages	disadvantages
<ul style="list-style-type: none"> <li>• Relatively straightforward</li> </ul>	<ul style="list-style-type: none"> <li>• not practical: efficient only if applied to every link</li> </ul>
<ul style="list-style-type: none"> <li>• only local information about traffic volumes on links is needed (No need for O-D matrices or route info)</li> </ul>	<ul style="list-style-type: none"> <li>• congestion is stationary</li> </ul>

Table 3.1: advantages and disadvantages of Static MCP models

### 3.2 First-Best Pricing and Second-Best Pricing

As several scholars observed (Small and Verhoef, 2007; Erath and Chakirov, 2012), “first-best pricing” models are characterized by total absence of constraints, and despite they represent a useful theoretical benchmark, they cannot be easily extended to real-world case studies. Thus, in reality “second-best” schemes applying tolls only to a limited fraction of links have been developed. For example in London, Singapore and Stockholm, a central area where riders are charged is defined by means of a perimeter (cordon).

Small and Verhoef (2007) highlight some important aspects of second-best pricing by means of a simple example of second-best toll determined as a compromise (average) of two first-best tolls for two different user groups. First, second-best pricing is less beneficial than first-best in terms of welfare gains. Second, tolls are still related to the marginal external costs, even if not equal. Last, second-best tolls are the result of trade-offs between non-optimal pricing in the different (sub) markets involved.

A classification of “second-best” pricing schemes can be made as it follows (de Palma and Lindsey, 2012).

#### 3.2.1 Facility-based schemes

Facility-based schemes represent the most basic form of pricing and they have been implemented in the past to toll roads, bridges and tunnels, although reducing congestion was not their main purpose. Precisely, these tolls were usually introduced to recover the financial

investments made for the realization of the project. In recent years, a new concept consisting of charging specific lanes of highways (HOT lanes) in order to achieve higher traffic flow performances (free-flow conditions) is becoming popular in countries like the USA. Tolls may be collected at the entry points or at multiple locations based on the distance travelled.

### **3.2.2 Cordons**

A cordon toll consists of a charge paid for crossing the perimeter of a certain area. The toll can be applied to inbound or/and outbound trips. Journeys that begins and finish entirely within the perimeter are not charged. Although cordon-schemes can be applied on multiple cordons, the existing ones are single cordons. The Norwegian toll rings implemented in Bergen (1986), Oslo (1990) and Trondheim (1991) represent the first experiences of cordon-based schemes, even though their main scope was collecting funds for infrastructure investments. Similarly, the EcoPass in Milan (introduced in 2008) is aimed to reduce more traffic emissions rather than traffic congestion.

The Stockholm congestion charge and the Singapore's Electronic Road Pricing (ERP) have been instead explicitly introduced as a traffic demand management measure. In Stockholm the cordon toll including the city centre is implemented during weekdays and it varies according the time of the day. The toll scheme in Singapore that also varies during the day can be considered more as a hybrid of facility-based and cordons because it involves some arterial roads and expressways

### **3.2.3 Zonal schemes**

Zonal or area-based schemes impose a toll not only to vehicles entering or exiting the zone, but also to those travelling within the zone without crossing its boundaries. However, once the toll has been paid, drivers are able to make unlimited journeys inside and across the zone without paying any extra fee. The only zonal scheme currently in use is the London congestion charge introduced in 2003. A flat (fixed) toll charge is applied to all trips within or across an area of 21 km<sup>2</sup> around the city centre during weekdays.

### **3.2.4 Considerations about area-based and cordon-based schemes**

The main advantages of area-based and cordon-based schemes are their transparency that makes them more "user friendly" (De Palma and Lindsey, 2012) and also their relative easy implementation from the administrative and technologic perspective (Santos and Rojey, 2004). But, on the other hand, May *et al.* (2002) point out some major drawbacks: they fail to intercept trips fully inside or outside (cordon-schemes), they induce rerouting of some journeys (hence congestion might increase outside the cordon) and they impose a single charge regardless of the distance travelled. The latter point can be interpreted as a theoretical drawback since the since

trips have increasing marginal costs of congestion with higher covered distance, but cordon and area based tolls make no distinction.

As concern the decision between cordon and area schemes, little research has been made thus far apart from a recent study by Sumalee and Maruyama (2007) where several area and cordon schemes with different boundaries and levels of charge have been analyzed. The results showed that in general area-based schemes perform better than cordon in terms of social welfare improvements and also determine higher levels of spatial equity. The reason for that lies in the fact that for the same coverage the affected number of trips is higher.

### 3.3 The Bottleneck Model

A series of aspects typical of peak-hour congestion neglected by the previously described static MCP models need to be considered. First, the travel choice entails not only the route, but also departure time. Time departures during the peak can be indeed affected by the introduction of a toll and determine changes of intensity and length of the peak. Second, total travel costs include not only monetary and travel time costs, but also costs determined by the deviation from the preferred timing of trips: scheduling costs. Third, static models cannot describe time varying tolls or policies tailored to variable demand.

Hence, in order to include these important characteristics, a new category of models where congestion is defined as queue at a bottleneck and vehicles able to move at free-flow speed in absence of queue has been developed. A presentation of this alternative typology of models called “bottleneck models” is given as De Palma and Fosgerau (2012).

The original model formulated by Vickrey (1969) considers a number of  $N$  identical travelers passing through a bottleneck with capacity  $s$ (travelers per unit of time). The total demand is inelastic.

All the travelers are assumed to have a preferred arrival time  $t^*$  and in addition to the travel time costs, they face some “schedule-delay costs” for arriving earlier or later than the preferred arrival time. The marginal cost of arriving earlier ( $\beta$ ) is always lower than the marginal cost of travel time ( $\alpha$ ). Hence for a trip starting at time ( $t_1$ ) and finishing at time  $t_2$  the total cost is given by:

$$c(t_1, t_2) = \alpha \cdot (t_2 - t_1) + \beta \cdot \max(t^* - t_2, 0) + \gamma \cdot \max(t_2 - t^*, 0) \quad (3.8)$$

Another important assumption is to consider the time of departure coincident with the time of entrance into the bottleneck and the time of exit from the bottleneck coincident with the time of arrival.

Travelers reach the bottleneck according to an aggregate schedule, described in terms of cumulative departure rate  $R$ , where  $R(a)$  is the number of travelers who have departed before time  $a$ . The departure rate  $\rho(a)=R'(a)$ .

If  $\rho(a)$  exceeds the capacity  $s$ , a queue forms and travelers are served according to the first-in-first-out (FIFO) principle.

### 3.3.1 No-toll equilibrium

In a situation of Nash equilibrium where no traveler can reduce his travel costs, all the users experience the same costs. In this situation, where only the first and the last user do not experience any queue and they incur the same schedule-delay costs, the equilibrium costs for every traveler corresponds to:

$$\frac{\beta\gamma}{\beta+\gamma} \frac{N}{s} \equiv \delta \frac{N}{s} \quad (3.9)$$

Where  $N$  corresponds to the number of users. Then the total cost is  $\frac{\delta N^2}{s}$  and the corresponding marginal cost  $\frac{2\delta N}{s}$ , of which half is internal to each traveler and the other half is external.

Figure 3.2 illustrates in the top-half the cumulative arrivals and departures under the no-toll equilibrium (arrivals are assumed instantaneous in free flow conditions). and in the bottom-half the composition of traveler costs as a function of the arrival time. By approaching  $t^*$ , the early schedule-delay costs are replaced by queuing costs, whereas after  $t^*$  late schedule-delay costs gradually replace the queuing costs.

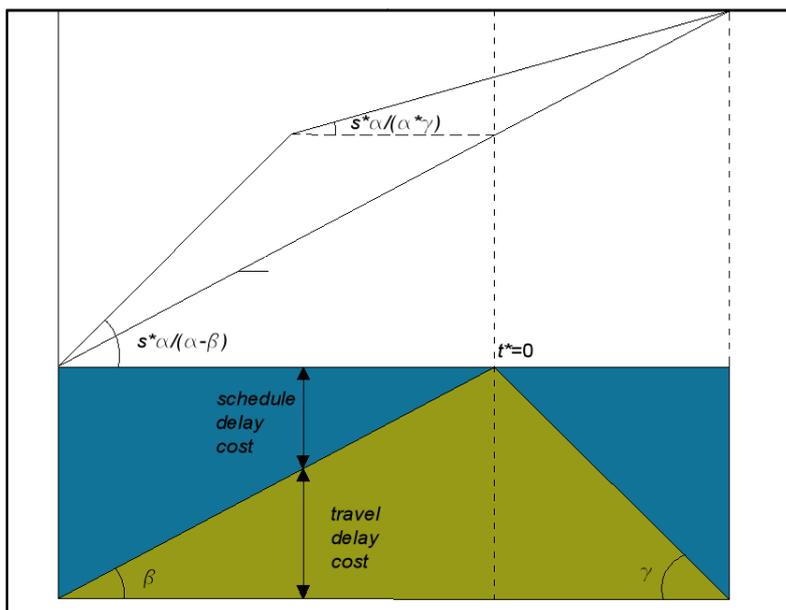


Figure 3.2 Cumulative arrivals and costs in the bottleneck models. Adapted from de Palma, Fosgerau (2010)

### 3.3.2 Optimal toll

An optimal toll that eliminates queuing delays can be introduced in order to change the departure schedules and keep the departures lower than the capacity rate  $s$ . The toll replaces queuing time by a money cost and it corresponds to:

$$\tau(a) = \delta \frac{N}{s} - c(a, a) = \delta \frac{N}{s} - \beta \cdot \max(-a, 0) - \gamma \cdot \max(a, 0) \quad (3.10)$$

The shaded area of Figure 3.2 represents the evolution optimal toll during the time. As it possible to see, the toll totally replaces the queuing costs and it generates efficiency gains through revenues in contrast with the complete waste represented by queuing.

As Small and Verhoef (2007) observe, the benefits from this approach are considerable. The optimal toll in bottleneck models in contrast with the MCP models completely eliminates travel delays. Furthermore, the average travel cost ( $c$ ) for each user is unvaried as the charge simply substitutes the cost of travel time delays. Finally, since arrival times at the destination hold the same both in the unpriced equilibrium and in the toll equilibrium, and no alternative mode to car is included, it is possible to solve morning peak commute problem only by adjusting departure times.

### 3.3.3 Extensions

The original model by Vickrey has given rise to a large series of studies aimed at exploring and including additional features such as the elasticity of demand and the heterogeneity of users. Other researchers have extended the original study with more realistic or practical applications like tolls varying on discrete time intervals.

Arnott *et al.* (1993) in order to account for inelastic demand assumed that the travel demand during the peak depends on the equilibrium cost. As de Palma and Fosgerau (2010) observe: "This is a very convenient way to extend the model: conditional on any equilibrium number of travelers, the properties of equilibrium are exactly the same as in the inelastic case".

Newell (1987) provided an extension of the basic bottleneck model by considering an equilibrium with heterogeneous commuters characterized by different travel scheduling preferences, where everyone minimizes his deterministic cost function. Consequently several slight changes in the outcomes occur as Small and Verhoef (2007) observe: "the queue grows while early travellers and it shrinks while late travellers are arriving. Second, a person exiting the queue exactly at his desired time must incur the maximum travel time incurred by users with the same characteristics. Third, under certain circumstances, Newell obtains for this more general model a key pricing result of the basic bottleneck model: if users have identical values of

time, a time-varying toll can be defined which has no locative effects on the number of travellers or their time of passage through the bottleneck” (p. 133).

Other studies have extended the original bottleneck to network of bottlenecks. For example, Arnott, de Palma and Lindsey (1998) demonstrate that the equilibrium and the optimum toll hold the same in case of two bottlenecks in series with different capacities and with no active origin and destination in between. Again Arnott, de Palma, Lindsey (1990) study the more complex case of two bottlenecks in parallel and they found an optimal toll that eliminates travel-delay costs. In this case, the optimal time-varying tolls are “analogous to those for a single bottleneck and the timing of exits and the route split are identical between the unpriced equilibrium and the optimum” (Small and Verhoef, 2007). Optimal tolls have been studied also in case of large scale networks by means computer simulation models such as METROPOLIS (see Section 3.5).

Another possible differentiation of tolls can be made by considering the time dimension. Indeed, with the development of bottleneck models it has been possible to develop pricing schemes tailored to peak-hour congestion. While in the theoretical studies of basic bottleneck models a time-varying toll could be implemented in order to eliminate queuing delays, implementing dynamic tolls in real world case-studies was not possible for several reasons. First, the technology required to realize such sophisticated schemes was not existing at that time (only in recent times systems like the electronic toll collection or plate recognition have been employed). Furthermore, too complex schemes could result confusing for drivers. As a result, several studies have extended the original time-varying toll with more realistic tolls characterized by “a time constraint”. An essential classification of these second-best congestion pricing schemes can be made as it follows:

- Static or flat tolls are the most basic typology consisting of a fixed charge that is usually applied during the peak hours. On one side, implementing flat toll is relatively easy from the technological and administrative point of view. On the other side, the outcomes may lead to way lower benefits compared with time-varying tolls.
- Step tolls consist of tolls varying over discrete time intervals, but constant within each interval. Arnott, de Palma and Lindsey (1990) proposed a simple single-step toll that “is zero except for a time interval (presumably a subset of the peak period) when it is a positive constant” (Small and Verhoef, 2007). A more sophisticated model characterized by several steps also known as multi-step toll has been introduced by Laih (1994) that introduced the concept of separate queuing facilities in order to address the issue of

mass departures (some users can wait in a separate queue for the toll to shift while others pass through the bottleneck and pay the toll).

### **3.4 The concept of hypercongestion**

The basis for an efficient congestion pricing is the marginal external cost of congestion and it involves essentially the value of travel time and the supply relation. Gaining a proper understanding of the supply side is of crucial importance for the design of congestion pricing schemes. The description of supply relationship has been considered the domain of engineering and it has been ignored by the majority of economic models where the 'supply' was given for example by the bottleneck capacity.

However, in reality congestion is a way more complex phenomenon as several studies in the last decades have demonstrated. The traffic state of a road can be expressed through three main variables, the productivity representing the amount of vehicles leaving the road (traditionally referred as flow), the accumulation representing the amount of vehicles present on the road (expressed as density) and the average speed of the vehicles. The fundamental identity of traffic flow implies that traffic flow equals traffic density times traffic velocity based on the conservation of vehicles equation. The relationship between traffic velocity and traffic density is such that outflow is increasing until capacity of the link is reached, and decreasing when density exceeds a critical level corresponding to the capacity. The diagram representing the relationship between these two variables is a concave curve with a clear maximum (see Figure 3.3). Traffic engineers refers to the left branch of the diagram as uncongested or free flow state and to the right branch as congested state. Hence, according to this view, the number of vehicles that are able to leave the link depends on the amount of vehicles on the link itself. In neither the static MCP models nor in the bottleneck models, traffic flows fall as demand increases.

Hence, traffic management and control measures that enable the system to work under optimal conditions (below the critical density) can significantly reduce time losses due to congestion and ultimately enhance higher benefits than traditional standard models.

A fundamental contribution to the development of congestion pricing models explicitly accounting for dynamic characteristics of congestion came from Verhoef (1999), Small and Chu (2003) who questioned the validity of the traditional supply curve used for equilibrium analysis.

The main idea emerging from "Hypercongestion" (2003) is that a major complication compromises the traditional model formulated by Walters (1961) based on the relationship between the "demand curve" representing the number of users for a given cost of using a facility

and the “supply curve” showing how the cost increase in relation the number of users (see Figure 3.1). The problem stems from the fact that the supply curve, traditionally derived from the time-flow relationship represented by Figure 3.3, may have one or more intersections with the demands curve (Figure 3.4). Particularly, a second or third equilibrium (points y and x in Figure 3.4) characterized by a downward sloping average cost curve can occur. The authors define this condition as “hypercongestion” in contrast with the “ordinary congestion” represented by the lower branch of the curve. A situation where an increase of traffic inflow determines a decrease of the average cost clearly does not make sense.

As “hypercongestion occurs as a transient response of a linear system to a demand spike” (Arnott, 1990; p. 200), this phenomenon is closely related to the peak-hour congestion problem.

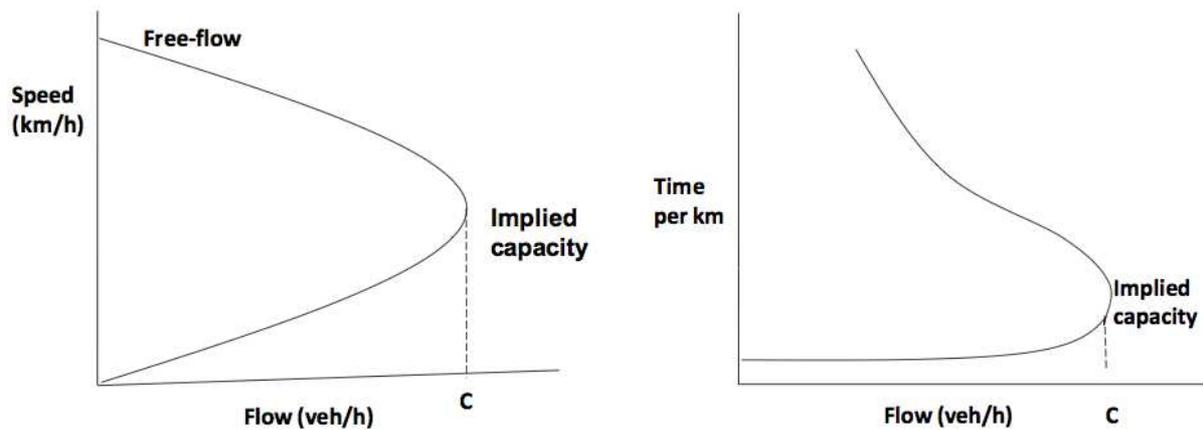


Figure 3.3: Flow-speed relationship and flow-travel time relationship. Source: Liu et al., 2012)

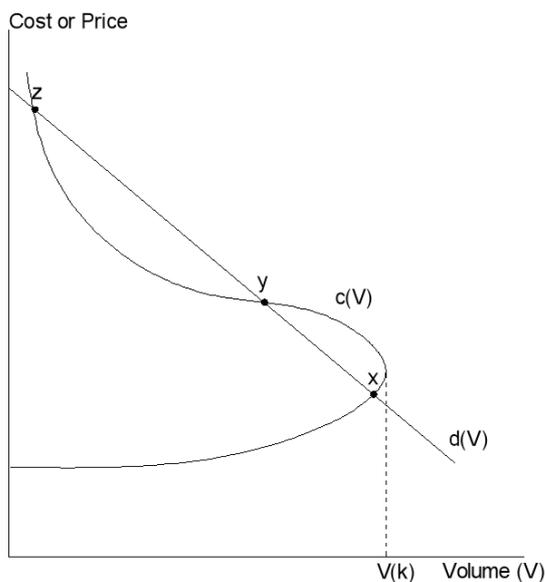


Figure 3.4 Hypercongestion and deriving demand-supply relationship. Adapted from Small and Verhoef (2007)

In recent years, arising from the considerations of Small and Chu, other criticisms have been directed at the traditional approach aforementioned.

According to Liu *et al.* (2011): “What is needed for a supply curve is an estimate of the time which would be spent by the demanded flow, at each of a given set of increasing levels of demand” (p. 230). While for low levels of demand and the system is relatively uncongested the flow-time relationship is quite reliable, during congested conditions the trips last longer and the travel times are strongly dependent of the facility performance.

Daganzo (2007) pushes the question further and debates on the validity of using the average travel time based on specific-demand curves, by proposing the following mass conservation equation to describe the state of the system:

$$\frac{dn}{dt} = I'(t) - o(n(t)) \quad (3.11)$$

where  $n$  is the accumulation (or density in terms of vehicles in the system),  $I'(t)$  is the input rate (inflow) to the system at time  $t$ , and  $o$  is the total outflow from the system as function of the accumulation. As it clearly emerges from the equation, the system is dynamic and in order to determine its conditions information about the inflows are not sufficient, but knowledge about the prior state of the system is also needed.

On the same page, Geroliminis and Levinson (2009) argue that: “the traditional network supply curve (desired or input demand vs. average travel cost) is not consistent with the physics of traffic” (p. 221). Indeed, given a certain demand, the corresponding travel times might significantly differ as it is shown in Figure 3.3. Furthermore, even small variations of the demand can generate large changes of travel conditions for the entire system. In such a context, aspects like the travel time choice and the route choice gain even more importance as they can determine the onset and offset of congestion. Consequently, the toll derived from traditional demand-supply curves may lead either a still congested system (if underestimated) or to a very uncongested system (if overestimated).

### **3.5 Macroscopic Traffic Models**

An innovative approach has been proposed in recent years by several scholars (mainly Daganzo and his colleagues) who combined the “classic” bottleneck model with macroscopic traffic models in order to account for the dynamic nature of congestion.

Daganzo (2007) and Geroliminis and Daganzo (2007; 2008) demonstrated by means of a micro-simulation of San Francisco Business District and a field experiment in downtown Yokohama

that traffic conditions in urban regions can be described at aggregated level by a “Macroscopic Fundamental Diagram” (MFD) if the street network is uniformly congested.

The MFD relates space-mean flow, density, and speed of an entire network of several links as Figure 3.5(a) shows. The MFD expresses a production (product of average flow and network length) as function of the accumulation (product of density and network length) and it exhibits a concave shape with a well-defined maximum representing the capacity of the network. The MFD is an intrinsic property of the network as it depends only on the infrastructure and control and not on the demand. Unless significant changes in the O-D patterns like those determined by big events or evacuations, the shape of the MFD is almost invariant during the day and across the days. Also the maximum space-mean flow is independent from the O-D matrices. Furthermore, the MFD can be determined relatively easily from available technologies like detectors and GPS.

Finally the two scholars observe that a network can be macroscopically modelled as a single bottleneck with state-dependent capacity given by a network’s exit function (NEF) represented as  $F(n)$  in Figure 3.5(b). Precisely, the network can be considered as a large and complex circle where vehicles enter from different origins and they interact with each other in order to reach their respective destinations. The rate at which vehicles enter the circle can be defined as inflow, while the rate at which vehicles leave the circle or complete their trip can be interpreted as outflow. The NEF express a state-dependent discharge rate of the network as function of the number of vehicles in the network. When the average trip length is about constant in the circle, the NEF expressed by  $F(n)$  can be directly derived from the MFD and it equals to:

$$F(n) = \frac{l}{d} Q\left(\frac{n}{l}\right) \quad (3.12)$$

Where  $l$  (lane-km) is the total length of the network, and  $d$  (km) is the length of a vehicle trip and  $n$  is the number of vehicles in the network.

In order to maximize the NEF the congested traffic state on the right branch should be avoided by means of traffic control measures, since equal performances can be achieved if density is kept below the critical value  $k^*$ .

According to Gonzales and Daganzo (2011), congested traffic conditions are not only inefficient, but also unstable. In fact, experiments have demonstrated that, when road users are not able to adaptively re-route their trips in real time to avoid congestion, the more congested the network become the more uneven vehicle distributions develop. Then, if the network is sufficiently congested, such uneven distribution may eventually generate gridlocks.

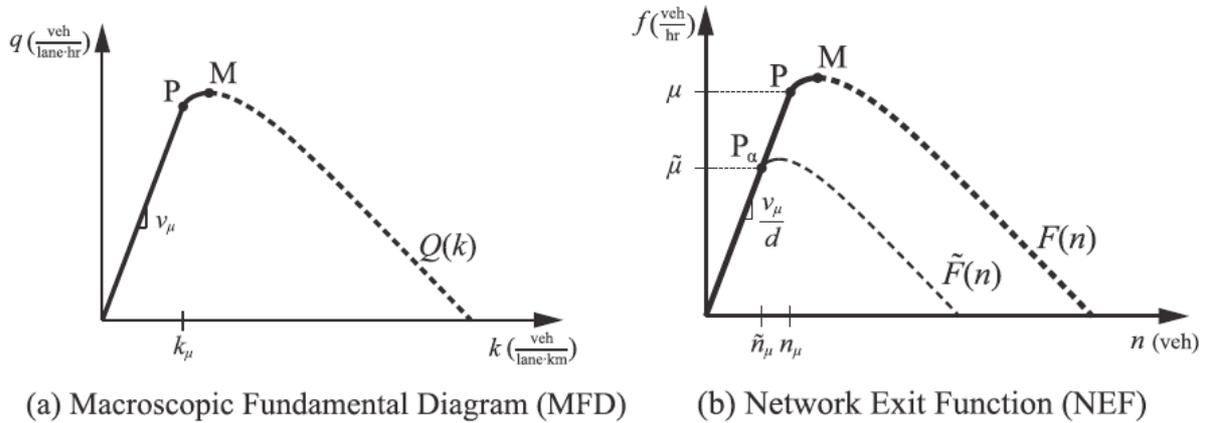


Figure 3.5: Macroscopic Fundamental Diagram and Network Exit Function. source Gonzales, Daganzo (2012)

Geroliminis and Levinson (2009) apply these findings to include in the traditional bottleneck model a supply curve determined by the MFD in order to evaluate the trade-off between efficiency and equity of a cordon-based toll. The paper investigates the possibility of allowing the network to operate at its maximum capacity with non-optimal average speed (“equitable solution”) or imposing a stricter control to reach lower density, but enabling higher average speed (“efficient solution”). Such policy dilemma is well represented by Figure 3.5(b) where the free-flow and capacity regimes are identified respectively by points P and M. State P is more reliable and efficient than state M, as the average speed is higher (and more likely to be stable), but the system operates below its maximum capacity. Thus, fewer people would be able to pay the toll and travel during the peak-hours. Contrarily, state M is more “equitable” in the sense that it allows more users in the system, but it has slower speed. Please, note that in real case studies the flow corresponding to free-flow conditions may be even 20-30% lower than the maximum flow. The final results in terms of total welfare gains of such policy decision depend on the distribution of the value of time and costs of schedule delay within the population. Further discussion will be provided in Chapter 4.

In order to account for the dynamic properties of congestion, the two authors propose an improved version of the classic bottleneck model with a variable capacity  $s$ (outflow) dependent of the average density  $k$  (accumulation) of the system. The no-toll equilibrium is derived in a similar way to that used in Vickrey’s approach and described in Sub-section 3.3.1, but accounting for travel delays determined by the outflow function of the system which in turn depends on the average density. The resulting departure curve determines the same triangular shape of queuing delay as in the original bottleneck model.

Finally, an optimal fine toll to reduce congestion is derived such as the time spent in the queue (travel delay) is replaced by a time-dependent toll according to the principle by Newell (1987) that: “one could convert the worthless expense of queuing into money” (p. 87). The main

advantages deriving from the introduction of pricing schemes based on traffic-state dependent supply model are: a complete discharge of travel delays; a reduction of the length of the rush hour; a resulting optimal toll smaller than the average delay cost, which implies savings in travelers' delay are significantly higher for the network model.

Gonzales and Daganzo (2012) use a similar approach to study the morning commute problem for a network served by both car and public transit (PT) sharing the same space in order to find an optimal fine toll and PT fare. The problem is addressed by assuming the network as a bottleneck whose outflow function (NEF) is derived from the MFD. In this case, despite the transit is fully segregated on its own lane and it is not subject to traffic congestion, the capacity of the bottleneck is shared between car and PT. In order to simplify the problem, the users are divided in two categories: captive car users and transit riders who are not able to switch mode and they can only adjust their trip departure, and non-captive users who can choose when and how travelling. The results of the study show that including public transit in the model represents a Pareto improvement as all the users experience lower costs. Furthermore, thanks to public transport the rush period shrinks both in the user equilibrium and in the system optimum.

Arnott (2013) recently introduced in his "Bathtub Model" the new findings about the MFD in order to derive a pricing scheme that allows the system to work beneath the critical density. An optimal time-varying toll is derived as difference between the marginal social cost and the user cost evaluated at the social optimum where the sum of total travel time cost, total time early cost, and total time late costs is minimized. The most important feature of the model can be found in the boundary conditions of the optimization problem where the arrival rate of trips explicitly depends on the traffic density of the system.

### **3.6 Traffic simulation models**

Another group of studies on congestion pricing can be identified in traffic simulation models. When studying in depth the effects on travel behavior determined by road pricing schemes implemented in large-scale transport networks, numerical methods represent a valid alternative to analytical models. When developing congestion pricing schemes an ideal model would: represent dynamically trip-timing decisions, account for different time-preferences, toll during specific periods of the day (peak hours), and consider relevant networks. However, conventional studies based on the MCP models or the bottleneck models could not cope with all these requirements at the same time (de Palma *et al.*, 2005). As a result, a new solution concept has been adopted in recent years by means of network micro-simulation models based on disaggregated information like METROPOLIS and MATSim. Both the simulations are based on

dynamic traffic assignment that allows effects such as timing and route decisions to avoid the toll.

De Palma *et al.* (2005) used the dynamic equilibrium simulator METROPOLIS that treats endogenously departure-time, mode and route choice to investigate congestion pricing schemes in the Ile de France (Paris metropolitan area).

METROPOLIS is a trip-based model that considers travel demand at the level of individual travellers (the number of trips, destinations and vehicle occupancies are exogenous). The generalized car travel costs can be divided in travel costs, schedule-delay costs and possible monetary costs (toll). A two-stage nested logit model is used to determine travel behaviour choices as shown in Figure 3.7. First, a choice between auto and public transit is made following a binary logit function. Then, if the car is chosen, a standard continuous logit specification determines the time-departure choice. Finally, the route choice is governed by a heuristic that minimizes the generalized costs (based on Wardrop’s first principle). In order to account for real traffic conditions, route choice decisions are revised at each intersection (travelers are able to observe travel costs on each downstream link). Congestion is represented by means of queue on each link occurring when the flow capacity of the link is reached. Although METROPOLIS could model horizontal queuing and spillbacks, these features were not enabled in those studies. As concern the simulation, travel behavior is characterized by a learning approach that involves a “day-to-day adjustment process” (de Palma *et al.*, 2004) from previous conditions in order to reach a stationary “equilibrium”.

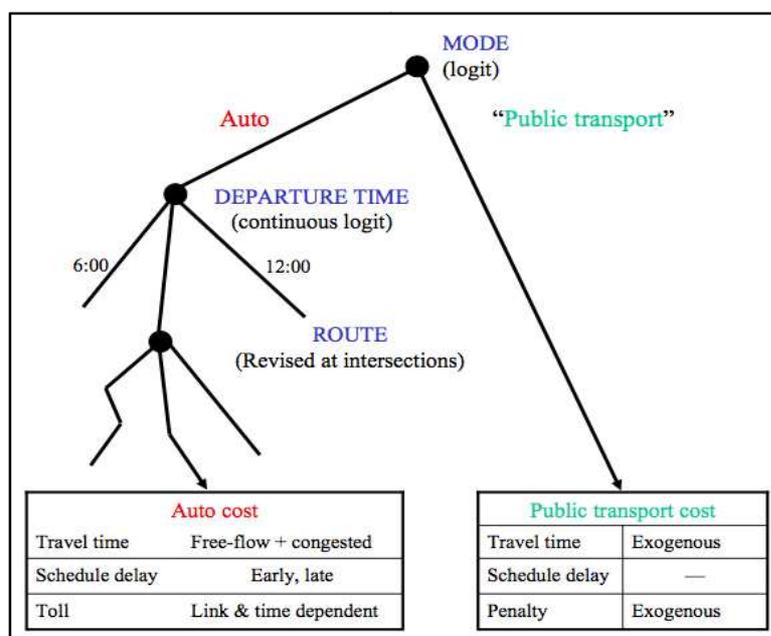


Figure 3.7: Overview of METROPOLIS simulation. Source De Palma

Tolls can be implemented by adding extra costs to each link (link-based). The effects in terms of welfare gains of six alternative schemes are investigated in order to maximize social surplus. First, a time-varying toll that eliminates queuing delays on every link to achieve *system optimum* (Wardrop's second principle) is implemented. The second scheme corresponds to a flat toll on the same set of links. The third and fourth schemes apply a flat and step toll to a cordon. The fifth and sixth schemes apply a flat and step toll to an area identified by the same of perimeter of the cordon-based tolls.

Two research groups in TU Berlin and ETH Zurich have adopted a similar approach consisting in pricing models developed in the agent-based transport simulation MATSim. MATSim: "integrates human behavioral models with queue-based traffic simulation and provides capability for the implementation of large-scale scenarios with several million agents" (Erath and Chakirov, 2012). Thanks to these characteristics it is possible to confront modeling and optimization issues that traditional analytical models could not easily address such as high heterogeneity of users, elasticity of demand et cetera. In addition, its capability of simulating huge amounts of users enables to study realistic scenarios. The main feature of agent-based simulation models is the representation of socio-economic characteristics of individual travelers like income, age, employment etc. that allows detailed analyses of the effects of transport policy measures or projects.

MATSim simulates an entire daily plan of every single user and it considers endogenously mode choice, departure time choice and route choice into a fully dynamic model. Differently from models that use single trips, this model allow predictions on reactions to tolls on the span of the whole day and it achieves higher level of realism. In reality, trips are typically linked to each other as a part of a daily plan and not that meaningful just as stand-alone trips (Nagel *et al.*, 2008). Often activities have higher importance in the daily schedule than trips that simply represent connections among them. For example, factors such as the working time, the opening and closing times of commercial activities play a crucial role in the daily schedule of people, and ultimately in the trade-off between different utilities (working eight hours, being at the shop when it is open, etc.) and disutilities (tolls, being late for work, etc.). Furthermore, each person in the model has several demographics features such as residential and (possibly) work or education locations that determine personal daily plans. Every plan is characterized by a scoring function, which is given by the sum of the utilities from performing daily activities and disutilities from travel times and deviations from preferred schedule (see chapter 5 for further explanations). Hence, in this way it is possible to award punctual performances and punishing long travel times and delays. During the simulation every user tries to optimize the utility of

their daily plan until a sort of equilibrium is reached by evolutionary algorithms. A more thorough description of the model is provided in Chapter 5.

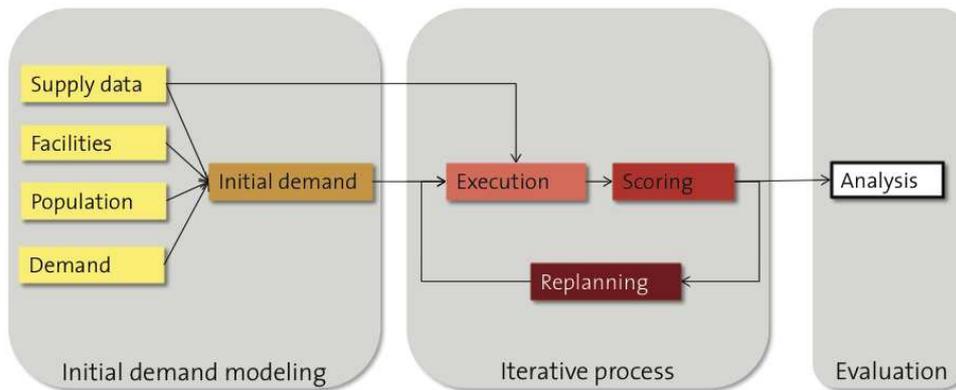


Figure 3.8: Overview of MATSim simulation process. source (Nagel et al., 2008)

Introducing pricing policies through additional costs in the scoring function is rather straightforward in MATSim. Nagel *et al.* (2008) are the first scientists who studied the effects of an evening toll implemented with MATSim in the Zurich metropolitan area. In this study the implemented toll does not derive from any specific rationale, but it is applied in order to demonstrate the potentials of the simulation. Kickhofer *et al.* (2010) took advantage of the highly disaggregated nature of the simulation to investigate the distributional effects on incomes of eight different fares of distance-based tolls. The authors achieved a higher level of heterogeneity of users by developing income dependent utility functions that determine different VOTs.

METROPOLIS	MATSim
<ul style="list-style-type: none"> <li>• trip-based</li> <li>• demand component: disaggregated (derived from OD matrix)</li> <li>• supply component: cellular automata model</li> <li>• traffic assignment: dynamic</li> <li>• home-to-work trips</li> <li>• homogeneous characteristics of traveler</li> </ul>	<ul style="list-style-type: none"> <li>• agent-based</li> <li>• demand component: disaggregated</li> <li>• supply component: queue based model</li> <li>• traffic assignment: dynamic</li> <li>• full daily plans (secondary activity choice)</li> <li>• socio-economic characteristics of travelers</li> </ul>

Figure 3.9: Comparison of METROPOLIS and MATSim

### 3.7 Tolls controlled by the MFD within agent-based models

An additional progress in modeling congestion pricing has been made recently by Zheng *et al.* (2012) who combined macroscopic modeling of congestion with agent-based simulator to develop a pricing scheme efficient from both the traffic engineering and economics perspective.

The authors claim that traditional traffic simulators: “consider demand as an input, i.e. inelastic to traffic conditions”, and that conventional congestion pricing models are based on a network supply curve that is not “consistent with the physics of traffic and the dynamics of congestion”. In order to overcome these issues, they propose a model that combines a MFD that well reproduces the dynamics of congestion with MATSim that well reproduces travel and it considers high levels of heterogeneity.

In the first part of the paper, the authors demonstrate that MATSim outputs are coherent with the MFD by empirical observations of the traffic conditions of the city centre of Zurich. The network is “filtered” to a 1.5 kilometer area that exhibits a well-defined MFD when traffic flow parameters of links in the region are aggregated. Zheng *et al.* (2012) also demonstrates that MATSim is able to reproduce the spillback effect that is the main cause of capacity decrease (right branch of the MFD) of the network. Then, a cordon-based pricing scheme is implemented to the identified area so that the traffic density during morning and evening peak hours never exceeds the critical level corresponding to the top of the MFD. The optimal toll is obtained by means of an “off-line” feedback control process where the toll is updated at the end of each simulation until congested regime is eliminated. Figure 3.10 represents the framework of the model. The authors demonstrate that such “aggregated approach of pricing” (Zheng *et al.*, 2012) is efficient also from the economical perspective as the benefits from travel-time savings are higher than the total cost paid. Finally, an investigation of changes in the travel departure time of “work-related” and “non-work related” trips is carried out.

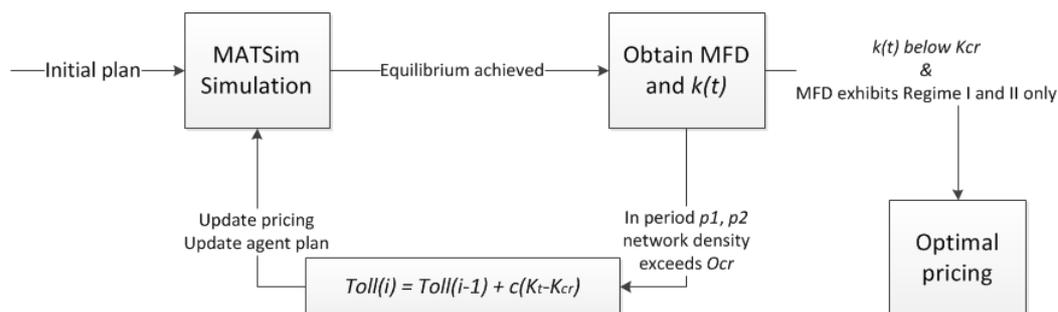


Figure 3.10: offline linear feedback process for the implementation of tolls. source Zheng *et al.* (2012)

### 3.8 The ideal model (and this thesis)

Based on the previous sections, it is possible to provide a basic outline of the main features that an ideal congestion pricing model should have.

First, the main aspects concerning travel behavior such as mode, time, route and travel choice need to be modeled in order to provide a realistic representation of the elasticity of demand (sensitivity to variations of tolls).

Heterogeneity of users is also an important factor for the study of congestion pricing as the impacts of tolls can vary significantly according to the different VOT of people. Indeed, the VOT varies not only across people with different socio-economic characteristics (typically depending on their incomes), but even for identical individuals with different travel purposes (e.g. work and leisure trips). Relying on simple averages can lead to an overestimation or underestimation of the effects as very high VOT of few might increase the population mean. Hence, heterogeneity need to be recognized and properly accounted for when setting the toll levels in order to reach adequate level of effectiveness and identify the distributional effects of the toll.

Furthermore, another factor whose importance has been highlighted in recent years by scholars like Fosgerau and colleagues is the increased travel time variability related to higher levels of congestion. Uncertainty of travel time can influence travel behavioral choices and ultimately the travel costs. For this reason it would be advisable to consider it when developing tolls and assessing their benefits. Simple decreases of time does not seem to be a sufficient indicator of the benefits of charging.

Consistency with the dynamic characteristics of traffic is another important feature that congestion models need to have in order to avoid the onset of congestion. As it has been described in the previous sections, phenomena related to the decrease of performance of entire networks after a critical density (MFD) have a strong influence on the equilibrium between demand-supply.

Finally, explicit consideration of second-best aspects when evaluating the system is important in order to identify possible constraints of charging instruments and practical issues that might lead to lower benefits than would be obtained in the ideal first-best schemes.

In this thesis congestion pricing models based on the MFD within MATSim will be studied. The combination of agent-based transport models with the concept of MFD represents an opportunity to develop efficient schemes from a social welfare perspective in line with innovative theories from traffic theory.

Agent-based models are chosen because they offer the possibility to reach high levels of realism of the model thanks to the reliable representation of travel behavior (choices) in large scale road networks with several thousand agents. Furthermore, the high disaggregation-

heterogeneity of the agent-based models offer the opportunity to investigate more in depth issues such the distributional impacts and acceptability of congestion pricing schemes.

The development of tolling schemes based on the MFD allows a more reliable reproduction of the supply side (performance of the system) and in theory a more accurate fares. In this particular study, the derivation of variable fares through the combination of the MFD with the classic MCP concept represents an additional effort to tie the engineering and economic approach. Finally, the concept of spread of density explicitly considered as additional dimension in the pricing model allows a deeper understanding of the congestion phenomena.

study	behavioral considerations	measures	evaluation	objective focus
Pigou (1920)	only car users; limited heterogeneity (VOT is an average of the population)	tax on congestion on a single road	static (MPC model)	economical: maximize net welfare
Walters (1961)	only car users; limited heterogeneity (VOT is an average of the population)	tax on congestion on a network of highways	static (MPC model)	economical: maximize net welfare
Vickrey (1969)	only car users; all the travellers have the same VOT; departure time decision	time dependent toll on a single road	dynamic: increase of demand during the peak period (Bottleneck Model)	economical: ease congestion
Daganzo (1985)	only car users; all the travellers have the same VOT; departure time decision	time dependent toll on a single road	dynamic: increase of demand during the peak period (Bottleneck Model)	economical: ease congestion
Newell (1987)	only car users; different cost functions relative to their work starting time	time dependent toll on a single road	dynamic: increase of demand during the peak period (Bottleneck Model)	economical: ease congestion
Arnott, de Palma et al. (1993)	treat explicitly users' behavioral decisions; elastic demand (price sensitive)	four different pricing regimes: no toll/coarse toll/step toll/fine toll	dynamic: increase of demand during the peak period (Bottleneck Model)	economical: determining gains of efficient pricing
Yang, Meng (1998)	elastic demand; departure time + route choice	toll on congestion on the network	dynamic: Dynamic Traffic Assignment; STEN	engineering/economical: optimal pricing in general queuing network to reach System Optimum
Verhoef (2002)	elastic demand; heterogeneity limited to different willingness to pay	toll on a sub-set of links/toll points of a network	static (MPC model)	economical: derive optimal tax
Levinson and Geroliminis (2009)	only car users; all the travellers have the same VOT; departure time decision	cordon-based pricing (time varying toll)	dynamic: throughput of bottleneck depends on level of congestion (bottleneck model+MFD)	engineering: eliminate delay
Gonzales and Daganzo (2012)	PT sharing the bottleneck with cars; +mode choice	pricing of a neighbourhood network	dynamic: throughput of bottleneck depends on level of congestion (bottleneck model+MFD)	engineering/economical: reach system optimum and derive system optimal pricing of cars and PT
Small (2012)	only car users; departure time decision	varying optimal toll in downtown a	dynamic: throughput of bottleneck depends on level of congestion (bottleneck model+MFD)	engineering/economical: reach system optimum
Zheng, Geroliminis et al. (2012)	additional reality of travel behavior: individual utility function, individual VOT, "interaction" with other users; +route choice	cordon-based pricing (flat toll)	dynamic: agent-based traffic model + MFD	engineering: optimize traffic performance of the network; economic: produce social gain

Figure 3.11: overview of major research papers on congestion pricing model from a static-dynamic perspective

## 4. The Macroscopic Fundamental Diagram

Providing a correct description and predicting traffic dynamics in complex transportation networks is nowadays an overriding priority to allow fluid traffic operations. The idea of a Macroscopic Fundamental Diagram (MFD) to describe network traffic conditions at aggregate level has become in recent years a well-established theory in the scientific community. The aim of this chapter is to describe the concept of MFD and passing through some of the major publications, discuss the critical points and main challenges encountered. In Section 4.1 the basic principles of the MFD are introduced and short historic overview is provided. Section 4.2 provides a more detailed description of the derivation of MFD. Particularly, some theoretical and empirical studies that allowed the theory to entrench in the academia are presented in Section 4.3. In Section 4.4 the question of hysteresis loops is discussed by looking at the different studies aimed at explaining this phenomenon. The topic of traffic distribution and the relationship with traffic performance in the context of the MFD is described in Section 4.5. Finally, Section 4.6 discusses some of the issues considered in this study.

### 4.1 The MFD in a nutshell

The Macroscopic Fundamental Diagram (MFD) in a few words, represents how vehicles travel through a network as a function of the number of vehicles in that network. The MFD extends the concept of Fundamental Diagram (FD) that relates traffic flux and traffic density of single roads to large clusters of links.

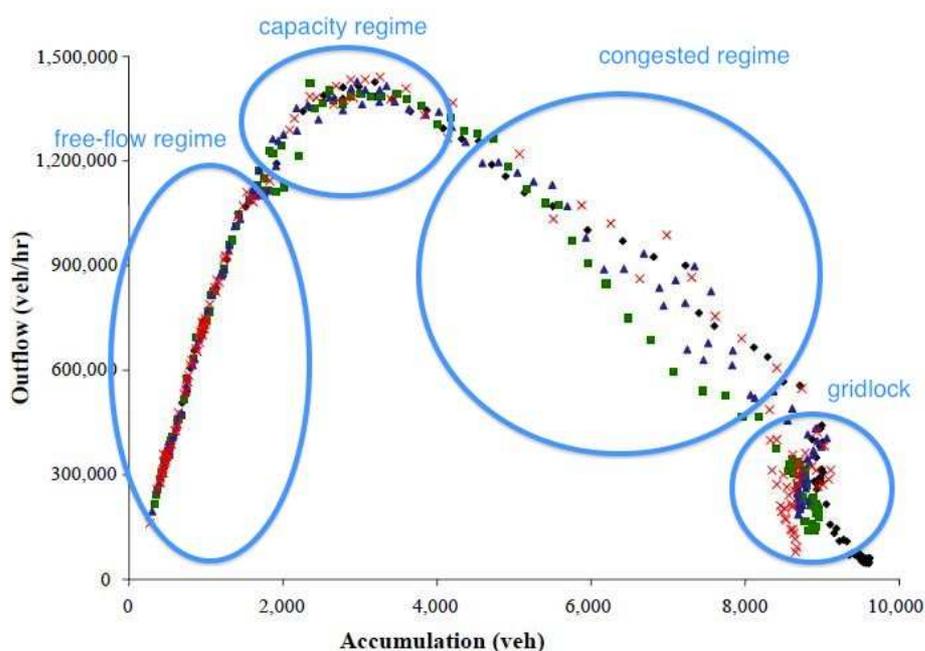


Figure 4.1: MFD and traffic regimes . Adapted from Geroliminis and Daganzo (2008)

As it possible to see in Figure 4.1, the diagram expresses the relationship between aggregated flow of all the links in the network (production) and the total amount of vehicles in the network (accumulation) by means of a concave function. The relationship exhibits a clear maximum representing to the maximum throughput of the network (capacity) to which corresponds a critical value of accumulation (critical density). Like in the FD, different traffic regimes can be identified on the diagram. On the left branch where production and accumulation have linear dependence almost all links of the network are in free flow condition (free-flow regime). The region characterized by slower increase of production with load until the capacity is reached is usually referred as capacity regime. If the critical density is exceeded, the network becomes heavily congested (congested regime) and as a result the production decreases with further increases of accumulation until the theoretical breakdown of the system corresponding to total gridlock. At this point all the vehicles in the network are completely standstill.

The relationship between production and accumulation has been expressed in different ways in the studies on the MFD. The former that is also referred as “Network flow” or “Performance” can be identified as average flow calculated in vehicle/hour or exit function (further explanations can be found in the next section). Depending on the complexity of the network in terms of length of the links and number of lanes, the production can be estimated as a simple or weighted average, which provides a more accurate description. The latter can be expressed in terms of vehicle/kilometres, vehicle/lane/kilometres and occupancy of detectors. Accumulation can be calculated as a simple or weighted average as well.

The idea of a macroscopic relationship between average flow and density was introduced by Godfrey (1969) who linked the number of vehicles circulating in a network and their average speed. Several studies in the following thirty years attempted to

Although during the following thirty years several research studies ((Ardekani and Herman, 1987 and Olszewski et al., 1995; Williams et al., 1987, Mahmassani et al., 1987 and Mahmassani and Peeta, 1993) from Geroliminis and Sun, (2011a)) have been carried out in order to explore more in depth this relationship, only recently the existence of an invariant macroscopic relation between network average flow, average density and average speed have been confirmed and formally expressed by Daganzo (2007; 2008) and Geroliminis (2007; 2008). Nowadays, this concept is catching on in the scientific community and it has given rise to a series of studies aimed at more detailed analysis of the topic and possible applications in the field of traffic management and control.

## 4.2 Original formulation of the MFD

Daganzo (2007) demonstrates the existence of macroscopic relation by deriving an exit function for single and multiple reservoir systems based on the following main experiment. An homogeneous looping road is characterized by the following assumptions:

“all the trips are internal (endogenous traffic); origin flows are uniformly distributed along the link and have priority; the average trip length is the same for all origins; [...] steady states  $(k, q)$  with the local density and internal flow should be uniform everywhere [...] and satisfy  $q = Q(k) = Q(n/l)$  (4.1) (where  $n$  is the number of users and  $l$  the length of the ring) for some unimodal FD that would capture the interference of entering and leaving flow” (p. 52).

The author defines the total distance travelled per unit time in two ways: as a product of the exit flow ( $g$ ) and the distance travelled by exiting vehicles ( $d$ ); and as sum of distances travelled by vehicles in the system at any given time:  $n \cdot v = l \cdot k \cdot v = l \cdot q = l \cdot Q \cdot (\frac{n}{l})$  (4.2). From the combination of the two, he derives  $g = l/dQ(\frac{n}{l})$  (4.3) that proves the linear relationship between exit flows and circulating flows.

Furthermore, the exit function, although it describes steady state, can be used for dynamic analysis. Under the assumption of slow transitions between steady states, the output can be predicted as:  $\frac{dL(t)}{dt} = g(t) = G(n(t))$  (4.4)

Then, Daganzo extends the abovementioned ideas to a city network composed by multiple reservoirs subject to gridlock. First he considers homogeneous systems that can be studied in the aggregate as single reservoirs. After, he considers inhomogeneous systems characterized by variable number of lanes and different input demands.

The author employs again a closed loop of length  $L_T$  for his demonstration and introduces the fundamental assumption that: “drivers looking for “opportunities” at the various exits take the first exit that satisfies their needs ... independently of where s/he comes from.”

Given these assumptions, the actual outflow in an interval  $(x, x+dx)$  can be expressed as a product of the circulating flow  $Q(k(x), x)$  and the fraction of vehicles  $p(x)dx$  exiting in the interval, as it follows:

$$g = \int_0^{L_T} Q(k(x), x)p(x)dx \quad (4.5)$$

The author finally derives that the two following insights: “If conditions do not change rapidly with the time, a road should not simultaneously have congested and uncongested portions”; “If

conditions do not change rapidly with time, system output is maximized when flow is at capacity only along road stretches with the greatest exit rates  $p(x)$ ; i.e. the greatest density of destinations,  $\lambda(x)$ ". As a result of that, confining clusters of congestion to locations where exit rates are low could be an efficient strategy for storing queues in congested roads.

Daganzo and Geroliminis (2007) recast this theory in terms of two postulates: (i) inhomogeneously congested "neighborhoods" exhibit an MFD relating "production" and "accumulation"; (ii) trip completion rate is proportional to the production.

### 4.3 Further developments of the theory

The first theoretical studies were followed soon by other research aimed at demonstrating the existence of the MFD, still produced by Daganzo and Geroliminis. In this section, two main works, both characterized by the same research objective are characterized. The proof of existence of the MFD in real cities is described in Sub-section 4.3.1. The theoretical approach focused on the description of the shape of the MFD is presented in Sub-section 4.3.2.

#### 4.3.1 The empirical evidence of the MFD

Daganzo and Geroliminis (2008) provided the first empirical proof of the existence of the MFD. Indeed, by means of detector data and taxi-paths collected in Yokohama, it was possible to determine relationship between the network average density and flow in line with the theory (Figure 4.2). Furthermore, a linear relationship between trip completion rate and total production could be derived.

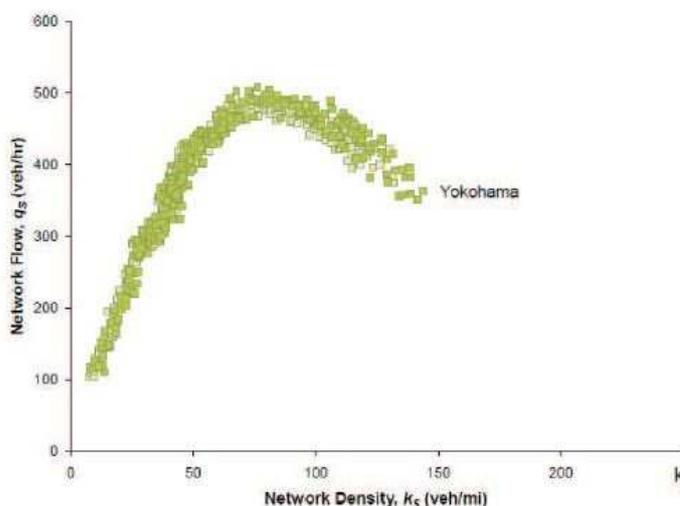


Figure 4.2: Macroscopic Fundamental Diagram of Yokohama. Source: Daganzo and Geroliminis (2008)

Finally, the two authors proved that the shape of the MFD is not affected by the demand, by analyzing different days and time periods (corresponding to different origin-destination tables). The importance of these findings is crucial in terms of implementation of large-scale traffic

management measures relying on real-time observations (in contrast with elaborated and burdensome prediction models).

#### 4.3.2 The demonstration of the existence of the MFD

Based on the previous studies, Daganzo and Geroliminis (2008) managed to describe the shape of the MFD with a theoretical approach based on the Variational Theory (VT). The authors apply “cuts” on the flow-density plane to define the outer boundaries of the MFD. The combination of three typologies of cuts defined according to the different speed and direction of the moving observer (forward, stationary and backward) results in the theoretical upper bound shown in Figure 4.3.

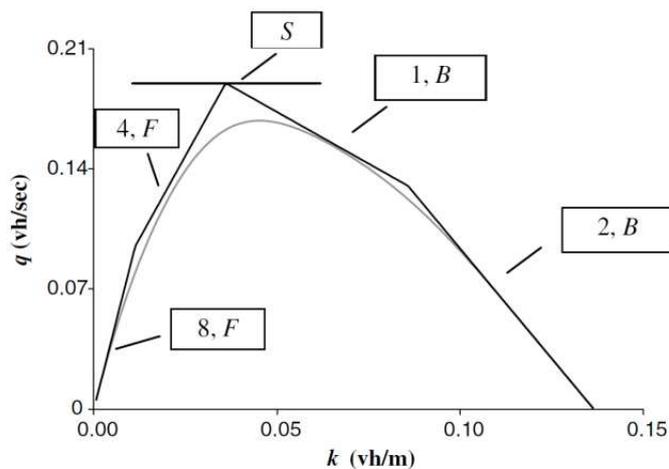


Figure 4.0.1: Theoretical cuts of the MFD. Source Daganzo and Geroliminis (2008)

In article, some important theoretical considerations concerning the ‘regularity conditions’ (sufficient but not necessary) to ensure a well-defined MFD are discussed as well. These conditions are expressed as:

“(i) a slow-varying and distributed demand; (ii) a redundant network ensuring that drivers have many route choices and that most links are on many desirable routes (iii); a homogeneous network with similar links (iv); links with an approximate FD that is not significantly affected by turning movements when flow is steady”.

As Buisson and Ladier (2009) notice, the above conditions imply the necessary redundancy to reach a Wardrop’s equilibrium and the homogeneity of links. Precisely, redundancy allows drivers to choose routes adaptively and since route are similar, the densities on each route should then be similar. This would imply that turns into and out of each route are steady and balanced. This ultimately entails that: the MFD can accurately describe the average flow in a network given the amount of vehicles in the network; it depends neither on the origins and destinations, nor on the drivers’ individual route choices.

#### 4.4 Hysteresis loops and network dynamics

Buisson and Ladier (2009) are the first to relax the ‘regularity conditions’ in their investigation about the MFD of the city of Toulouse. Data are collected from loop detectors of a network encompassing a large variety of roads including highways and residential roads. The authors investigate the influence on the MFD of aspects such as the homogeneity of data measurement location, the homogeneity in the typology of selected roads, and the spatial evolution of the demand. In this study the resulting MFD does not exhibit any congested branch and it presents high scatter forming hysteresis loops (Figure 4.4). Such a phenomenon can be explained by the spatially heterogeneous evolution of congestion, which is likely related to the presence of slow-moving platoon of trucks. The authors also notice that the MFD derived from highways present a triangular shape rather than the trapezoidal one theorized by Daganzo and Geroliminis (2007) probably because of the absence of traffic signals. Furthermore, they found that in a signalized network, the distance between loops and traffic signals influences the shape of the MFD. The authors conclude that, since shape of the MFD is also correlated with the type of network, reservoirs should be identified not only on geographic basis, but also accounting for the type of road. The experiments actually highlighted the difference between penetrating highways and ring road in terms of scatter and hysteresis phenomena.

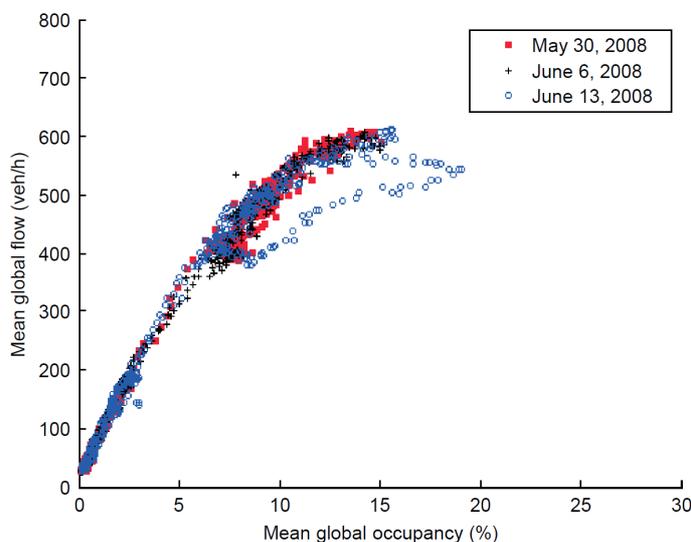


Figura 4.4: scattered MFD. Source: Buisson and Ladier (2009)

Soon after, a series of studies mainly (co)authored by Daganzo, Geroliminis and Mahamassani further investigated the effect of the hysteresis.

Gayah and Daganzo (2010) gain a deeper understanding of hysteresis phenomenon by analyzing the qualitative behavior of a two-bins system. In this study, clockwise loops characterized by higher inflows during the onset of congestion are observed and related to the

drivers' adaptivity. The results suggest that, when drivers are less adaptive "small imbalances in the system grow and clockwise hysteresis loops arise".

Daganzo *et al.* (2011) demonstrate that even in the "most favorable scenario possible for the MFD" (perfectly homogeneous, time independent and not granular), the MFD is characterized by the presence of multiple traffic states for the same value of density. This phenomenon is defined by the authors bifurcation. By using a two-bin system, the authors show that even small disturbances determined by turning maneuvers can make the system unstable and generate gridlocks. No matter how well balanced the amount of turning is, the network will always breakdown sooner or later with higher probability corresponding to increasing density and turning rate.

Geroliminis and Sun (2011b) investigate in depth the reasons of hysteresis loops by analyzing real data of a freeway network. The authors identify to major causes behind this phenomenon. First, the degree of spatial heterogeneity is different at onset and offset of congestion. Typically at the onset the outflow is higher than in the offset as the traffic is better distributed. For this reason, in a system characterized by presence of hysteresis loops, the knowledge of the accumulation is not sufficient to allow prediction of the outflow. In order to do that, information on the path followed by the network density before reaching the current value is necessary. A second reason of hysteresis can be explained by the synchronized occurrence of transient periods and capacity drop in the offset of congestion.

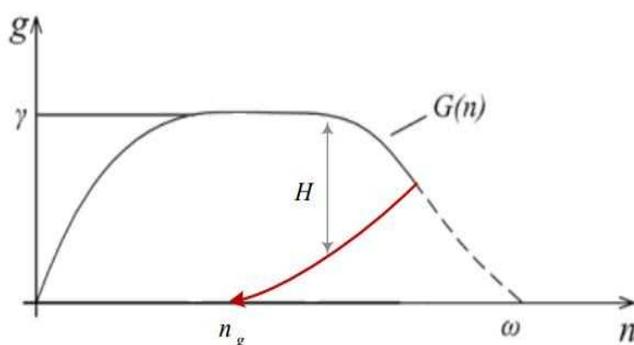


Figure 4.5: Hysteretic Network Exit Function. Source: Mahmassani *et al.* (2013).

An interesting theoretical approach to deal with the presence of hysteresis in the determination of the network performance has been recently proposed by Mahmassani *et al.* (2013). In this study the authors argue that "Daganzo's exit function" is an idealized description of the equilibrium behavior that holds only in case of slow changes of the input and homogeneous distribution of traffic. Then, in order to account for the hysteresis phenomenon, an hysteretic Network Fundamental Diagram (Figure 4.5) is derived from a formula that combines the output

flow  $G(n)$  and the deviation from steady-state due to the hysteretic behavior ( $H$ ). This component depends on both the accumulation and the spatial inhomogeneity of congestion distribution.

#### 4.5 The influence of traffic distribution on the network performance

An important progress in the study of the aggregated traffic relationship of road networks has been made in 2010 by Mazlounian and Geroliminis. The two scholars were the first ones to formally recognize a relationship between the inhomogeneity of traffic conditions, the spatial distribution of density and scatter of the MFD. The authors observe that congestion in urban networks is by nature inhomogeneous in time and space because of several reasons like the clustering of demand inside the network and differences in the typology of roads and control. For this reason, they investigate how the existence, shape and scatter of the MFD are affected by the spatial distribution of traffic. In particular, a macroscopic relation among these variables is sought. By means of two innovative modeling techniques, macroscopic flow quantization and memory less traffic flow routing, several tests with increasing realism are performed on an artificial city centre represented by a lattice-like uni-directional road network. The authors demonstrate that a unique monotonously falling relationship exists between the average flow and the standard deviation of the number of vehicles in all the links in the network (Figure 4.6 a). For this reason, there is a wide variation of network production of the MFD that exhibits scattering. Furthermore, the amount of full links, which is related to the uneven distribution of congestion, is identified as the major responsible for the decrease the performance (Figure 4.6 b). Indeed, inhomogeneity of spatial distribution of density increases the probability of spillover that ultimately decrease the network flow. As a consequence, traffic managers should reduce the variability of density to enhance traffic performance.

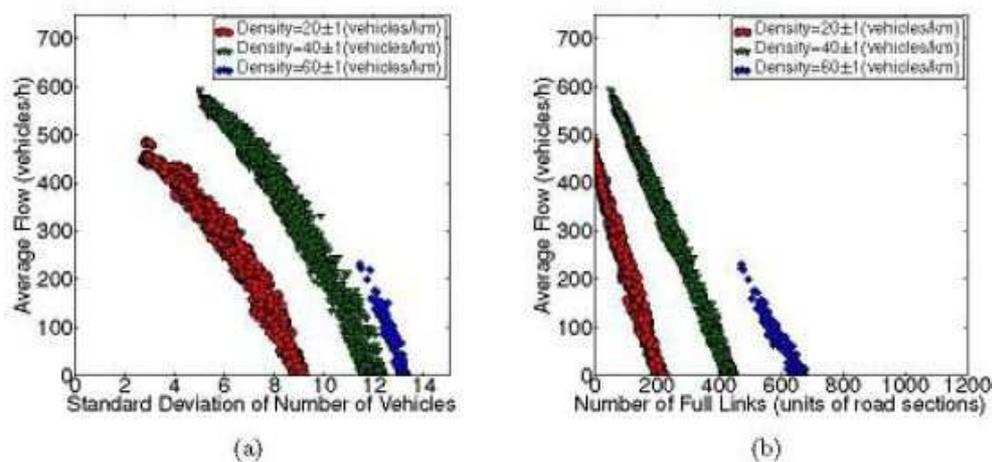
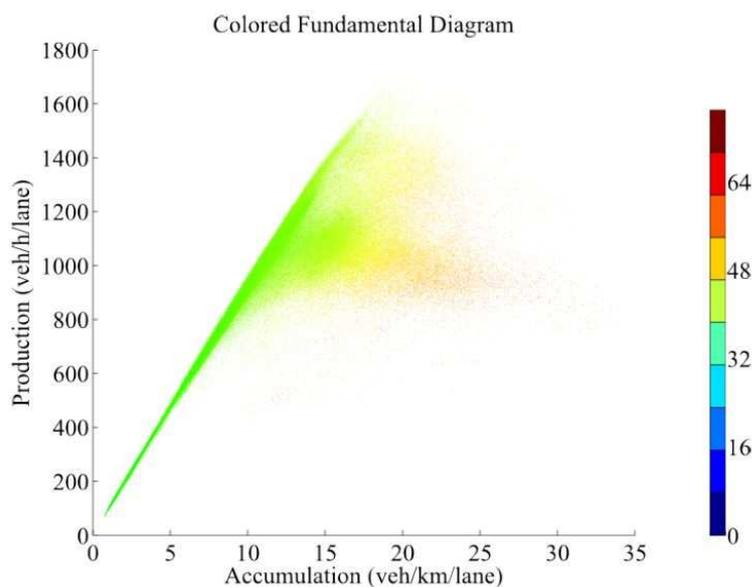


Figure 4.6: relationship between production and distribution of vehicles, and amount of full links. Source: Mazlounian and Geroliminis (2010)

On the same page, Geroliminis and Sun (2011a) investigated the relationship between properties of a network and the MFD by analyzing the Yokohama case study. The authors show that: “if the spatial distribution of link density is the same for two different time intervals with the same number of vehicles in the network, then these two time intervals should have the same average flow” (p. 617). This is a sufficient condition for the existence of a MFD with low scatter. Another interesting finding is that, in the case of Yokohama, the coefficient of variation given by the ratio of standard deviation and mean flow, is almost constant for occupancies (density) above 20%. Then, according to the authors a relationship between the level of spatial heterogeneity and the variance of the MFD can be derived if flow’s CoV is considered constant. Such a relationship might provide a useful hint about the possibility to reproduce a well-defined MFD in different cities.



**Figure 4.7: Generalized Macroscopic Fundamental Diagram. Source: Knoop and Hoogendoorn (2013)**

Another recent contribution came from Knoop and Hoogendoorn (2013) who analyzed real traffic data of an urban freeway network around Amsterdam to investigate the effect of inhomogeneity on the traffic production. The authors highlight that near the maximum production of the MFD, there is high scatter because of the mix of different traffic regimes (congested, uncongested). This phenomenon is not the result of a random combination, but it obeys some important properties. The spatial spread of density increases with a larger accumulation and more importantly, for the same accumulation, higher spread corresponds to lower production (Figure 4.7). In one sentence, the spatial spread significantly affects the production, especially near its critical value. Then, Knoop and Hoogendoorn (2013) derive a quantitative description of traffic production from the accumulation and spatial spread of

density, called generalized macroscopic fundamental diagram (GMFD). This function, which is approximated as a third order polynomial, is tested and found to describe the relationship as good as a non-parameterized average. The GMFD presents a similar concave shape to the MFD and can be used for traffic control strategies like ramp metering and dynamic speed limits.

#### **4.6 This research study**

In this study the aggregated traffic relationship for the urban network of Zurich will be analyzed in order to develop effective congestion pricing strategies controlled by the MFD. The investigations will be carried out by means of an agent-based transport model (MATSim). A previous study (Zheng *et al.*, 2012) has shown encouraging results in terms of consistency of the model with traffic flow theory and effectiveness of the pricing measure implemented. This research aims at gaining additional understanding of macroscopic traffic phenomena and providing different traffic management solutions to cope with them.

The simulation of a complex traffic network characterized by irregularity in the number of lanes, link lengths, counter flows and bottlenecks like bridges and tunnels, offers the opportunity to investigate more in depth the relationship between spatial distribution of congestion and traffic performance of the system. Since previous studies have demonstrated that decent MFDs can be derived even in heterogeneous networks and accurate results can be obtained when considering the variance (spread) of density, this study will follow the same direction.

Furthermore, alternative pricing strategies characterized by different traffic management approaches will be implemented to investigate their effects on the traffic performance of the network and reduce the levels of congestion. Particularly, questions regarding the impacts of pricing measures aimed at reducing the accumulation and distribution of congestion will be made during the design phase.

## 5. Set-up of the experiment and model framework

The previous chapters offered the necessary theoretical background to develop congestion pricing schemes based on the MFD aimed at achieving the efficiency, a fair distribution of benefits and ultimately high levels of public acceptability. Chapter 5 brings the discussion to a more practical level in order to develop alternative tolls within a multi-agent based transport model coordinated by the MFD. A description of MATSim and the studied scenario is provided in Sections 5.1 and 5.2. Then, the derivation of macroscopic traffic variables is presented in Section 5.3. Section 5.4 investigates the effects of certain parameters on the traffic reproduction of the simulation. In Section 5.5 the macroscopic traffic characteristics of the city centre of Zurich are studied. Sections 5.6 and 5.7 focus on the analysis of these properties, investigate traffic flow phenomena such as the hysteresis and the scatter of the accumulation-production relation and finally derive a generalized macroscopic fundamental diagram. To conclude, three different tolling schemes based on the previous findings are developed in Section 5.8.

### 5.1 Overview of MATSim: a Multi-Agent Transport Simulation

MATSim represents traffic behavior at highly disaggregated level by modeling individual agents. A schematic overview of the simulation structure is illustrated in Figure 5.1.

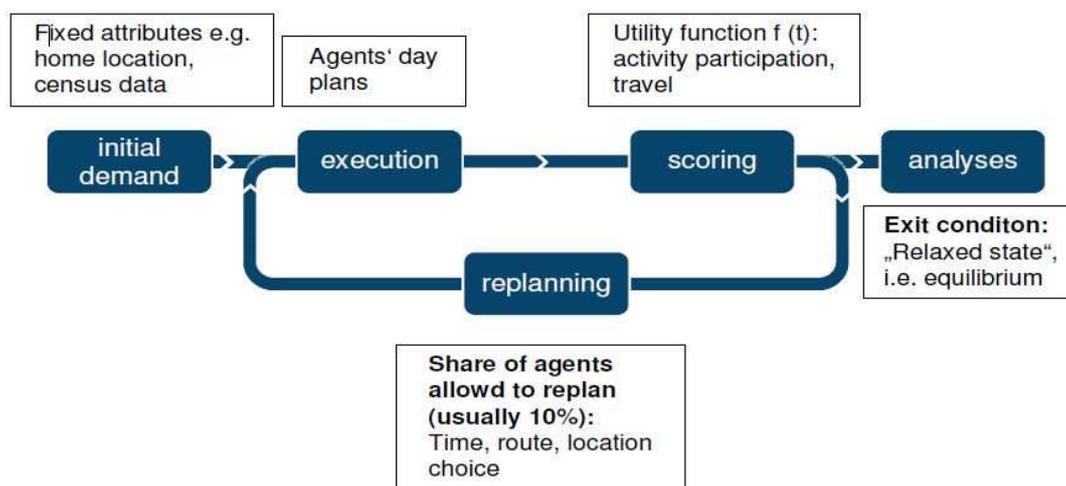


Figure 5.1: MATSim simulation structure Source: [www.matsim.org](http://www.matsim.org)

As Kickhofer *et al.* (2011) highlight, the overall process can be described by the following main features:

- Each agent independently develops a plan that expresses its preferences in terms of activities and their schedules during the day (Plans).

- The agents simultaneously perform all the plans in the physical system in the mobility simulation (Execution)
- To compare the performance of different plans, each one is associated to a score given by a utility function (Scoring)
- Agents are able to memorize their plans and during the simulation improve them by means of a learning algorithm (Replanning). During the implementation the system iterates between plans generation and traffic flow simulation
- The cycle continues until the system has reached an equilibrium where no agent can improve anymore his score (Analyses). This state is often referred as “relaxed”.

### 5.1.1 Development of plans

Plans contain the itinerary of the activities that need to be performed during the day and the necessary trips to reach the activity locations (facilities). Each plan contains details about the typology and location, duration, time constraints and chronological order of each activity. Trips are expressed by mode, route, expected departure and travel time. Every plan can be associated to a score given by the gains of performing the activities (at the desired times) and losses of travelling (travel costs and possible PT fares or tolls). Plans can be improved by changing the time of departure, varying the route and choosing different transport mode through modules. The Time Adaptation module changes the timing of the agents' plans by randomly mutating the duration of agents' activities (Balmer *et al.*, 2005). The Router is a time-dependent best path algorithm (Kickhofer *et al.*, 2011; Lefebvre and Balmer, 2007) that uses link travel times of the previous iteration to derive the link's generalized costs. Plans and their respective scores can be stored in the system so that at every iteration new ones can be created by modifying the old ones. The simulation works as an iteration cycle where the outputs of the traffic simulation influence modules decisions to update plans. This feedback cycle is controlled by the agent database (Kickhofer *et al.*, 2011) where multiple plans generated by the agent are stored. In every iteration each agent decides which plan to execute from its memory from the following options (Meister *et al.*, 2010): With a probability of  $p_{replan,m}$  the agent is chosen to replan a random plan of his memory and execute it in the next iteration by means of the replanning module  $m$ . While the probabilities  $p_{replan,m}$  may have different values, and may differ from iteration to iteration, in this study the same value is used for all replanning modules across all iterations. For the rest of the agents, the probability to change the selected plan is based on the utility of the plan and it is calculated as:

$$p_j = \frac{e^{\beta \cdot U_j}}{\sum_i e^{\beta \cdot U_i}} \quad (5.1)$$

where  $p_j$  is the probability for plan  $j$  to be selected among  $i$  plans of a choice set  $I$  and  $\beta$  an

empirical constant. The choice model used in this paper is thus equivalent to the standard multinomial logit model.

Since the amount of plans in the memory of agents is limited, at each iteration the worst performing one is replaced by the new one. Thanks to this feedback mechanism agents are able to improve their plans over several iterations until the system reaches the “relaxed” state when agents cannot significantly improve their plans and the outcome of the system becomes stable. This state is also referred as agent-based stochastic user equilibrium (Nagel and Flotterod, 2009).

Once all the agents’ plans of every agent are selected, the traffic flow simulation executes them simultaneously on the network and returns as output the events occurring during that iteration. The car traffic simulation is implemented as a queue simulation where each road segment (link) is modeled as a First-In First-Out queue with a minimum service time given by the length of the link divided by the maximum travel speed (Zheng *et al.*, 2012). The capacity of links is a predetermined value (in other models it is derived from the maximum outflow which is influenced by the amount of vehicles (density) and their interactions). The storage capacity of the link limits the number of vehicles that can enter the link. Additional details about the traffic simulator will be provided in the following sections (Section 5.1.3 and 5.4).

### 5.1.2 The utility function

Every plan is associated to a quantitative score, also referred as utility, determined as it follows:

$$U_{total} = \sum_{i=1}^n (U_{act,i} + U_{travel,i} + U_{wait,i} + U_{short,i}) \quad (5.2)$$

Where  $U_{total}$  represents the total utility for a given plan;  $n$  is the number of activities;  $U_{act,i}$  is the (positive) utility earned for performing activity  $i$ ;  $U_{travel,i}$  is the (usually negative) utility earned for traveling during the trip  $i$ ;  $U_{wait,i}$  is the (negative) utility earned for waiting instead of performing an activity  $i$ ; and  $U_{short,i}$  is the (negative) utility earned for performing an activity  $i$  for a too short duration.

The positive utility earned by performing the activity  $i$  is derived from the following algorithm:

$$U_{act,i} = (U_{c,j} - U_{c,j-1}) \cdot f_p \quad (5.3)$$

Where  $U_{c,j}$  is the cumulative score of all the activities of the same type as activity  $i$ , of which  $i$  is the  $j$  instance when ordered ascending temporally.  $f_p$  denotes a factor which may negatively influence the score of a shop or a leisure activity if the related facility is too crowded (Horni *et al.*, 2009). This cumulative score is given by the following equation:

$$U_{c,j} = \begin{cases} \max(0, \beta_{perf} \cdot t^* \cdot \ln\left(\frac{\sum_{k=1}^j t_{perf,k}}{t_o}\right)) & j > 0 \\ 0 & j = 0 \end{cases} \quad (5.4)$$

$\beta_{perf}$  is the marginal utility of another unit of time spent performing an activity (the marginal utility is the same for all the activities that differ from each other by the opening times and typical duration).  $t^*$  is the externally defined desired time budget of the agent for spending time performing one or more activities of the same type as activity  $i$ .  $t_{perf,k}$  is the time spent performing the  $k^{th}$  activity of the same type as activity  $i$ .  $t_o$  is the zero of the logarithmic function and is proportional to  $t$ . In other words, this scoring function cumulates scores by activity type and does not require a time allocation between several activities of the same type beforehand. The advantage is that one does not have to define a large set of parameters for desired activity durations as it was necessary in previous studies (Meister *et al.*, 2010).

The (dis)utility of waiting corresponds to:

$$U_{wait,i} = \beta_{wait} \cdot t_{wait} \quad (5.5)$$

Where  $\beta_{wait}$  is the marginal utility of waiting time,  $t_{wait}$  and is the waiting time  $i$ .

The penalty for performing activity  $i$  shorter is:

$$U_{short,i} = \beta_{short} \cdot \max(0, 0.5h - t_{perf,i}) \quad (5.6)$$

with  $\beta_{short}$  being the marginal utility of another unit of time which is missing to the minimum activity duration of 0.5h, and  $t_{perf,i}$  being the time spent performing activity  $i$ . This penalty is necessary in order to prevent that activity  $I$  is assigned the entire desired activity duration if there are further activities of the same type as activity  $I$ , which would possibly have a duration of 0s.

The (dis)utility of travelling to the activity  $i$  by transport *mode* is assumed as:

$$U_{travel,i,mode}(t_{travel,i}) = U_{access-egress,i,mode} + \beta_{tt,mode} \cdot t_{i,mode} + \beta_{cost,mode} \cdot c_{i,mode} \quad (5.7)$$

with  $\beta_{tt,mode}$  being the marginal utility of another unit of time traveling by mode,  $\beta_{cost,mode}$  being the marginal utility of another unit of money spent on traveling by mode.  $t_{mode}$  denotes the in-vehicle travel time, and  $c_{mode}$  is the expenditure of money required for traveling with mode.  $U_{access-egress,mode}$  denotes the (usually negative) score of access/egress to/from a vehicle if the mode is different than walk. The related costs for access and egress for the modes car and bikes is derived as:

$$U_{access-egress,car} = U_{access-egress,bike} = (\beta_{tt,walk} - \beta_{perf}) \cdot (t_{access,car-bike} + t_{egress,car-bike}) \quad (5.8)$$

For the mode *walk*, the value of this score constant is zero. This is also the case for the mode *pt*, where access to and egress from the closest pt stop are explicitly part of the route description.

As concern the disutility of arriving early, no explicit disutility is associated to it since: “waiting times are already indirectly punished by foregoing the reward that could be accumulated by doing an activity instead (opportunity cost)” (Kickhofer *et al.*, 2011). Hence, the disutility of waiting can be expressed as  $-\beta_{perf}$ .

The standard values used for our case study are derived from Meister *et al.* (2010).

In particular, the marginal utility values for performing an activity, waiting and travelling by car are derived from parameter estimation exercise for a bimodal (car/pt) MATSim study of Switzerland (Kickhöfer, 2009) and corresponds to:

- $\beta_{perf} = 2.26/h$ ;  $\beta_{tt,car} = 0.0/h$ ;  $\beta_{wait} = 0.0/h$ ; (from Kickhofer, 2009)

As long as  $\beta_{perf} > \beta_{tt,car}$  and  $\beta_{perf} > \beta_{wait}$  the opportunity costs of traveling by car and waiting are negative, which is an incentive for the agent to spend as much time as possible performing activities.

- $\beta_{short} = -180.0/h$ ;

Performing an activity shorter is penalized very hard, compared to the value of  $\beta_{perf}$ , in order to maintain minimum durations in the fixed activity chain.

The values of the marginal utilities for spending time traveling by modes other than car, as well as the marginal utilities for monetary expenditures of all modes are the result of a manual calibration procedure.

- $\beta_{tt,pt} = -2.0/h$ ;  $\beta_{tt,bike} = -16.0/h$ ;  $\beta_{tt,walk} = 0.0/h$ ;
- $\beta_{cost,car} = 0.0/h$ ;  $\beta_{cost,pt} = -0.8/h$ ;  $\beta_{cost,walk} = -0.1/h$ ;

The average monetary expenditures per kilometer for motorized modes of transport are based on Swiss values documented in Vrtic *et al.* (2007b):

$$c_{km,car} = 0.12 \text{ CHF/km}; c_{km,pt} = 0.28 \text{ CHF/km};$$

The speed of pt transport is estimated by means of matrix derived from a survey, while the speed of bike and walking is set to:

$$v_{bike} = 14.0 \text{ km/h}; v_{walk} = 2.8 \text{ km/h};$$

The access and egress time is set to:

$$t_{access,car/bike} = t_{egress,car/bike} = 5 \text{ min}$$

### 5.1.3 JDEQSim: a queue-based traffic simulator

MATSim is a model in constant development and during its evolution several changes have been made in terms of scenarios, replanning modules and traffic models. As a result, during the last years different versions of simulations characterized by alternative traffic models have been implemented. In this specific study the employed traffic simulator consists of JDEQSim (Charypar et al., 2007; Waraich et al., 2009).

JDEQSim, which stands for “Java Deterministic Event-Driven Queue-Based Traffic Flow Micro-Simulation”, derives from classic queue-based microsimulations. The main assumption of these models is that intersections alone determine the main features of traffic (Charypar et al., 2007). Here, links can store cars travelling through them and collaborate with each other in order to move car through the networks according to several constraints like capacity, free speed travel time, intersection precedence and space available at the next link. Queue-based models allow higher computational performances than cellular automata as the number of simulated units is smaller (links instead of cars).

An important feature of JDEQSim consists of the event-based approach instead of the traditional queue-based one. While in the latter each link is simulated in every time-step even though no cars are present on it, in the first one only links where an event (entering or leaving car) occurred are processed. The gains in terms of processing time are evident. A similar event-based approach is used in METROPOLIS (De Palma and Marchall, 2002).

Other additional features have been implemented in JDEQSim in order to make the simulation more realistic. For example, the presence of gaps travelling backwards to better reproduce phenomena of acceleration and deceleration between vehicles in the queue. As a consequence, when the front car in a queue starts driving it leaves behind a gap that the cars behind will need to wait for in order to start driving. The speed of gaps is given as a parameter that can be modified in the simulation settings (see Subsection 5.4.2).

In the real world, when a gridlock occurs, vehicles are able to interact in such a way to resolve it by changing routes, turning back, doing unusual maneuvers or cooperating with each other as Mahmassani *et al.* (2013) observe. In the simulation instead, when this phenomenon occurs, vehicles might wait for each other forever. In order to avoid this situation cars are temporarily

allowed to enter the consecutive link (even if its capacity is exceeded). A threshold parameter called “squeeze time” corresponding to the amount of time spent waiting in the link before being able to move the next one is set to push stuck vehicles forward. Vehicles that are “squeezed forward” are then penalized with a very low score in order to reduce the intensity and number of gridlocks in the following iterations. On the other hand this expedient might compromise the accuracy of the traffic performance of the model if the network is severely congested.

## 5.2 Case Study Scenario

The simulation scenario (from Meister *et al.*, 2010) consists of an area of 30 km around the city of Zurich (Greater Zurich Area) represented in the figure below (Figure 5.2). Agents residing outside the study area, but entering at some time during the day are also included in the simulation (Waraich and Axhausen, 2011).

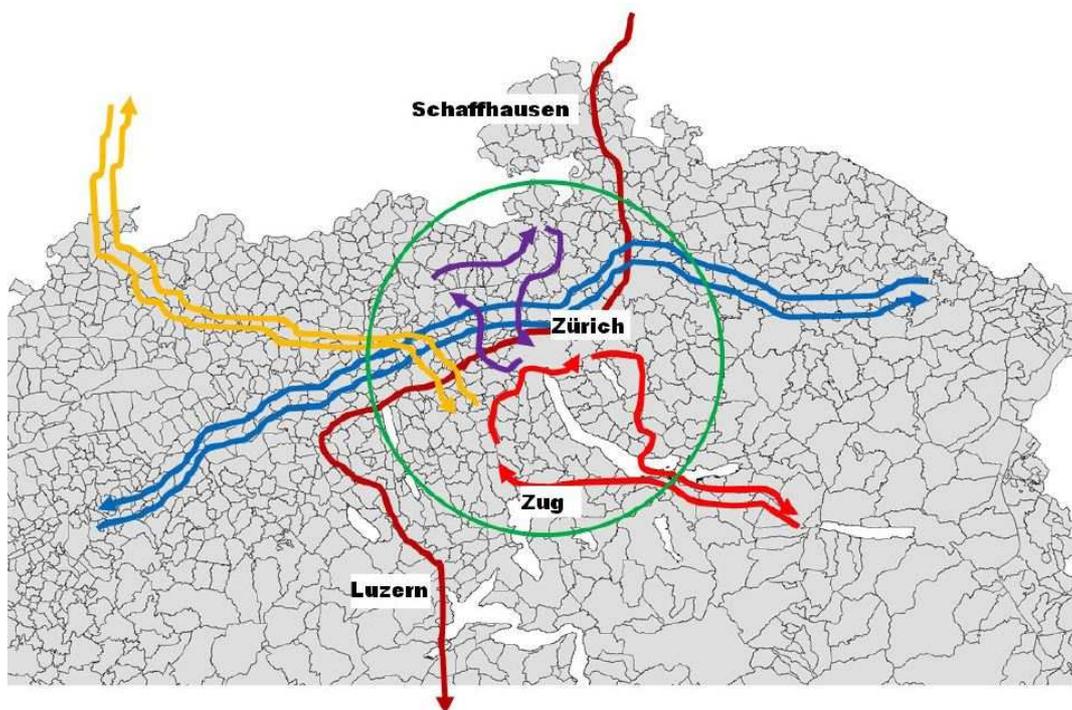


Figura 5.0.2: Simulation scenario. Source Ciari (2010)

The road network used consists of an high resolution navigation network including about 1.035.305 road segments (links) and 472.819 junctions (nodes). Instead of simulating the full population of agents, a sample of 10%, equivalent to 180.000 agents, is used for the experiments of this thesis. Normally, each of them needs to travel at least once per day to execute his plans. In order to deal with smaller samples, it is common practice to scale down link capacities to match them with the sample size. Although such a approach might generate side effects particularly in terms of traffic simulation, also known as “artifacts” among MATSim users they

do not compromise the quality of the experiments, if properly considered (see Section 4.4 and 4.5).

The available transportation modes in the simulation are car, public transport, bike and walk. Only cars are “physically” simulated along the roads, while the other modes are “teleported” from the origin to the destination. The duration of the PT trips is constant and corresponds to the origin-destination (OD) travel time derived from a matrix estimated for the whole metropolitan area of Zurich by surveys. The duration of walk and bike trips is calculated by means of average speeds (see Section 5.1).

### 5.3 Derivation of the macroscopic variables

In order to perform macroscopic traffic analyses, traffic accumulation (space mean density) and traffic production (outflow) data at single link level are required.

Data for every link are collected in interval of 15 seconds and after aggregated in larger intervals of 5 minutes.

The average density  $k_j$  for the single link is derived as:

$$k_j = k_{j-1} + \Delta k_j \quad (5.9)$$

Where  $k_{j-1}$  corresponds to the sum of the density derived at the time  $j-1$  and  $\Delta k_j$ , which is the change of density occurring during the time interval between  $j-1$  and  $j$ . Such a variation is calculated with the following formula that explicitly recalls the Cell Transmission Model:

$$\Delta k_j = \frac{e_j - o_j}{l \cdot n} \quad (5.10)$$

Where  $e_j$  indicates the number of vehicles entering the link;  $o_j$  indicates the number of vehicles leaving the link;  $l$  corresponds to the length of the link and  $n$  to the number of lanes of the link.

The outflow  $q_j$  is simply estimated as number of vehicles leaving the link during the same intervals:

$$q_j = o_j \cdot \frac{3600}{\Delta t} \quad (5.11)$$

Where  $\Delta t$  represents the time interval.

The average speed for single links is calculated by means of the well-known relationship of traffic flow theory with density and outflow:

$$u_j = \frac{q_j}{k_j} \quad (5.12)$$

The fact that speed is derived from the outflow might determine little imprecision when during a certain interval a vehicle drives into the link, but it exits during the following one. However, it should be noted that, this problem entails particularly long links (over the hundred metres) that are a minority. Furthermore, the problem is resolved by aggregating the measurements in larger intervals.

Finally, the average density and the average outflow of the network composed by  $n$  links can be determined as the following weighted averages of the  $l$  individual link values:

$$K = \frac{\sum_l^n (k_l \cdot l_l \cdot n_l)}{\sum_l^n (l_l \cdot n_l)} \quad (5.13)$$

$$Q = \frac{\sum_l^n (q_l \cdot l_l \cdot n_l)}{\sum_l^n (l_l \cdot n_l)} \quad (5.14)$$

The average network speed  $U$  can be derived as:

$$U = \frac{Q}{K} \quad (5.15)$$

The average speed could be calculated also by aggregating the average speeds of all the links. This estimate would slightly differ (particularly when the network is almost unused) from the previous one as free flow speeds of empty links would now be directly included in the estimation.

It has emerged from the literature review (see Section 4.5) that the deviation in density of the different links through the network plays also an important role in the shape of the MFD and it is considered as a cause of scatter by several scholars (Buisson and Ladier, 2009; Knoop and Hoogendoorn, 2013). Hence, the additional traffic variable called spatial spread of density, representing the “distribution” of congestion inside the cordon is introduced. Similarly to Knoop and Hoogendoorn (2013), the spread of density is estimated as the square root of the weighted variance of densities in all sections:

$$\gamma = \sqrt{\frac{\sum_l l_l (k_l - K)^2}{\sum_l l_l}} \quad (5.16)$$

Where  $l_l$  corresponds to the length of the link  $l$ ,  $k_l$  to the density of the link  $l$  and  $K$  the average density of the network.

#### 5.4 Consistency of MATSim with traffic flow theory

A deeper understanding of the possibility of MATSim to reproduce the main properties of traffic at macroscopic level is necessary to implement at the second stage tolling schemes based on a

sound theoretical background. Hence, before working on the case study scenario, a detailed investigation of the ability of MATSim to reproduce traffic conditions is carried out on a small test scenario (Figure 5.3). Particularly, the presence of the fundamental relationship between density and flow for single links is verified in order to extend the analyses at macroscopic level afterward.

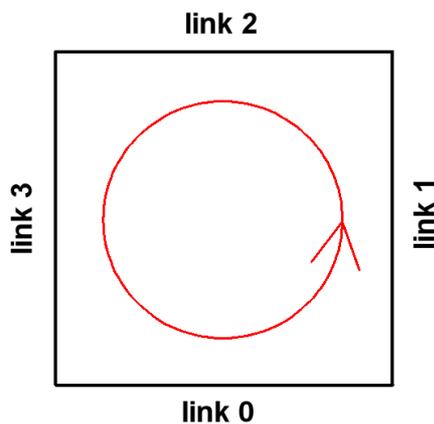


Figure 5.3: test scenario.

Such as scenario is composed of four links of equal length (200 metres) and capacity (2000 veh/h). Drivers are loaded on the networks from Link 0 and drive continuously over the ring. Every hour the number of agents is increased in order to create congestion. Several parameters of the simulation such as the time span between measurements, the rate of traffic increase and the speed of the gaps travelling backward are changed in order to gain a deeper insight of congestion phenomena within MATSim.

The most interesting aspect emerging from this experiment consists of the influence of the “gap speed” factor on the shape of the fundamental diagram. As it possible to see from Figure 5.4, 5.5, 5.6, representing the density-outflow relationship on a single link with different speeds of the gap, the patterns are considerably different. For high speed (e.g. 15 m/s) the diagram is characterized by a transition area where the production is constant and a steep decrease for very high density values (above 100 veh/km). On the contrary, when speed is low (e.g. 1 m/s) the decreasing branch of the fundamental diagram seems to be flattened almost completely. The most “reasonable” result seems to be given by speed of about 5 m/s. Under this condition it is possible to see a clear maximum around 25 veh/km and a decreasing branch until the jam density of 130 veh/km. This speed is in line with the typical backward speed of shockwaves and the resulting shape of the fundamental diagram corresponds to the theoretical one proposed by Daganzo (1997).

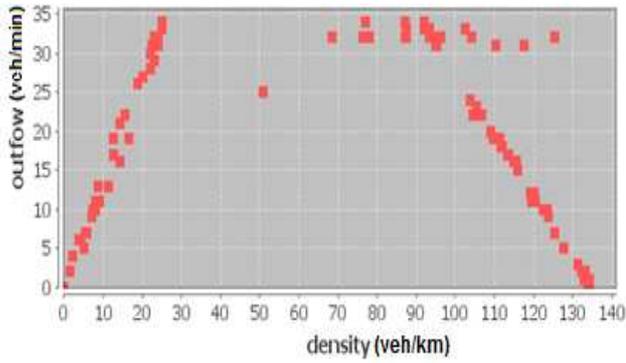


Figure 5.4: FD with gap speed of 15 m/s

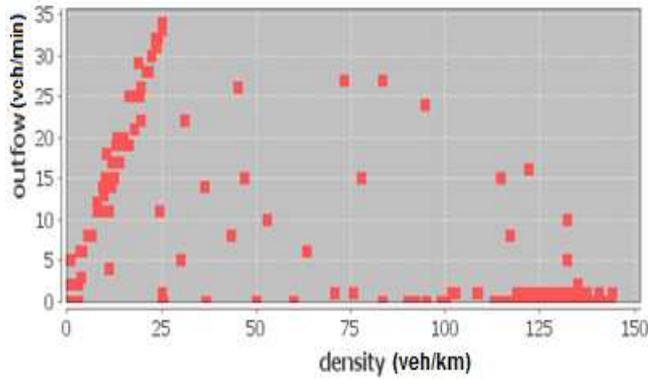


Figure 5.5: FD with gap speed of 1 m/s

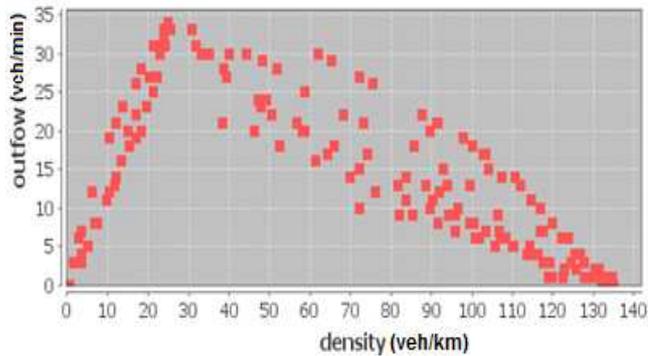
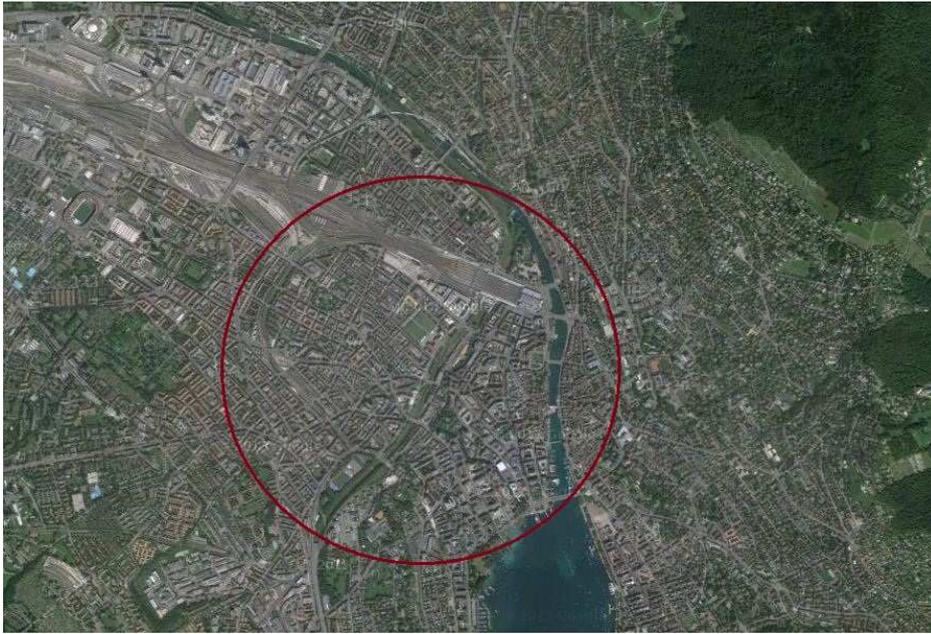


Figure 5.6: FD with gap speed of 5 m/s

## 5.5 Derivation of the MFD for the city centre of Zurich

Since all the findings in the previous section are consistent with former studies on macroscopic characteristics of traffic, we move now to the case study scenario. A ring of 1.5 km around the city centre is identified as cordon where agents will be charged regardless of their direction (Figure 5.7).



**Figure 5.7: Cordon area**

The table below (Table 5.1) provides a synthetic overview of the major quantitative characteristics of the analyzed data. From a qualitative point of view, the studied network is characterized by different typologies of road (arterial and collector roads, local streets) and by the presence of several bottlenecks like bridges and tunnels. Hence, such a typology of network will hardly satisfy the homogeneity conditions formulated by Daganzo and Geroliminis (2008) necessary to derive a well-defined MFD. However, as we will explain in the following sections, we will explicitly deal with this issue rather than looking for the optimal partitioning of the networks in regions that exhibit homogeneity of traffic conditions. The main rationale behind this choice is to provide a complementary approach to derive control strategies that cope with the (uneven) distribution of congestion.

<b>Key characteristics of the data analyzed</b>	
<b>total length</b>	175.5 km
<b>number of links</b>	1224
<b>number of intersections</b>	550
<b>simulation time</b>	00:00-24:00
<b>aggregation time</b>	5 minutes

**Table 5.1: Key characteristics of the data analyzed**

A careful observation of the simulation reveals the importance of certain simulation settings such as the “Flow Capacity Factor”, “Storage Capacity Factor” and “Squeeze Time” for the consistency of traffic simulation. The first one denotes the scale of link capacity and it clearly affects travel times. The second one denotes the scale of vehicle storage length of a link and it influences aspects such as the heaviness of congestion in terms of queue length and spillbacks.

The latter in particular might spoil the real throughput of links and ultimately affect the performance of the network if used not correctly (see Subsection 5.1.3).

For example, the possibility of congested links to “push forward” stuck vehicles might become a problem when two upstream links converge in a downstream link that is already full. This could result in an higher density than allowed one. The extent of this anomaly is considerably amplified when using smaller samples (e.g. 10%) instead of the full population. Indeed, in this situation link capacities are usually scaled down (typically to 0.1 with 10% samples) to match the demand and supply. Then, if a link is characterized by a capacity of few vehicles, every additional one represent a large increase of value of density. As a consequence, during the most congested periods of the day, few links are affected by substantial increases of density, sometimes higher than the theoretical value admitted.

Another important issue related to this factor consists in the resulting production-accumulation relationship. Although the possibility to “squeeze” vehicles makes the simulation more fluid and stable, it seems to affect the performance of the network at aggregated level especially in case of severe traffic congestion.

For low squeeze times, standstill vehicles in front of the queue are automatically pushed forward after few seconds. As a result, in congested conditions the outflow of links might correspond to the inflow and vehicles be able to move around the network with relative ease. Hence, as Figure 5.8 shows, no drop of performance of occurs and high values of production are maintained for relatively high densities.

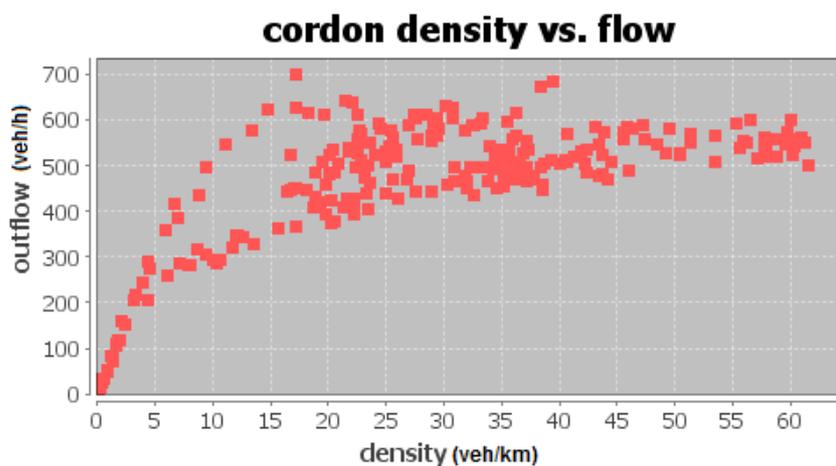


Figure 5.8: resulting MFD with low squeeze times.

On the contrary, for high squeeze times, the network does not seem to be able to deal with gridlocks and discharge all the traffic. As a result, the drop of performance occurs as it possible to see in Figure 5.9, but congestion does not resolve and the network is not able to recover.

Ultimately, several vehicles seem to be “squeezed” anyway.

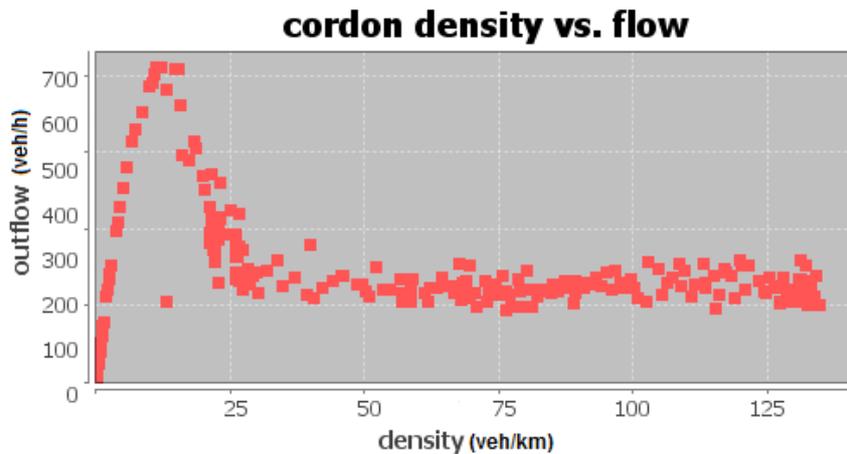


Figure 5.9: Resulting MFD with high squeeze times

A “sweet spot” that allows the derivation of a consistent MFD and enables the network to discharge the traffic by the end of the simulation is identified through a trial-and-error process where the parameters: “FlowCapacityFactor” (FCF), “StorageCapacityFactor” (SCF) and “Squeeze time” are tuned. As previously mentioned, these factors clearly influence the output of the simulation and quality of the model. The FCF directly affects the travel times of users and determines the onset of congestion phenomena. The SCF and Squeeze Time determine the intensity and propagation of congestion. In theory, when using a 10% sample the FCF and SCF should be set to 0.1, but in practice they are typically increased to 0.2-0.3.

In this study, the optimal settings consist of slightly increased FCF and SCF (0.13) and squeeze time equal to 150 seconds for a sample size of 10%. The resulting production-accumulation relationship is represented in the figure below (Figure 5.10). From a traffic flow theory perspective, the results seem to be in line with the main principles (well recognizable traffic regimes) and with the previous studies (comparable values of accumulation-production). Furthermore, the problem of those extremely high densities generated by squeezed vehicles in already congested links has almost disappeared (the phenomenon involves less than 0.5% of the links during the peak hours). With regard to the realism of the model, the chosen values do not compromise the quality of the simulation as some main indicators show (counts, modal split, average travel time).

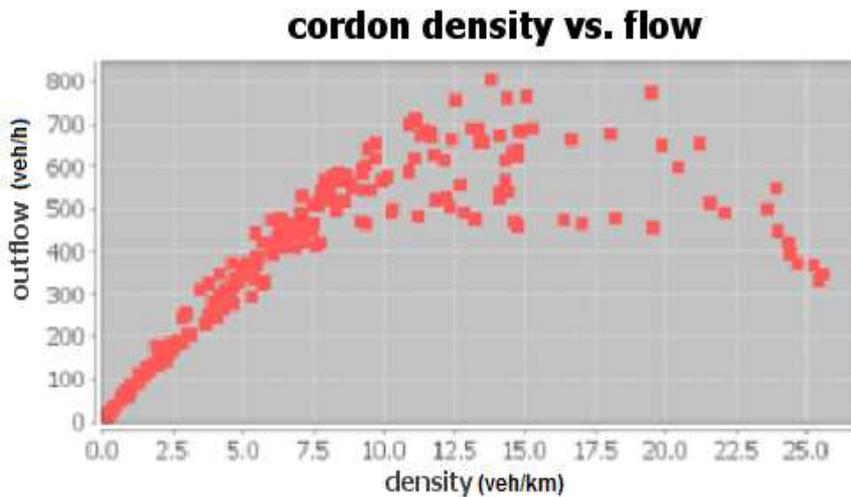


Figure 5.10: MFD derived at the “sweet spot”

An additional confirmation is given by the fair consistency with a simulation of the same scenario with a larger sample size (20%). In this case, the capacity factors are scaled up to 0.25 and the squeeze time is slightly “relaxed” to 100 seconds. The resulting accumulation-production relationship (Figure 5.11) is similar from a quantitative (corresponding maximum outflow, critical density and free flow speed) and qualitative (presence of scatter) point of view.

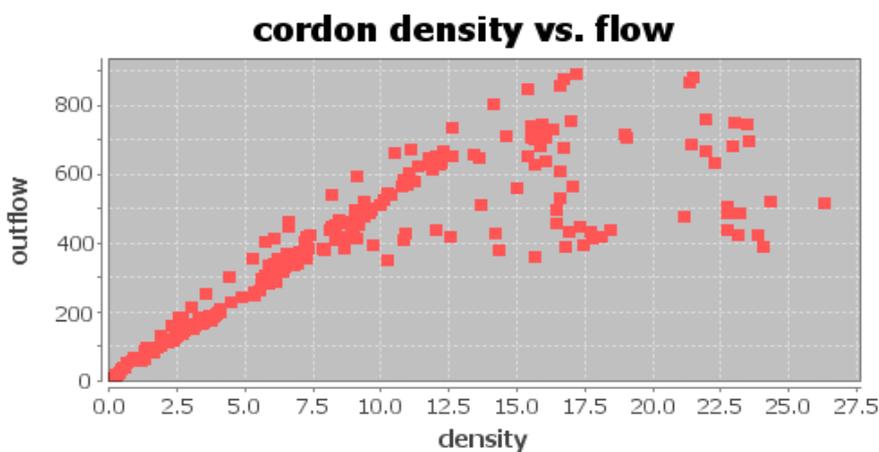


Figure 5.11: MFD for a sample size of 20%

The main issue encountered during the simulation consists of “network breakdowns” determined by changes in the agents’ plans that lead to overloads of congestion. In this situation congestion cannot be absorbed anymore by the network (since squeeze times have been increased) and it generates severe gridlocks. A similar phenomenon is described in the article “MatSim-T: Architecture and Simulation Times” by Balmer *et al.* (2009).

In order to make the simulation more stable the following strategy has been adopted. First, the “squeeze time” has been gradually increased (from 100sec in the first fifty iterations to 150sec

in the last fifty ones) in order to allow the system to reach equilibrium more smoothly. Furthermore, “innovative strategies” have been deactivated in the last 10th iterations (from it.90 to it.100) in order to reduce fluctuations. In this way, agents are able to change only the plans stored in their memory so far (the learning mechanism is interrupted). Although the combination of the two adopted solutions allows higher stability of system, the equilibrium conditions are not reached very “naturally”. As result few agents might have not optimal plans (not the best possible) and some deviations in the final utility (score) of agents might occur due this unnatural procedure. Nevertheless, given the high amount of iterations, it is plausible to think that such a procedure does not compromise the overall quality of the simulation. Alternatively in the future more elegant approaches targeted to the re-routing “on the fly” and the propagation of spillbacks may be adopted.

## **5.6 Investigations of the macroscopic traffic relations**

The first two figures represent respectively the accumulation-production and the accumulation-average speed relationship estimated by means of the previously discussed formulas (Section 5.3). As concern the capacity, the value derived (around 700-800 veh/h/lane) is in line with the outputs of previous similar studies: Daganzo *et al.* (2011); Mazloumian *et al.* (2010); Mahamassani *et al.* (2013). The identified critical accumulation (around 15 veh/km/lane) is, instead, 30-40% lower than the other studies. However, this does not constitute a problem since every network is characterized by a different “shape” with its own specific threshold values, depending on several factors such as the typology of roads, the mix of traffic and the traffic control (although in our simulation no traffic signal is implemented).

It possible to see from Figure 5.12 that, although the diagram exhibits the typical concave shape, the situation is not as clear as for the experimental relationship determined in homogeneous conditions (see previous Chapter 4). In this case, when density approaches values closer to capacity, a “bifurcation” takes place and both uncongested and congested state can coexist for the same value of accumulation.

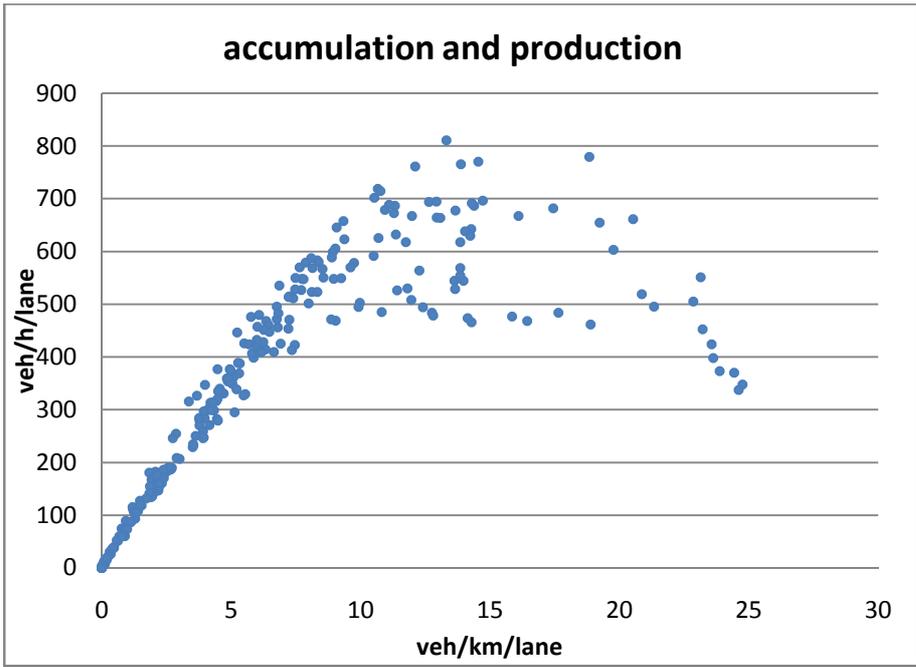


Figure 5.12: accumulation-production diagram

Figure 5.13 representing the macroscopic relationship between average accumulation and average speed of the network, despite the presence of some outliers, identifies an average free-flow speed of about 70 km/hour. The presence of some arterial roads, characterized by high speed is probably the main reason for such an high value.

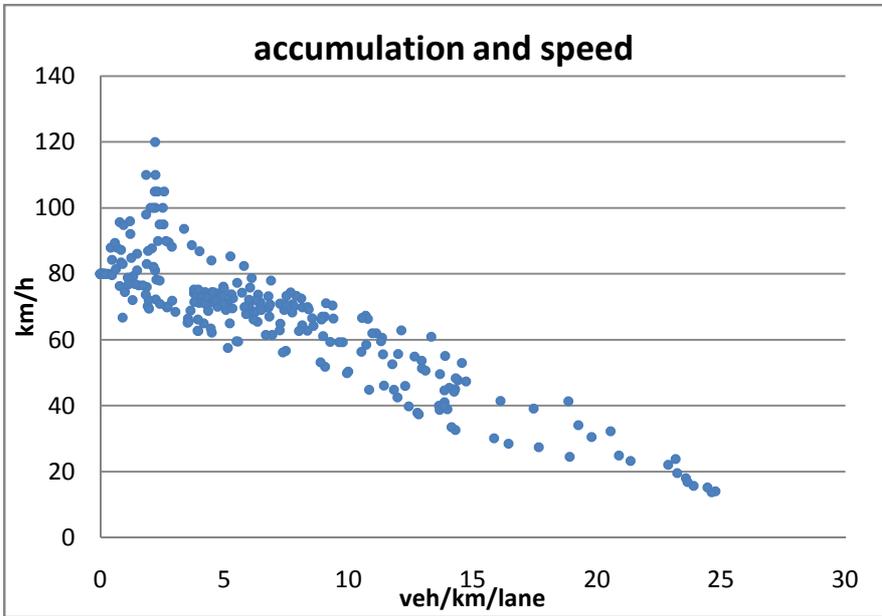


Figure 5.13: accumulation-speed diagram

The relationship between average density and spread of density (Figure 5.14) highlights a parabolic trend and it shows that different values of spread might correspond to the same value of accumulation. This fact can be physically interpreted that the same amount of users in the

system (density) could be evenly distributed on the network or more concentrated in specific zones (higher spread). Interestingly, only an upward trend is observed, while after a certain level of density a decrease of spread would be expected as the network becomes congested everywhere. However, since in this simulation during the most congested period only 25% of the links result congested (the network is far away from being “uniformly” congested) the resulting pattern of spatial spread seems plausible.

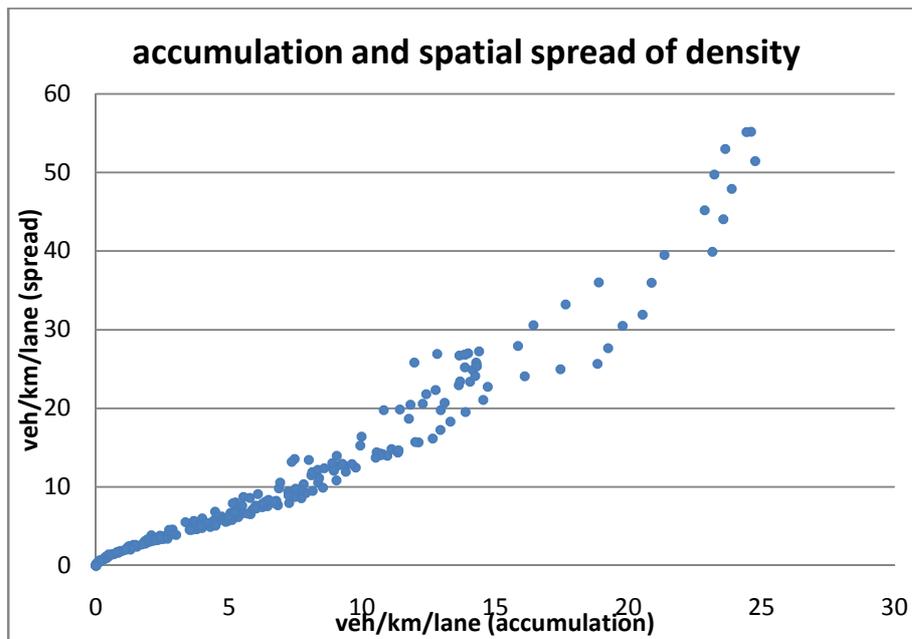


Figure 5.14: accumulation-spatial spread of density diagram

A more careful observation of the hysteresis phenomenon shows that the inhomogeneous distribution of congestion can play a role in the performance conditions even at densities below the critical one. This is in line with Daganzo *et al.* (2011) and Mahamassani *et al.* (2013) who discussed the eventuality of this phenomenon (bifurcation of the MFD) and identified as possible reason behind that uneven distribution of traffic (ultimately related to the degree of turning rate). The figures below represent the average density-outflow and spread-outflow relationships during the morning (Figure 5.15, Figure 5.16) and evening peaks (Figure 5.17, Figure 5.18). The results seem to be in line with the study by Geroliminis and Sun (2011) who identified a clockwise pattern in the density-outflow diagram corresponding to an anticlockwise pattern in the spread-outflow diagram (red circles in the figures). It is worth mentioning also that during certain time intervals drops of production seem to be determined by increases of spread rather than increases of density (“vertical” slopes in the black circles). Physically this would correspond to decreases of production generated by clusters of congestion rather than increases of users in the network.

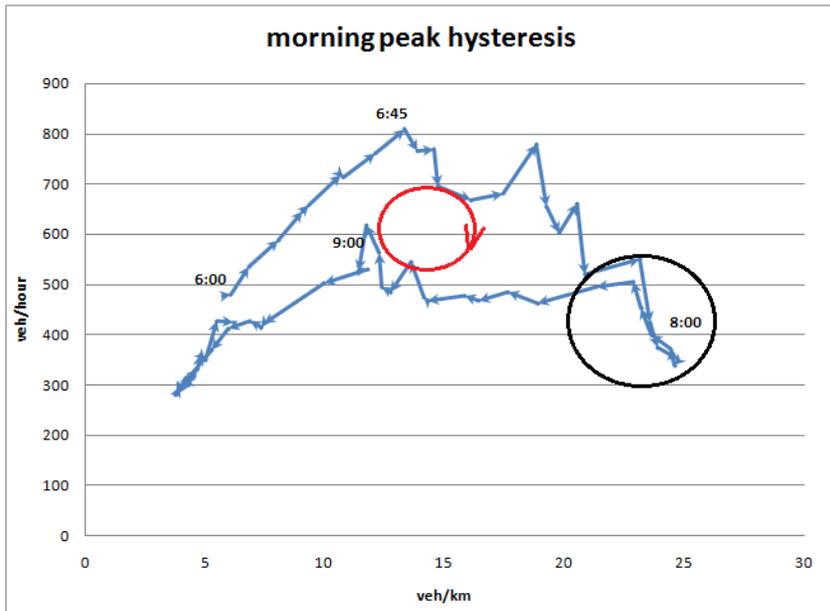


Figure 5.15: Morning peak accumulation-production hysteresis characterized by a clockwise loop.

The figure above represents the morning peak hysteresis in the accumulation-production diagram.

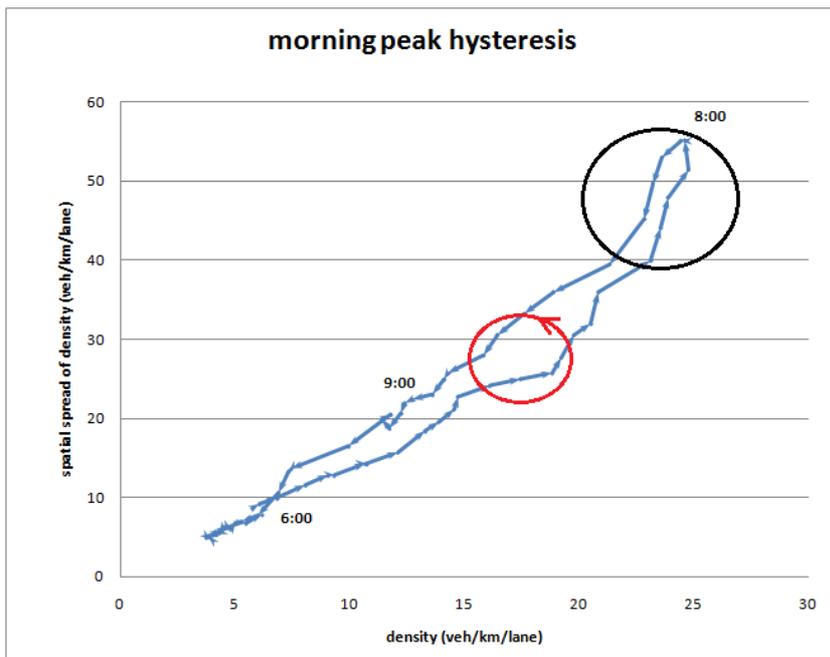


Figure 5.16: Morning peak spatial spread of density-production hysteresis characterized by an anticlockwise loop.

The figure above represents the morning peak hysteresis in the spread-production diagram.

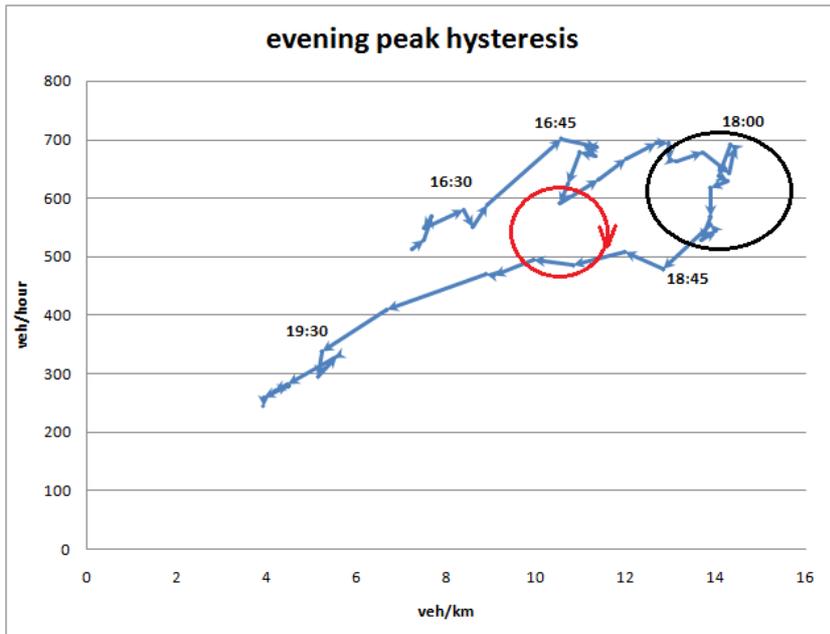


Figure 5.17: Evening peak accumulation-production hysteresis characterized by a clockwise loop.

The figure above represents the evening peak hysteresis in the accumulation-production diagram.

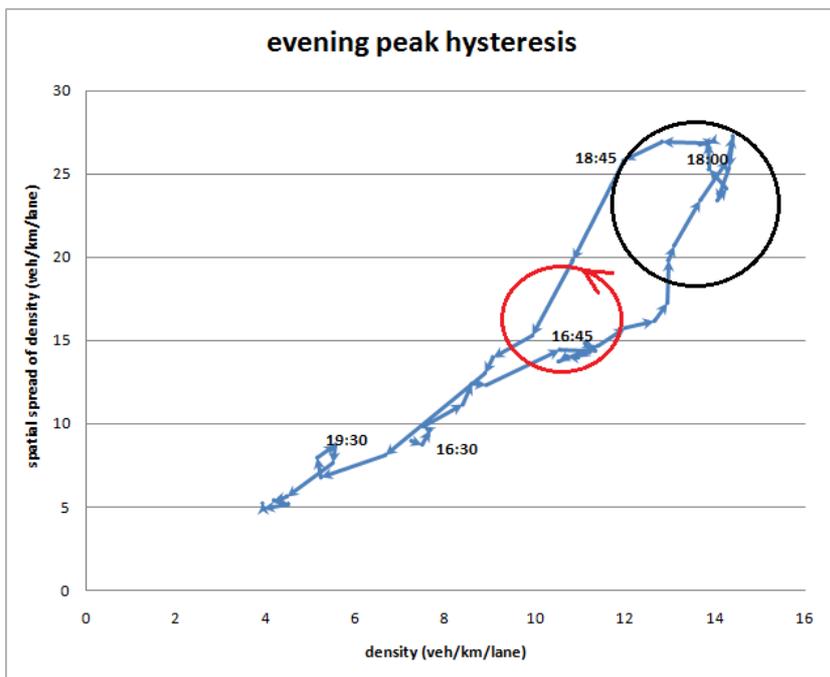


Figure 5.18: Evening peak spatial spread-production hysteresis characterized by an anticlockwise loop.

The figure above represents the evening peak hysteresis in the spread-production diagram.

## 5.7 Identification of a Generalized Macroscopic Fundamental Diagram

The presence of scatter in the macroscopic traffic relationship seems to be related to inhomogeneous traffic conditions inside the cordon. Qualitative and quantitative observations confirm the coexistence of very congested links and free ones inside the cordon during the same time intervals. For example, the two figures below (Figure 5.19 and Figure 5.20), representing the density of links in the cordon at 7:30 and 18.30 show that during the morning peak the links are more homogeneously congested, while during the evening peak, few spikes of congestion are present.

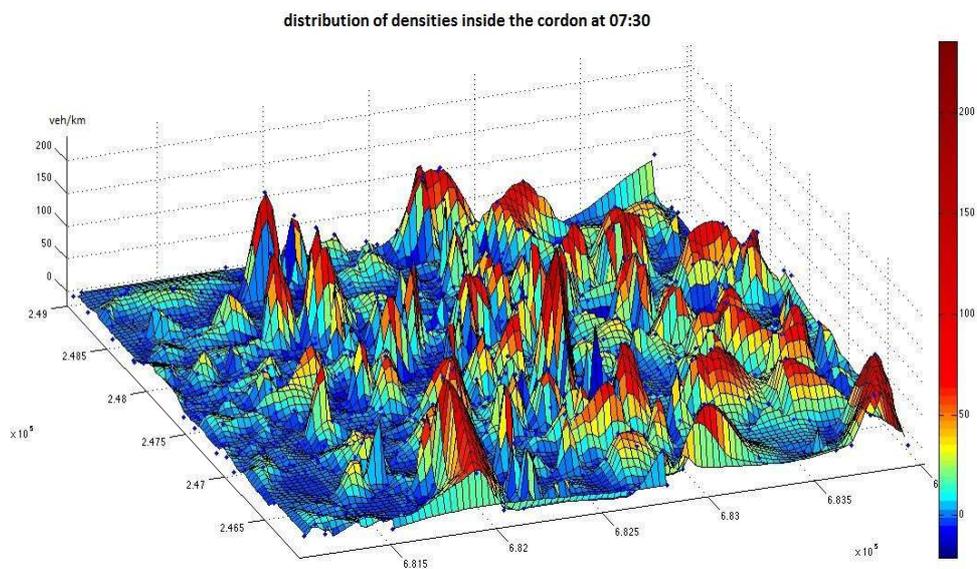


Figure 5.19: Distribution of densities inside the cordon at 07:30

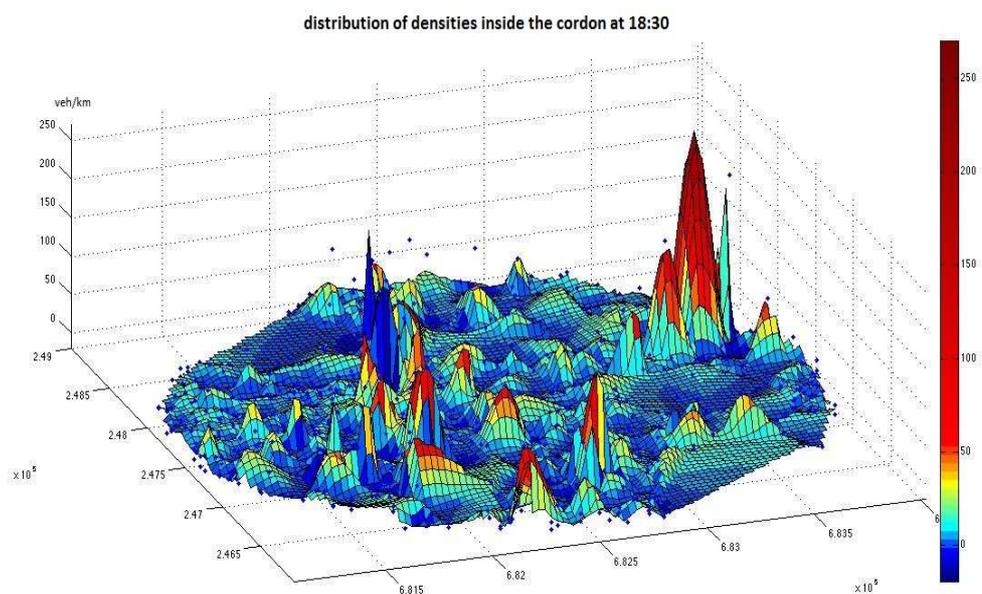


Figure 5.20: Distribution of densities inside the cordon at 18:30

The frequency distribution of single link densities reported in Appendix I seems to confirm the idea that, for similar amount of users (accumulation), traffic is distributed more evenly during the morning. Indeed, during the evening peak there are several links close to the jam density whereas during the morning there are more links in the congested regime (but not so severely congested). Consequently, the spread of density is relatively higher during the evening and the traffic accumulation-production follow different patterns.

These findings are consistent with the conditions expressed by Daganzo and Geroliminis (2007; 2008) and Buisson and Ladier (2009) according to whom homogeneity of traffic inside the network is fundamental to derive a well-defined MFD, which is actually a theoretical upper bound.

In this thesis, rather than identifying “reservoirs” characterized by homogeneous traffic conditions, the presence of spread has been explicitly considered and an empirical relationship with the “typical” traffic variables (average density and outflow) has been identified in a similar way to the study by Knoop and Hoogendoorn (2013).

A Generalized Macroscopic Fundamental Diagram (GMFD) that includes also the measure of spread of density is represented below (Figure 5.21). The accumulation and production are plotted respectively on x and y axis, while the spread of density is represented by the color.

The diagram has been derived by using the output of 5 simulation “seeds” characterized by very similar agents’ plans, but slightly different origin-destination patterns. The term “seed” here refers to an alternative output obtained from different simulations with slightly different amount to replanning. The number of seeds has been set arbitrarily in order to simply verify the macroscopic behavior of the network in slightly different situations. Under homogenous conditions, a well-defined invariant MFD would show up under any circumstance, but in this case study the diagram exhibits high scatter. However, such a phenomenon should not be considered as a pure coincidence. In fact, as it possible see from the colored plot (Figure 5.21), higher accumulation values correspond to higher spread values and more importantly, for the same value of accumulation the lower production the higher the spread.

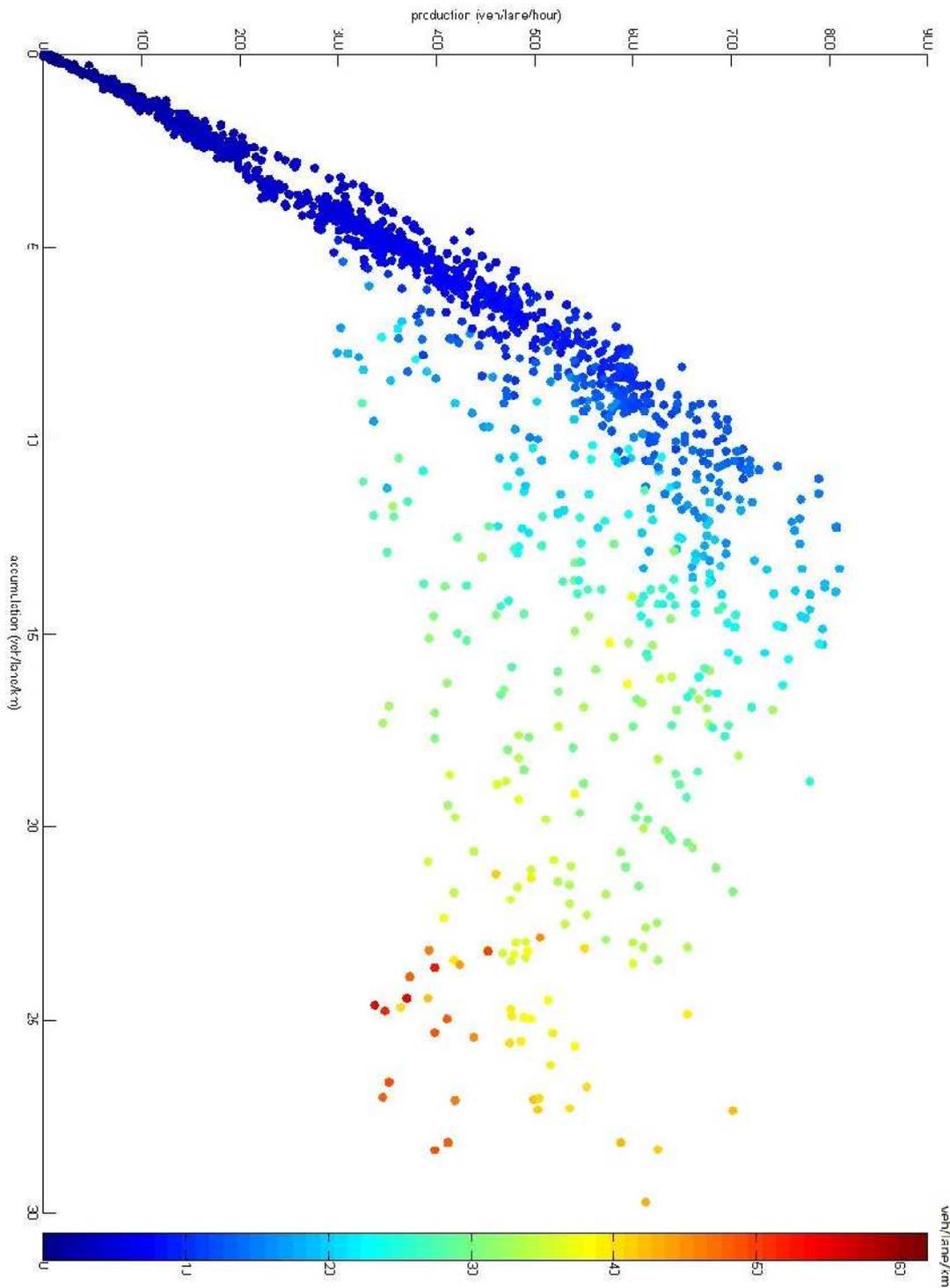


Figure 5.21: Generalized Macroscopic Fundamental Diagram with spread of density in colors

### 5.8 Development of three alternative tolling schemes based on the (g)MFD

Once the macroscopic characteristics of the cordon have been identified, it has been finally possible to derive alternative pricing schemes. In this study, three main schemes have been implemented:

1. Flat Toll
2. Step Toll
3. Step toll accounting for spread (Spread Toll)

Although the three alternative tolls are all based on the optimal control of the aggregated traffic relationships of the network, they are characterized by different conceptual approaches. The Flat Toll aims at operating the system below a threshold value corresponding to the critical accumulation. The Step Toll consists of a time-varying toll that aims at eliminating delays in the cordon by means of the MFD. The Spread Toll is a variation of the Step Toll where the inhomogenous distribution of traffic and its influence on performance are considered as well in the estimation of charges by means of the (g)MFD.

The cordon-based tolls involve both the inbound and outbound trips. Although penalizing trips exiting the area might seem counterintuitive, this decision is supported by the following arguments. First, drivers determine delays inside the cordon also when driving outbound and they need to pay for that. Second, since a large portion of the trips during the evening peak-hours will be likely directed from the centre to the suburbs, it would be useful to regulate their flows by applying an outbound toll. Third, a bi-directional toll is meant to be an additional disincentive for those drivers that just cross the cordon, as they would pay twice.

#### **5.8.1 The Flat Toll**

The first scheme corresponds to a fixed charge derived by means of a feedback control process like in the study by Zheng *et al.* (2012) where the toll is updated until the average density is below a threshold value (critical accumulation). The main difference in this study consists of using of two different threshold values for the morning and evening peaks as the drop occurs at lower density during the evening (also because of the inhomogeneous conditions).

#### **5.8.2 The Step Toll**

The second scheme corresponds to a time-varying toll in discrete intervals of half an hour called Step Toll. In this case, the price is derived such that the new users are charged for the additional delay they create inside the cordon. This principle is in line with the MCP approach where the external costs generated by each traveler are internalized by means of the toll. By means of the MFD it is possible to measure the change of average speed determined by the corresponding changes of accumulation and production. The approach is illustrated in the figure below (Figure 5.22) and expressed by means of the following equations.

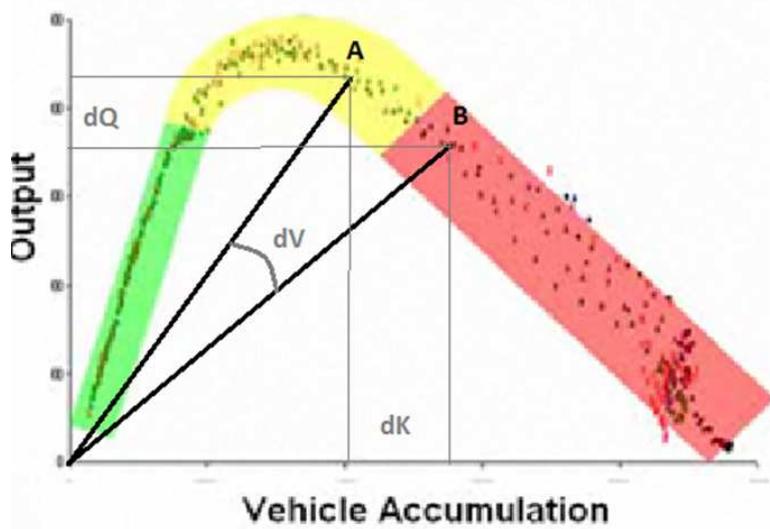


Figure 5.22: Macroscopic derivation of delays inside the cordon. Adapted from Geroliminis and Daganzo (2008)

The time loss per user determined by a decrease of speed corresponds to:

$$\Delta d = t_1 - t_0 = \frac{s}{u_1} - \frac{s}{u_0} \quad (5.16)$$

Where  $s$  corresponds to the average trip distance travelled inside the cordon.

The total delay for users on the network inside the cordon (Figure 5.23) is given by:

$$\Delta D = \Delta d \cdot K \cdot L \quad (5.17)$$

Where  $L$  corresponds to the total length of the network.

The number of additional users  $N$  is derived from the change of average density

$$N = \Delta K \cdot L \quad (5.18)$$

Finally, the toll is derived by dividing the product of total delay with an average value of time (25 CHF) by the amount of additional users.

$$\tau = \frac{VOT \cdot \Delta D}{N} \quad (5.19)$$

The final toll has been determined by means of an iterative process where at the end of each simulation, the delays and the time-varying toll were derived and updated until significant drops of production were eliminated. A minimum time interval of half an hour has been applied as a time constraint to derive the steps. It is worth mentioning that the Step Toll could have also been determined “more dynamically” entirely within the simulation process, with levels of charge determined according to real time traffic conditions, similarly to traffic management measures like perimeter control or ramp metering. Probably, this typology of scheme would

enhance higher benefits as it would be able to deal better with the variability of traffic conditions. However, because of the high load of programming involved and its low applicability (it might be too hard for road users to accept a charge that varies everyday and every few minutes according to different traffic conditions), the former static approach has been preferred. For certain aspects this strategy reminds the one used in Singapore where tolls are set on intervals of half an hours (during the peak hours) and monthly updated in order to achieve the desired levels of average speed. Nevertheless, it might be theoretically interesting to test the effects of a more dynamic toll based on the (g)MFD in the future studies.

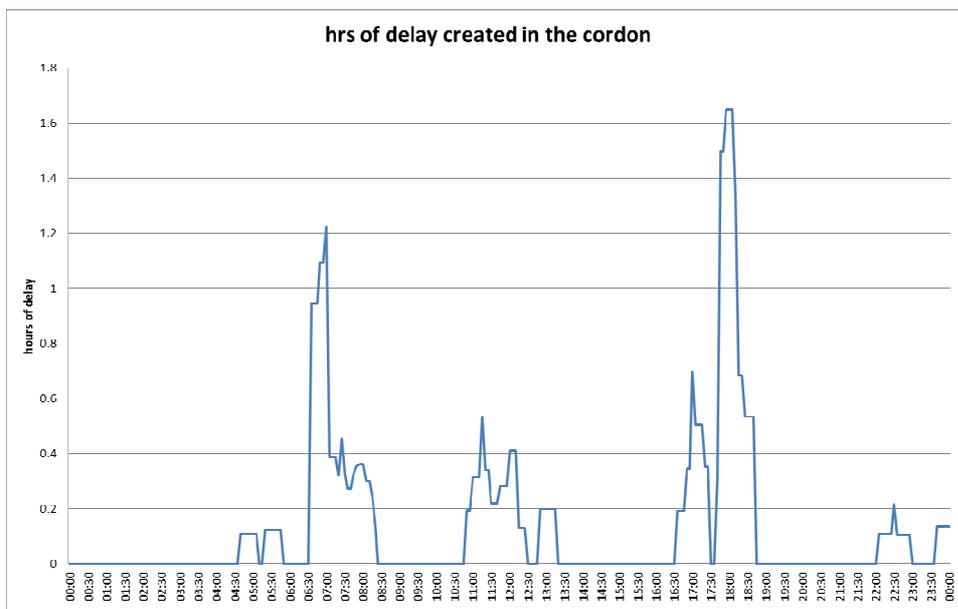


Figure 5.23: Total delay of users inside the cordon

The figure below (Figure 5.24) represents the initial estimation of the Step Toll during the evening peak.

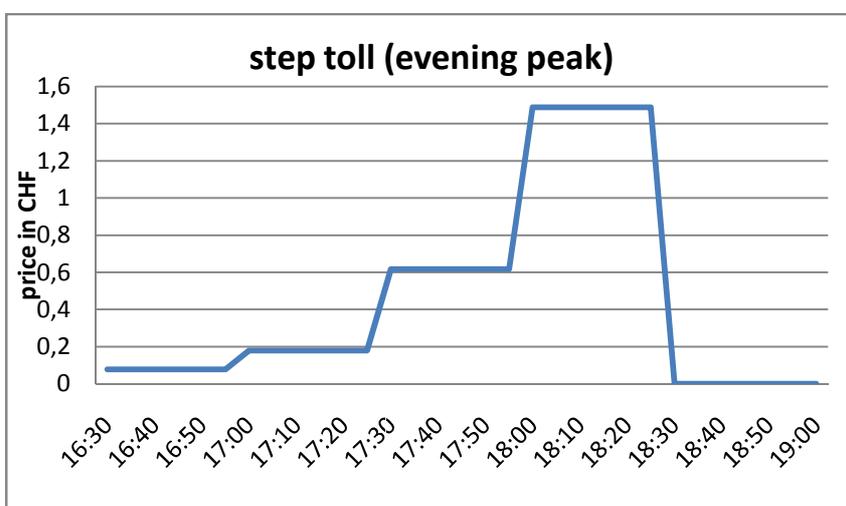


Figure 5.24: Preliminary estimation of the Step Toll (evening peak)

### 5.8.3 The Spread Toll

The third scheme consists of a step toll that accounts for the issue of uneven distribution of traffic by means of the spread of density. In practice, as it has been already shown, the accumulation-production relation is characterized by scatter and it is influenced by the distribution of congested links. For example, as it possible to see in the bottom-left quadrant of the figure below (Figure 5.25), higher drops of performance might occur even for lower increases of density (AC compared to AB) because of clustering of congestion. Typically these decreases are characterized by high increase of spread of density (red colored dots). This hypothesis finds support in the previous studies (see Section 4.5) that highlighted how the distribution of congestion affects the production of the system.

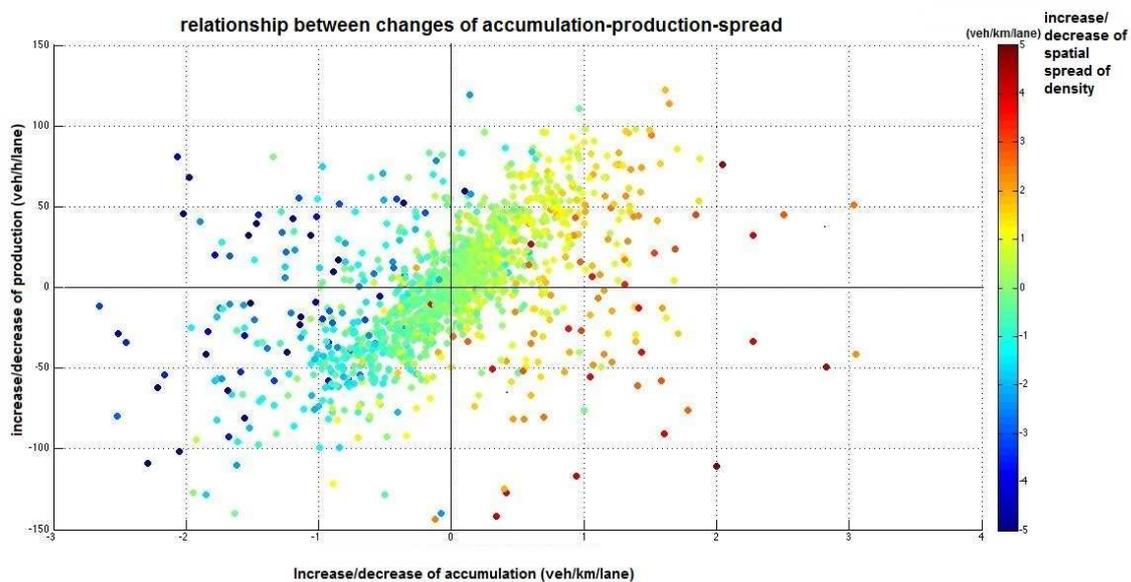


Figure 5.25: Relationship between changes of accumulation, production and spatial spread of density

The second tolling scheme applies a toll regardless of the loss of performance due to the heterogeneity of traffic conditions inside the cordon. As a result, it cannot properly internalize the cost of delay related to new entrants. Indeed, few entrants might pay high tolls only because clusters of congestions (of users already entered) have occurred during the same time interval. The third scheme explicitly considers this issue and it applies a toll that internalizes only the delay determined by the increase of density.

The resulting fares and improvements will be lower than the other two tolling schemes, as the uneven distribution of congestion proved to be a major cause of the decrease of performance of the system. On the other hand, from a social perspective this scheme represents a “fairer” approach as it “punishes” users only for the actual drop of performance they determine by entering the cordon.

In order to derive this kind of toll, it is first necessary to identify the extent of drop determined by the increase of density and increase of spread. Hence, a relationship between average density, spread of density and average outflow is expressed by means of the “best fitting” function.

Following the example of Knoop and Hoogendoorn (2013), the following polynomial form is adopted for reason of simplicity:

$$Q(k, \gamma) = a \cdot k + b \cdot k^2 + c \cdot k^3 + d \cdot \gamma \quad (5.20)$$

Where  $Q$  corresponds to the total production,  $k$  corresponds to the accumulation and  $\gamma$  corresponds to the spread of density. The figure below (Figure 5.26) represents a tridimensional representation of the GMFD and its fitting function.

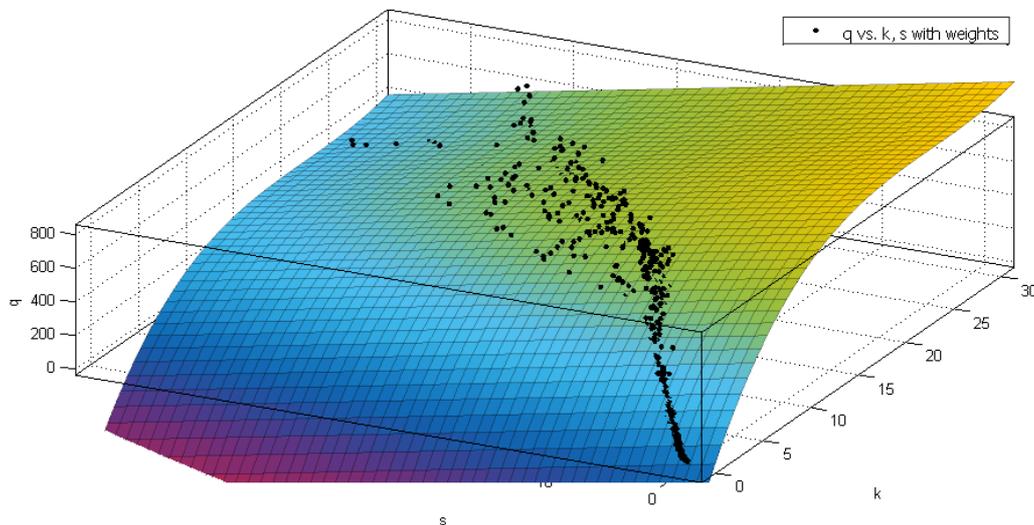


Figure 5.26: tridimensional representation of the polynomial function and the dataset.

The focus of this estimation process has been mainly on the quality of approximation of the drops of production determined by the combination of increases of density and spread rather than providing a function that explains the behavior for every traffic state. The coefficients  $a$ ,  $b$ ,  $c$ ,  $d$  are estimated by means of a weighted polynomial regression so that the deviation (RMSE) between the measurement and the fit is minimum. A sample of 1440 measurements and a control group equal to 20% of the sample to test the goodness of fit have been used. The resulting values are reported below.

$$a = 127.6 ; b = -5.61 ; c = 0.082 ; d = -11.75$$

With RMSE=13.25 and coefficient of determination  $\bar{\rho}^2=0.69$ . Although the  $\bar{\rho}^2$  is not particularly high, the function gives a reasonable approximation (around 8% of average error) during for

congested and nearly congested traffic conditions. As it possible to see from Figure 5.27 showing the goodness of fit of the derived function, although free-flow regime looks slightly overestimated, the congested regime seems to be fairly reproduced. Note that the assumed relationship is only an approximation that seems to provide a good estimation of decreases of performance due to variations of accumulation and spread in this specific study, but it may not represent the actual form of the Generalized Macroscopic Fundamental Diagram. Further research will be needed to identify a sound form to express the relationship between the three variables.

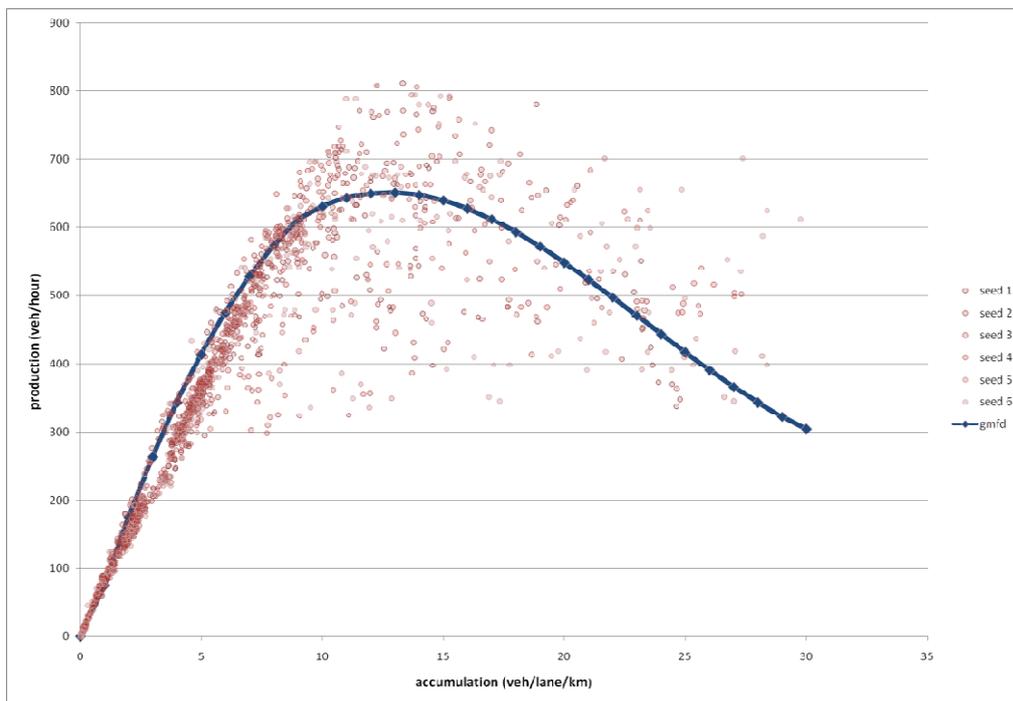


Figure 5.27: Fitting of the derived gMFD function

Finally, once the production is expressed as a polynomial function of density and spread, it is possible, by calculating its gradient  $\nabla Q(k, \gamma) = \frac{dQ}{dk} + \frac{dQ}{d\gamma}$ , to identify the variation of outflow “strictly” due by the variation of density and spread. Hence, a toll aimed to internalize solely the decrease of performance determined by additional users can be identified.

The initial toll will updated with the same mechanism of the Step Toll until all the delays determined by additional users will be eliminated.

#### 5.8.4 Final fares

After several iterations it was possible to derive the corresponding fares for the three alternative tolling schemes (Figure 5.28). A time constraint of thirty minutes has been applied to determine the discrete intervals of change (steps) for the time-varying tolls. Such a constraint

has been chosen for reason of understandability and with reference to the currently operating systems in Singapore and Stockholm.

The final Flat Toll corresponds to an amount of 1.5 CHF from 06:30 to 08:30 and 2 CHF from 16:30-19:00. The final Step Toll reaches a maximum amount of 1.3 CHF between 07:30-08:00 and 2.5 CHF between 18:00-18:30. The final step toll accounting for spread (Spread Toll) entails significantly lower fares (less than 1 CHF) and it applies during shorter intervals (07:00-08:00 and 18:00-19:30). The resulting fares are in line with the expectations, especially the Spread Toll that is by definition a “milder” approach when drops of performance are determined by cluster of congestions. It is worth mentioning that the levels of fares in the evening peak are higher than in the morning peak, even though the overall levels of traffic are lower. This outcome confirms again the different nature of congestion of the two peaks.

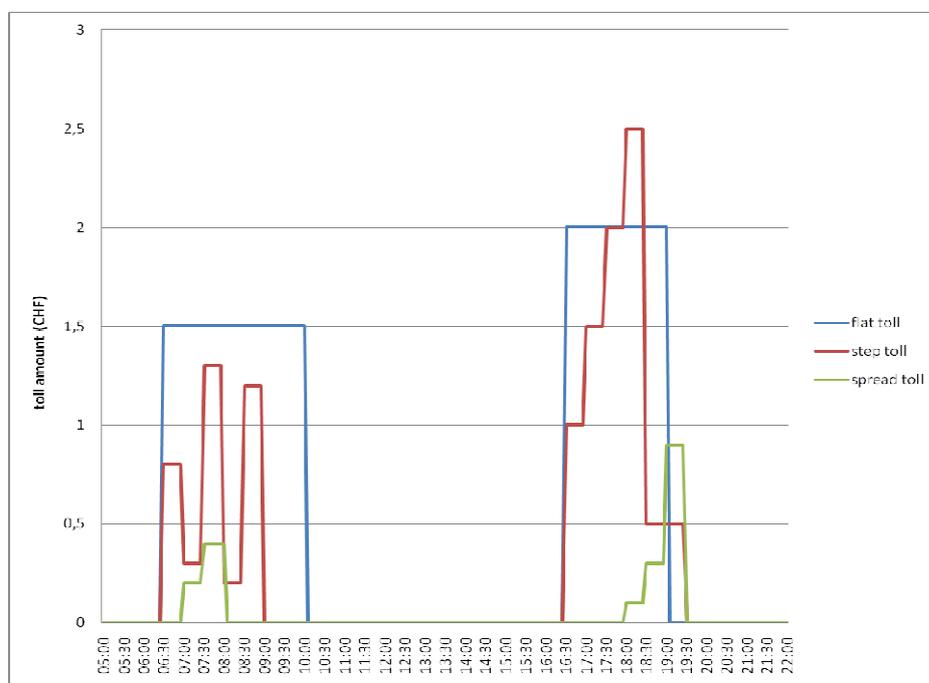


Figure 5.28: Final fares of the three tolling schemes

## 6. Comprehensive evaluation of tolling schemes

Assessing congestion pricing schemes requires not only to consider their efficiency, but also their distributional impacts and their achievable level of public acceptability. For these reasons, it is fundamental to evaluate tolling schemes with a comprehensive approach that integrates different aspects like traffic performance improvements, travel behavior responses and economic impacts. The following conceptual framework, reported in Figure 6.1, will be adopted to evaluate and compare the alternative schemes from a broader perspective. The investigations will be driven by a series of hypotheses in order to allow in the end a conclusive discussion about the overall performance of the tolling schemes. The final considerations concerning the equity and public acceptance of the alternative tolls will be focused on some specific socio-demographic categories identified on the basis of the previous review of these issues (Chapter 2).

The analyses reported in the following sections will try to give light to the following questions:

- Do the alternative tolling schemes determine similar improvements from a traffic flow theory perspective?
- How people react to the different schemes in terms of travel behavior changes?
- Are certain travel behavior choices influenced by any socio-demographic characteristic?
- Is it possible to identify any relationship between travel behavior responses and economic impacts?
- Do the tolls produce significantly different economic impacts on the whole population and across specific socio-demographic groups?
- Is it possible to identify any relationship between the economic impacts of the tolls and any of identified socio-demographic characteristics?

Then, a thorough evaluation of the effects of the schemes by means of the typical traffic flow performance indicators will be carried out in Section 6.1 and 6.2. After that, the investigations will focus on the economic aspects of congestion pricing. First, the economic interpretation of the results will be illustrated and an explanation of the choice of socio-demographic groups will be provided in Section 6.3 and 6.4. The impacts of the schemes on the travel behavior will be discussed in Section 6.5 to give an insight of the various ways how different groups respond towards the tolls. The economic efficiency and the distributional effects will be examined in Section 6.6. Finally, on the basis of the previous investigations, the profile of winners and losers will be described in Section 6.7. In the end, some considerations about the public acceptability of schemes will be made with the support of additional analyses in Section 6.8.

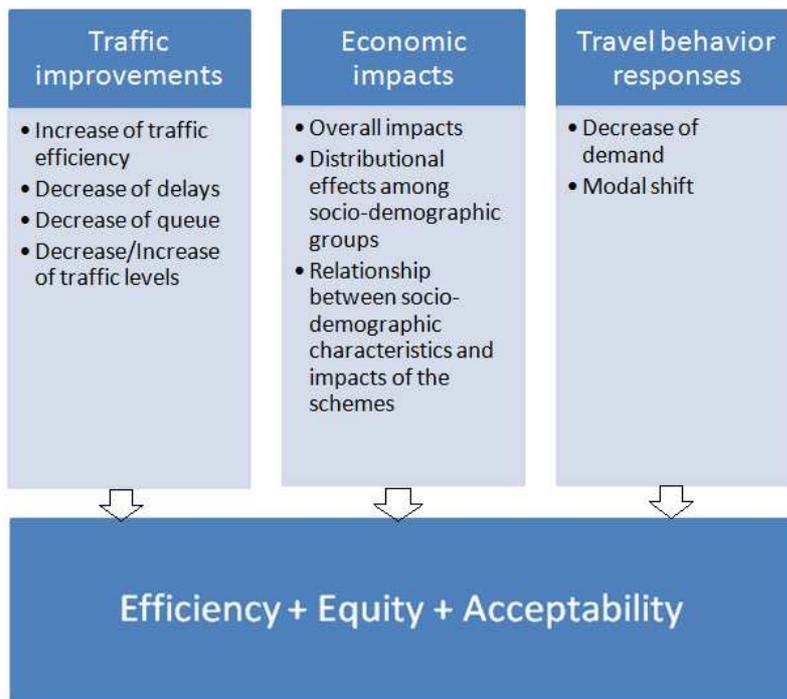


Figure 6.1: Overview of the comprehensive appraisal framework

### 6.1 Resulting production-accumulation relationships

The Flat Toll schemes produces a significant improvement of performance. The congested branch disappears and it exhibits no drop of production (Figure 6.2).

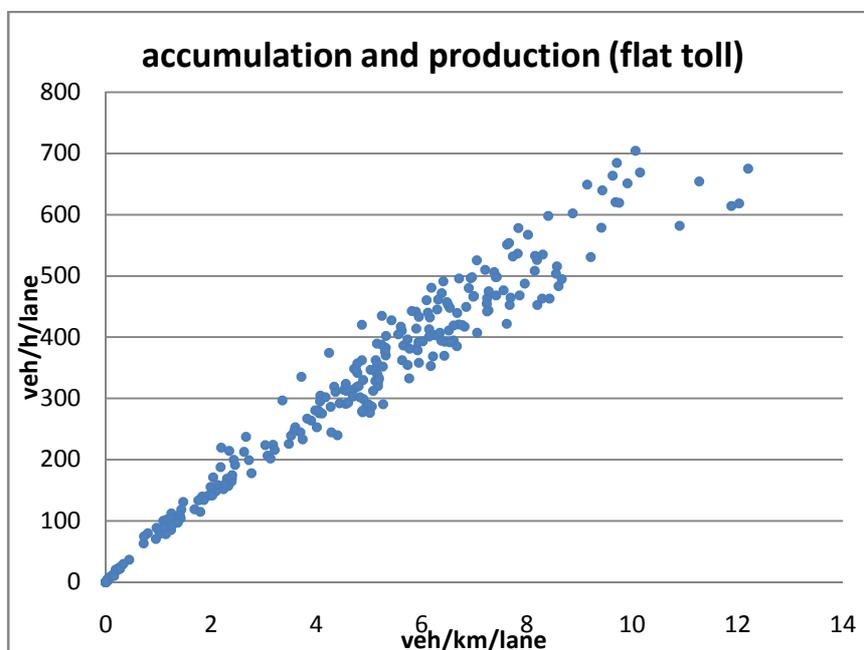


Figure 6.2: Resulting accumulation-production relationship from the Flat Toll

Also the Step Toll seems to enhance higher performances of the network as it does not exhibit any congested branch (Figure 6.3). However, on closer inspection it presents differences from

the Flat Toll scheme. Indeed, the free-flow branch presents higher scatter that can be explained by hysteresis phenomena characterized by frequent loading and unloading cycles (Figure 6.4). At every cycle the performance of the system seems to progressively deteriorate. This result suggests that hysteresis phenomena might be a reason of decrease of the performance of the system as well.

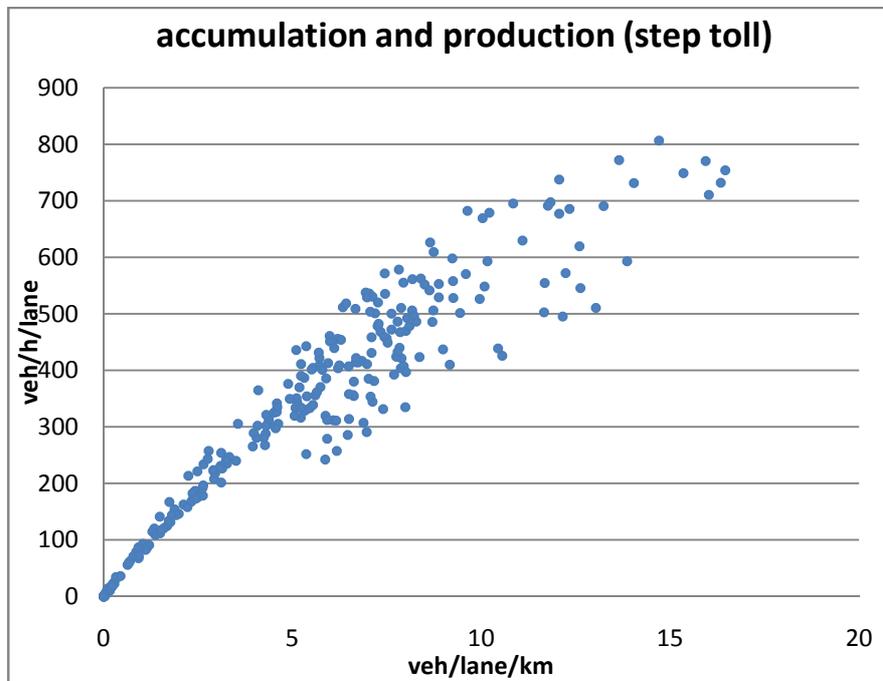


Figure 6.3: Resulting accumulation-production relationship from the Step Toll

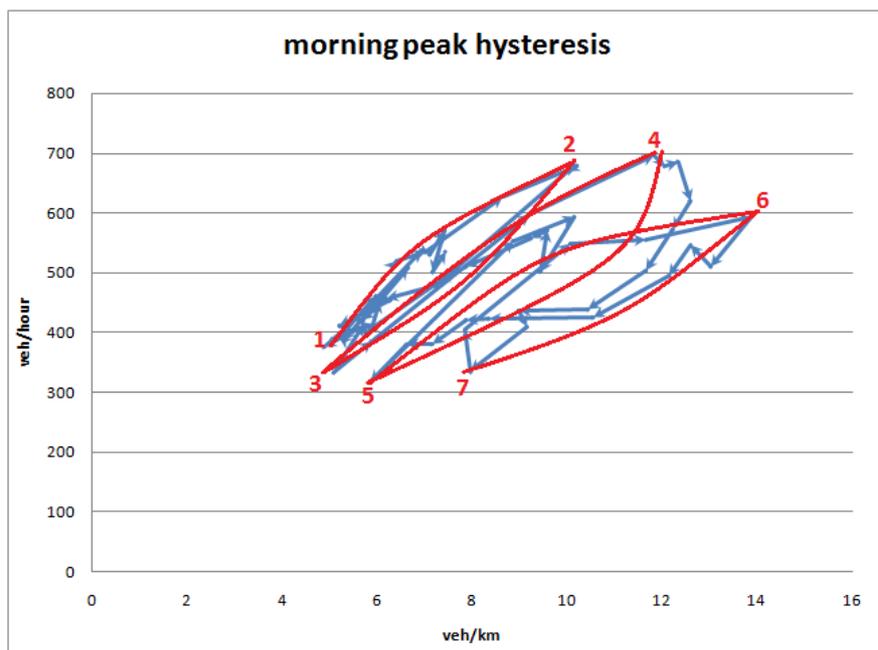


Figure 6.4: Overview of hysteresis patterns associated with the Step Toll

The accumulation-production diagram resulting from the Spread Toll still exhibits considerable decreases of performance (Figure 6.5). All in all the traffic improvements seem to be less than the other two tolling schemes.

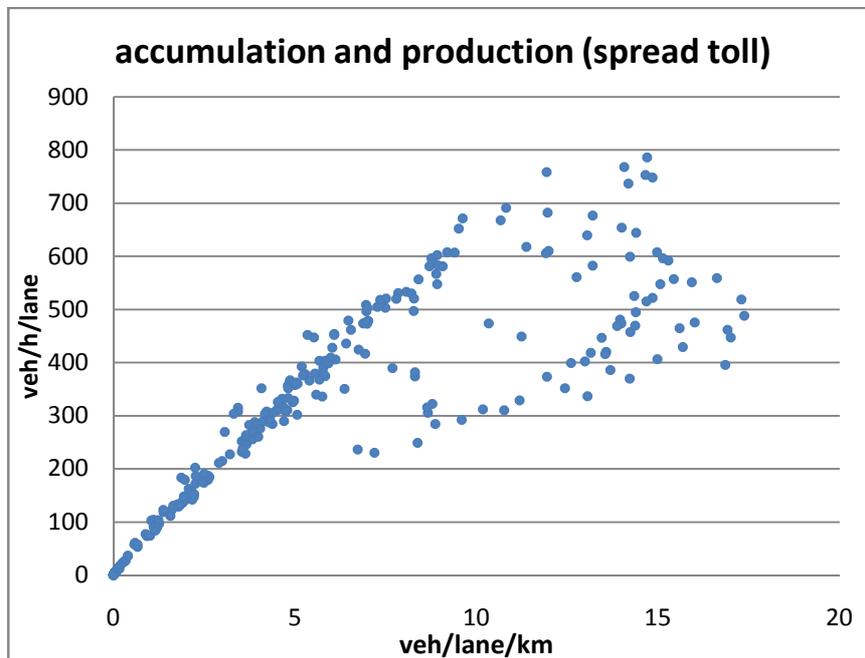


Figure 6.5: Resulting accumulation-production relationship from the Spread Toll

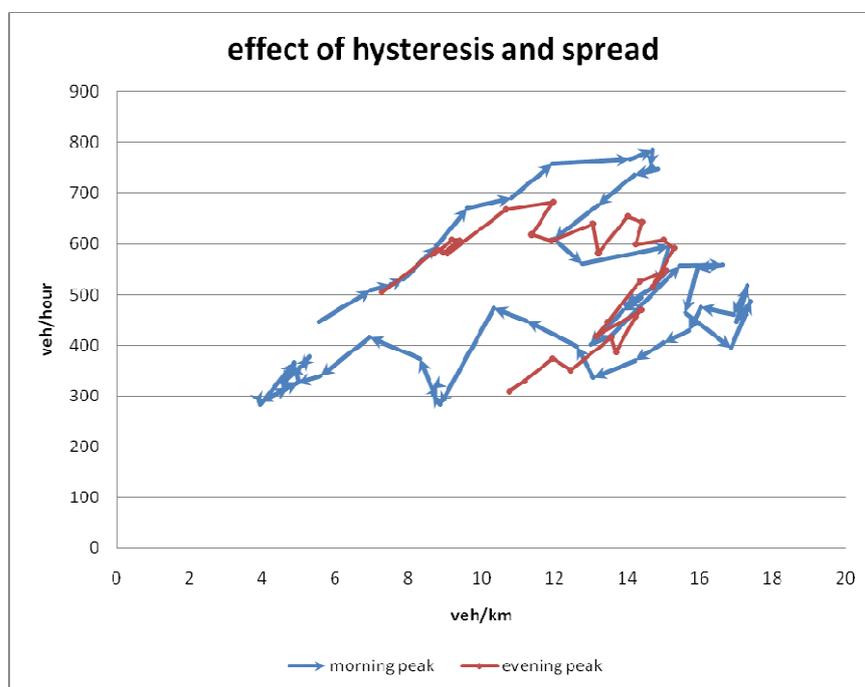


Figure 6.6: Overview of hysteresis patterns associated with the Spread Toll

A more careful analysis (Figure 6.6) reveals that the combination of hysteresis and increases of spread of density play an important role in the drop of performance.. Such a result appears

reasonable as the overall amount of the charge is lower than in the other two schemes and it raises the question whether it would be preferable to apply a “stricter” schemes with higher traffic performances or “milder” ones with lower improvements. Further investigations concerning the economic impacts of the three charges will be conducted in the following sections (6.6 and 6.8).

As to the accumulation, the graph below (Figure 6.7) shows that both the morning and evening peaks are significantly smoothed down. The demand seems to be more elastic in case of Flat Toll, as the amount of traffic travelling into the city centre is lower than in the other two scenarios. Interestingly, the Step Toll seems to generate a shift of trips to the lunch period, that still, however, does not create any decrease of performance. This additional peak might be related to a slight increase of car trips mainly directed to shopping-leisure activities late in the morning. The Spread Toll produces an important decrease of accumulation during the morning peak, while no appreciable reduction is achieved during the evening peak.

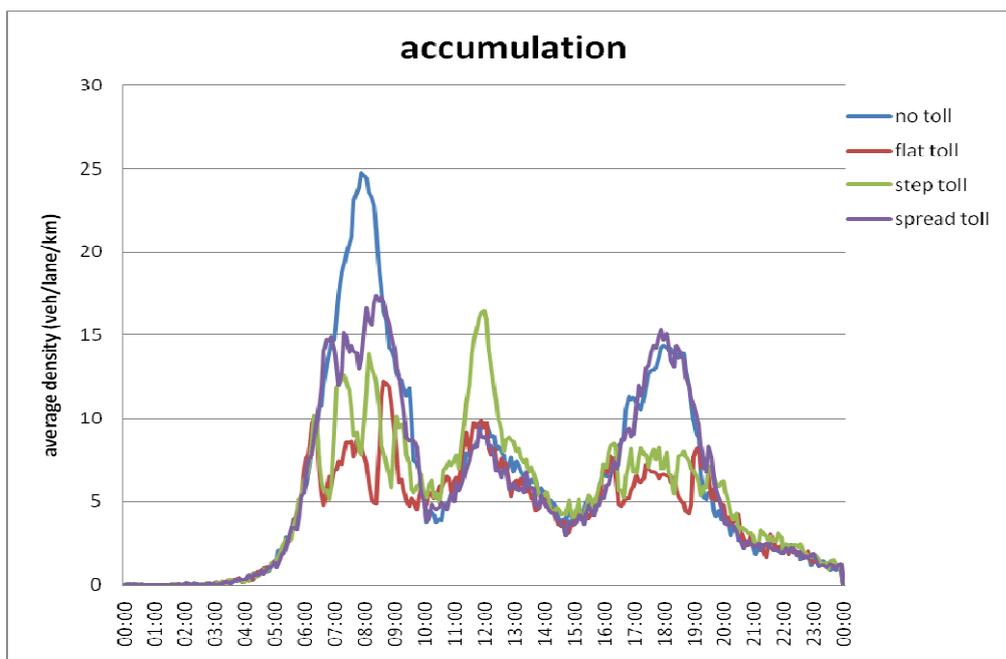


Figure 6.7: Evolution of the accumulation during the day for different tolling schemes

Figure 6.8 representing the evolution of spread of density during the day highlights the different behavior of the network (cordon) determined by the three schemes. It clearly emerges that while the Flat Toll and Step Toll produces an appreciable decrease of spread, no significant improvement is achieved with the Spread Toll. This outcome is not surprising since the main goal of the latter was, indeed, to charge users only for the drops of performance related to the additional traffic they produced.

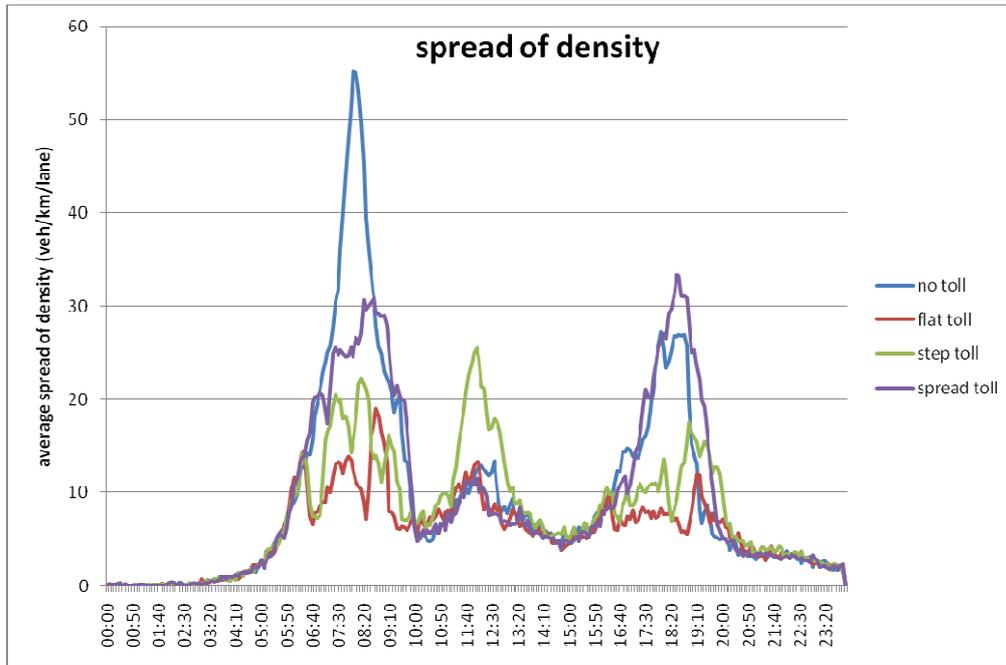


Figure 6.8: Evolution of the spatial spread of density during the day for different tolling schemes

## 6.2 Traffic performance enhancements

The performance of a traffic network can be examined by several perspectives. Each of them can provide a different vision and be interpreted in different ways. For example, a reduction of delays might be determined by a reduction of number and duration of traffic jams or by an increase of average speed during congestion or free flow regime. A decrease of the heaviness of congestion can be determined by both a reduction of the number and length or duration of traffic jams. In order to investigate the effectiveness and efficiency of the tolling schemes, several aspects are considered: traffic efficiency, time savings, decrease of travel demand and heaviness of congestion. Two series of analyses are performed to evaluate the impacts of schemes during the whole day and during the peak hour periods.

### 6.2.1 Traffic improvements for the whole day

In order to investigate the optimal utilization of the network intended as a trade-off between network utilization and network performance, the indicator **Traffic efficiency** is calculated from Brilon *et al.* (2005) as:

$$E = Q \cdot V \cdot T \quad (6.1)$$

where  $Q$  represents the average production (veh/h),  $V$  the average speed (km/h) over the network during a time interval  $T$  of 5 minutes. The factors  $Q$  and  $V$  are easily obtained from the equations 5.14 and 5.15. The results of different tolls are reported in Table 6.1.

The Flat Toll determines a slight overall reduction of efficiency, mainly due the fact that Q is significantly reduced. The Step Toll determines a slight increase of traffic efficiency, while the Spread Toll is the worst performing one.

Traffic Efficiency (veh*km/hour)	Total (x10 <sup>3</sup> )	Total difference (x10 <sup>3</sup> )	change (%)
No toll	473.49	—	—
Flat Toll	468.25	-5237.6	-1.1
Step Toll	483.60	+10.109	2.0
Spread Toll	418.89	-54604.0	-11.5

Table 6.1: Changes of traffic efficiency determined by the tolling schemes for the whole day

Travel delay decreases determined by the alternative tolling schemes are reported in Table 6.2. Both the Flat Toll and the Step Toll produce a significant reduction of **travel delays** (vehicle loss hours) calculated as difference between real travel time and travel time in free flow conditions (with average speed of 70 km/h). Also the Spread Toll determines a little decrease of travel delays.

Travel delays (veh loss hours)	Total (x10 <sup>3</sup> )	Total difference	change (%)
No toll	493.00	—	—
Flat Toll	88.53	-404.47	-82.0
Step Toll	226.11	-266.88	-54.1
Spread Toll	461.26	-31.74	-6.4

Table 6.2: Changes of travel delays determined by the tolling schemes for the whole day

As to the **traffic demand**, expressed as total travelled kilometers by all the vehicles inside the cordon (veh-km), all the tolling schemes seem to determine a decrease (Table 6.3). As expected, the Flat Toll determines the highest reduction, while surprisingly the Step Toll does not produce considerable decreases. A possible reason for that might be the occurrence of an additional peak of traffic during the lunch period that partly counterbalances the reduction produced during the morning and evening peaks.

Traffic demand (veh-km)	Total (x10 <sup>3</sup> )	difference	change (%)
No toll	1378.2		
Flat toll	1211.5	-166.7	-12.1
Step toll	1332.8	-45.4	-3.7
Spread toll	1256.3	-121.8	-9.1

Table 6.3: Changes of traffic demand determined by the tolling schemes for the whole day

In order to represent the extent of congestion both in space dimension (by means of queue length) and in time dimension (by means of duration of congestion) the indicator **heaviness of congestion** is introduced. The index is calculated as a product between the total kilometers of congested links (with density above a threshold value of 35 veh/km) and the time interval of 5 min (0.083 hour). All the schemes seem to determine a significant reduction of heaviness of congestion as Table 6.4 shows. The Flat Toll seems to dissolve almost completely congestion, the Step Toll determines also a considerable decrease and even the Spread Toll almost halves the original value.

Heaviness of congestion (km*hour)	Total	Total difference	change (%)
No toll	120.0	—	—
Flat Toll	26.8	-93.2	-77.6
Step Toll	50.3	-69.7	-58.0
Spread Toll	75.2	-48.8	-40.6

Table 6.4: Changes of heaviness of congestion determined by the tolling schemes for the whole day

### 6.2.2 Traffic improvements during the peak hours

A more specific investigation of changes during the morning and evening peak hours (06:00-10:00 and 16:00-22:00) provides some additional insights about the impacts of tolling schemes. As to the traffic efficiency (Table 6.5), the introduction of tolls seems to be more beneficial in the morning rather than in the evening. While in the morning the congestion is severe and the decrease of performance large, it seems that in the evening traffic is not so heavy to compensate the lower utilization of the network and justify the application of such a demand management measure (traffic efficiency might be intended as a proxy indicator of the optimal usage of transport infrastructures). Probably, this is the reason why the Spread Toll performs equally to other two schemes. However, this is just a single perspective to evaluate the effects of the tolls, as the improvements in terms of travel delays (Table 6.6) and heaviness of congestion (Table 6.7) show. The first ones are significantly reduced both in the morning and evening by all the schemes except of the Spread Toll during the evening peak hours. As to the heaviness of congestion, all the schemes determine important decreases for both the morning and evening peak. Considering that the Spread Toll determines a substantially lower decrease of demand compared to the other schemes these results might be even more appreciated. It is interesting to see that traffic improvements enhanced during the evening peak are equal or even lower than those achieved during the morning peak despite the higher corresponding decrease of demand (Table 6.8).

Traffic efficiency (veh*km/hours)	flat toll	step toll	spread toll
morning peak variation (%)	29.4	26.7	-5.8
evening peak variation (%)	-17.9	-21.6	-21.1

Table 6.5: Traffic efficiency improvements achieved by different tolling schemes during the morning and evening peak

Travel delays (veh loss hours)	flat toll	step toll	spread toll
morning peak variation (%)	-89.9	-75.9	-26.8
evening peak variation (%)	-71.2	-44.0	54.1

Table 6.6: Traffic delays reductions achieved by different tolling schemes during the morning and evening peak

Heaviness of congestion (km*hours)	flat toll	step toll	spread toll
morning peak variation (%)	-85.1	-73.8	-47.8
evening peak variation (%)	-83.8	-73.8	-12.3

Table 6.7: Heaviness of congestion reductions achieved by different tolling schemes during the morning and evening peak

Traffic demand (veh-km)	flat toll	step toll	spread toll
morning peak variation (%)	-13.6	-9.7	-7.7
evening peak variation (%)	-21.7	-20.9	-13.0

Table 6.8: Traffic demand decreases determined by different tolling schemes during the morning and evening peak

The figure below (Figure 6.9) reports the evolution of queue (estimated as a sum of links with density higher than 35 veh/km) during the day under different tolling regimes.

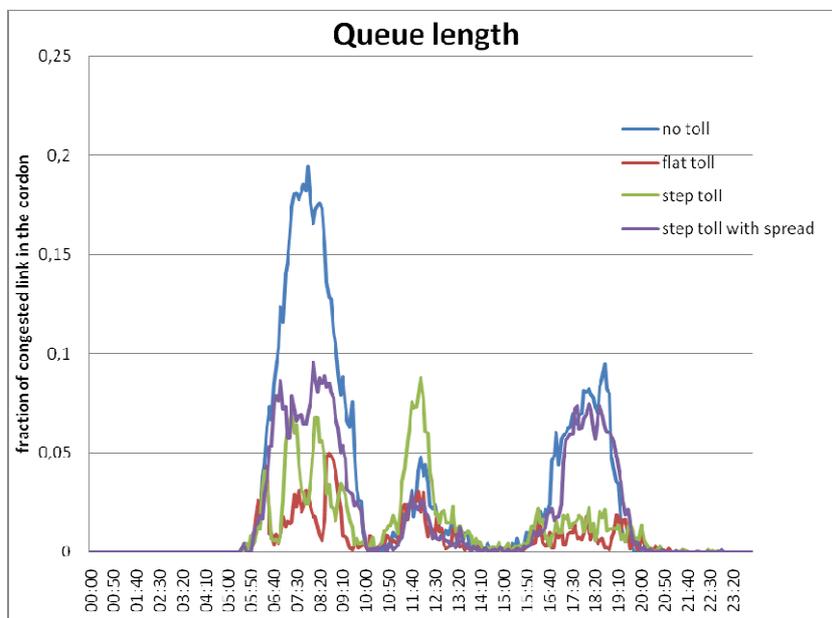


Figure 6.9: Evolution of total level of queue during the day for different tolling schemes

A first inspection clearly shows that the Flat Toll and the Step Toll determine the highest reduction, while the Spread Toll has positive effects only during the morning peak. This outcome seems to confirm the hypothesis that, traffic conditions in the morning are largely

influenced by the entering volumes whereas in the evening they are mainly affected by clustering of flows exiting the area.

### 6.2.3 Considerations of the traffic performance of tolling schemes

When looking at the different performance indicators of the tolling schemes all together (Figure 6.10 and Figure 6.11), it clearly emerges the gap between the Flat Toll/Step Toll and Spread Toll. While the first two produce benefits to the same extent, the latter determines appreciable improvements only during the morning peak. Although the main explanation for that can be found in the significant lower level of fare (higher in case of Flat Toll), the previous analyses have identified some additional influencing aspects like the hysteresis and spread of the density. Above all, the uneven distribution of congestion appears to be a major responsible for the drop of production, especially during the evening peak. In this case, cordon-based toll become rather inadequate in controlling the dynamics of the network and only severe charges can significantly improve traffic conditions. Under these circumstances, the Spread Toll, which on the contrary addresses only those delays ascribable to increases of volumes, becomes totally inefficient.

Furthermore, the occurrence of hysteresis loops related to loading-unloading cycles seems to deteriorate progressively the production like in the case of the Step Toll (Figure 6.4). Further considerations can be found in Section 7.2.

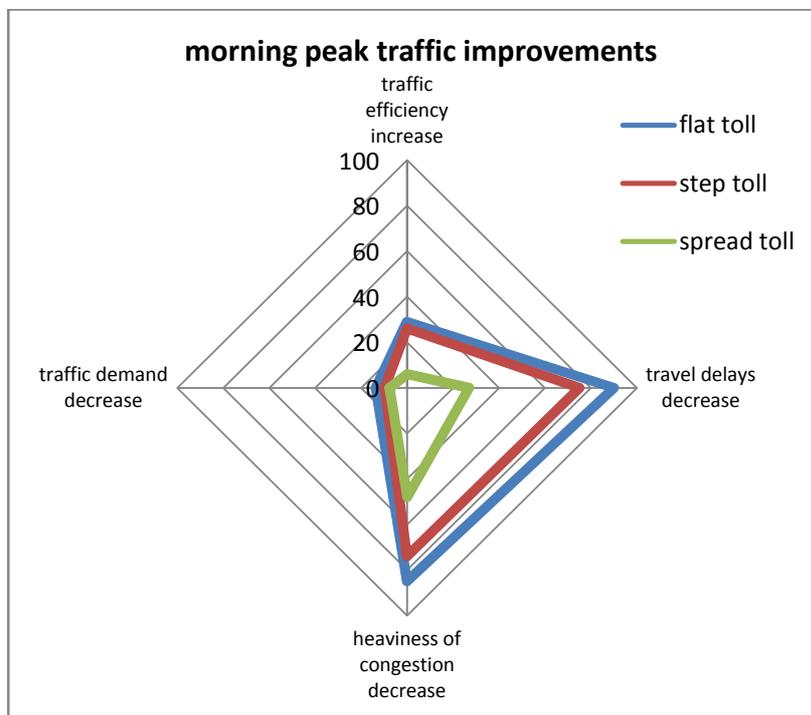


Figure 6.10: Radar chart of traffic enhancements achieved by the alternative tolling scheme during the morning peak

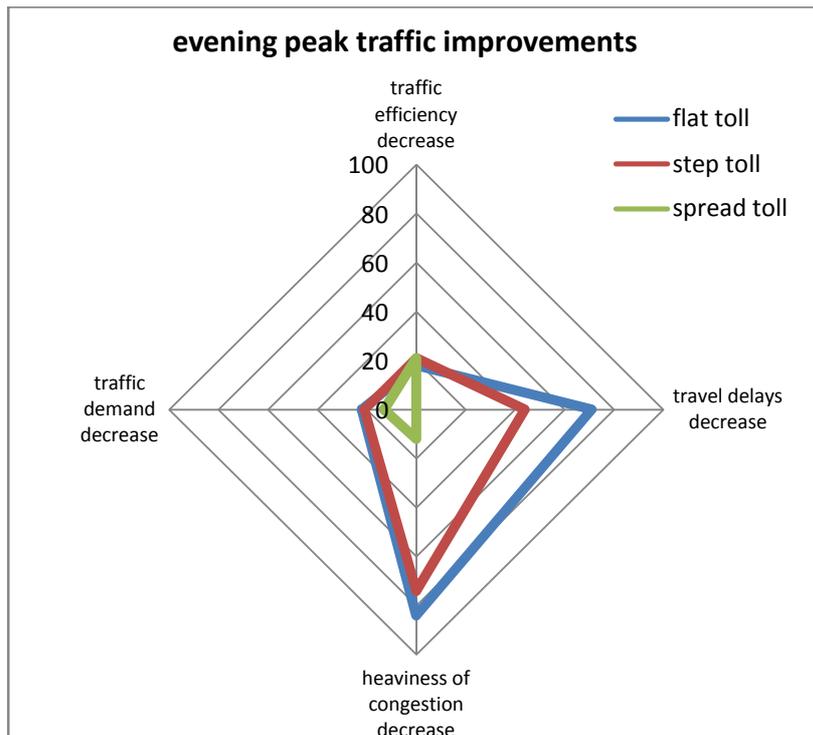


Figure 6.11: Radar chart of traffic enhancements achieved by the alternative tolling scheme during the evening peak

As it will be better illustrated in Section 6.5, these differences in traffic flow operations are reflected by dissimilar travel behavior impacts. The Flat Toll determines a slightly higher reduction of demand compared to the Step Toll, but they both achieve similar traffic improvements. This is an interesting outcome, since the two schemes are characterized by a different conceptual approach and might determine different impacts in terms of economic gains, distributional effects and public acceptability. On the other hand, the fact that the Spread Toll determines lower reduction of demand and smaller improvements is not necessarily a negative point, as it might compensate with higher economic gains (for example, people might benefit from a “softer” approach). For these reasons, in the following sections (Section 6.5, 6.6, 6.7 and 6.8), the economic impacts deriving from such approach will be investigated in depth.

### 6.3 Economic interpretation of results

In the standard economic appraisal resulting travel patterns are typically associated to an economic evaluation of them. For example, as Kickhofer *et al.* (2011) explain, the base case and the tolled scenario can be compared by analyzing for each link the difference in number of users and travel time. Typically, travel time for a link would go down and the number of users would decrease as well. The social welfare change can be estimated by means of the well-known “rule of the half”, which is given by the contribution of the gains/losses of existing users and gains/losses of new users. Let 0 and 1 denote “before” and “after” the congestion-pricing

scheme has been implemented. Then, assuming that the main changes deriving from the toll concern the travel costs and travel time, the total welfare change might be estimated as:

$$W = \frac{1}{2} \cdot (P_0 + P_1) \cdot (c_0 - c_1) + \frac{1}{2} \cdot (P_0 + P_1) \cdot (t_0 - t_1) \quad (6.2)$$

Where  $P_0$  and  $P_1$  correspond to the amount of users before and after the toll;  $c_0$  and  $c_1$  correspond to the average travel costs before and after the toll;  $t_0$  and  $t_1$  correspond to the average travel time before and after the toll.

However, the abovementioned “standard approach” presents three main drawbacks. First, as Arnott *et al.* (1990) observe, not only travel time changes, but also schedule delays effects that might contribute more than half of the economic effects of time-dependent tolls need to be considered. Second, the standard approach does not allow any equity consideration as it is not able to differentiate between subgroups. Finally, the approach needs to assume a uniform value of time for everyone, since neither travelers nor trips purposes are differentiated.

Multi-agent simulations like MATSim partly overcome these issues by allowing a “non-conventional” economic appraisal based on the agents’ utilities as an economic performance indicator of the system. Indeed, the utility (score) of each agent is given by the difference between the gains from performing the planned activities during their preferred schedule and for their duration, and the losses from delays, early and late arrivals, and fares. Hence, the score can properly account for the main effects of congestion pricing such as higher travel costs, changed travel behavior, different travel times and revenues. Thanks to this approach, it is rather straightforward to identify winners and losers of a policy measure by simply looking at the increase/decrease of utility determined by the tolls for every agent or socio-demographic category of users. Furthermore, utilities can be easily aggregated, monetarized and summed to revenues to make some preliminary considerations about the potential level of public acceptability of the tolling scheme. However, the utility approach presents some limitations as well (Kickhofer *et al.*, 2011). First, since the logit model is stochastic, agents are also stochastic (logit-switchers). As a result, some may decrease the systematic part of their utility function. Then, results cannot be interpreted at individual level, but need to be aggregated in sub-populations. Second, it is questionable whether the human behavior really follows such a complex optimization process of the utility (RUM) and even in that case, it is not clear whether some ex-post happiness is optimized by this process.

Individual utility changes can be monetized in different ways for economic evaluations such as the cost-benefit analysis. According to Kickhofer *et al.* (2011) all the individual utility changes

can be summed up and then monetized, or first, individual utility changes can be converted into individual willingness-to-pay, and then summing up.

Following from the assumption that welfare is equal for all the individuals, individual utility changes  $\Delta U_j$  can be summed up and converted by means of an average monetary value of utility  $\frac{1}{n} \frac{\sum_j^n y_j}{\alpha}$  where corresponds to daily income of a person and  $\alpha$  is an empirical coefficient of conversion. The toll revenues  $c_{toll,i}$  can be summed up indifferently as well. Hence, the resulting overall welfare change of the society  $\Delta W$  corresponds to:

$$\Delta W = \frac{1}{n} \frac{\sum_j^n y_j}{\alpha} \cdot \sum_j^n \Delta U_j + \sum_i^n c_{toll,i} \quad (6.3)$$

In practice this approach is widely used in the evaluation of transport projects where usually travel time saving are aggregated and then multiplied for an average Value-of-Time (VOT).

An alternative approach might consist in using the individual willingness-to-pay as indicator to describe changes in the society's welfare level. In this case, the conversion from utils to money is done at individual level with person specific values. As a result, the overall welfare change of the society  $\Delta W$  would correspond to:

$$\Delta W = \sum_j^n \Delta U_j \cdot \frac{y_j}{\alpha} + \sum_i^n c_{toll,i} \quad (6.4)$$

Where the first summand represents the sum of monetized utility changes. In this case, it is dependent on the individual utility change and on the reciprocal value of the income dependent marginal utility of money,  $y_j/\alpha$ . The second summand is just the sum of all the toll payments like in (6.3).

#### **6.4 Identification of socio-demographic categories**

As it has been already discussed in Chapter 2, the concept of equity can be considered from several perspectives (opportunities, wealth, utility, etc.), each one implying a different approach in the analyses and ultimately disparate evaluations. While some transportation planning literature focuses on the distributional effects on individual characteristics like gender and age, the transportation economics literature mainly refers to equity in regards to the impacts on different income categories and working/living locations. Since our investigations will deal with the question of equity from the economics perspective the three main following socio-demographic classifications of have been defined: the living location relative to the cordon, the neighborhood of residence and the purpose of the trip.

The first one, concerning the living location, groups the agents in three zones (see Figure 6.12): Zone 1 which corresponds to the area circumscribed by the cordon (radius of 1.5 km); Zone 2 which corresponds to an external circumference of 3 km; and Zone 3 that includes the outer area (broader region including the Canton of Zurich and other ones). Such differentiation has been conceived mainly to identify possible “inequities” between people living inside the cordon, in the proximity and further distant. For example, people living within the cordon, given its relatively limited extension (about 7 km<sup>2</sup> in comparison with other cities where congestion pricing has been implemented like Stockholm: 47 km<sup>2</sup> ; London: 16 km<sup>2</sup>; Singapore: 7 km<sup>2</sup> <sup>6</sup>), might be particularly worsened and suffer from a sort of “segregation effect”. Alternatively, people living in the proximity of the cordon might be affected by increased traffic due to rat-running phenomena, and people living at long distances from the city centre might be forced to continue to enter the cordon by car because of inadequate transport alternatives.

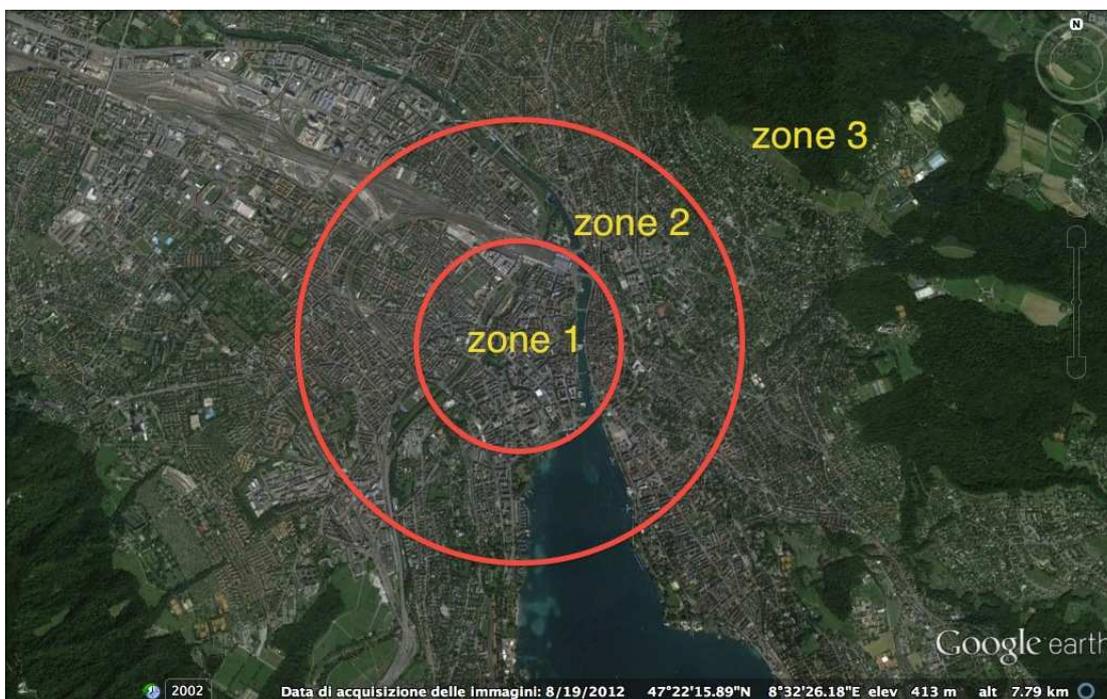


Figure 6.12: socio-economic differentiation in three living areas

A second geographic classification consists of a more detailed subdivision of the municipality of Zurich in the current administrative districts (Kreis in German). The city has been split in twelve areas (see Figure 6.13) with population varying between 5.000 and 50.000 agents in order to identify possible clusters of discomfort deriving from the schemes and make further considerations upon the question of public acceptance. This focus has been adopted with explicit reference to the previous experiences of congestion pricing described in section 2.1 and

<sup>6</sup> Note that cities Singapore and London characterized by very high densities of activities in the central area do not run the same risk of segregation

section 2.2. Several proposed tolling schemes in the past were actually stopped by users from those areas who felt most disadvantaged by the charge (e.g. Manchester, Edinburgh and NYC). Hence, identifying the most problematic clusters and potential opponents during the design stage might help to identify the most suitable solutions to achieve “fairer” schemes and ultimately higher levels of acceptability.

Finally, the third classification concerns the purpose of the trips (crossing the cord) that have been divided in four main categories: home, work, education and shopping-leisure. Each one is characterized by a different scoring function, minimal duration and opening times (see Section 5.1). This classification might provide further insights of the typology of agents affected by the toll such as commuters or people directed to the centre for shopping activities. The importance of this aspect is clear, as it higher burden to work trips might generate resistance of commuters, while worsened conditions for shopping and leisure trips might raise opposition from owners of commercial activities.

Although information concerning the income of agents was not available in this scenario, it would be interesting in future studies to investigate more deeply the vertical equity issues by including in the analyses a classification based on agents’ salary (and different VOTs).

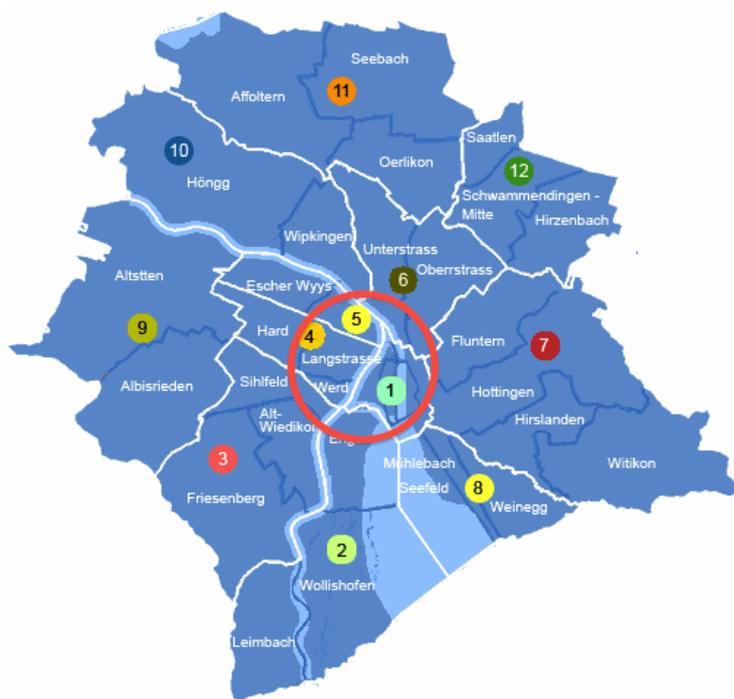


Figure 6.13: The residential districts of Zurich

## 6.5 Travel behavior changes

About 36,000-38,000 agents out of 180,000 are travelling across (inbound and outbound) the cordon. Car share represents about 50% of these trips. Hence, only 17,000-18,000 of the agents are more less directly affected by the tolling schemes (if only trips during the peak hours are considered the amount even drops to 8,000-9,000 agents).

The Flat Toll and Step Toll determine a significant reduction of demand (trips crossing the cordon in both directions), respectively about 10% and 7%, while the Spread Toll has almost no influence. The Flat Toll causes an overall decrease of 1975 trips during the entire day. The Step Toll determines a lower decrease equal to 1250 trips. The Spread Toll determines a slight increase of 200 trips. These different behavioral trends are confirmed the number of payments<sup>7</sup> corresponding to the tolling schemes. The amount of payments in the Flat Toll, Step Toll and Spread Toll scenarios corresponds respectively to 6,887, 8,970 and 5,328. This result is consistent with previous studies (Van den Berg, 2012) that showed how a uniform toll such as the Flat Toll determines an higher reduction of demand than time-varying tolls. The Spread Toll is characterized by a low number of payments because it operates during a shorter time period. On the other hand, given the lower level of fare, users will be more willing to pay and cross the cordon.

Concerning the travel behavior changes of those agents who stopped travelling inbound and outbound the tolled area (payers), the major impact in all the three scenarios consists of rerouted trips (between 50-70%). It is worth mentioning that all these trips have origins and destinations outside the cordon (they simply travelled through city centre), since no relocation of activities is implemented in the simulation. About 10%-20% of the trips are made at different times and only a small percentage of the car trips is replaced by public transport (PT) trips (between 0% and 15% according to the scheme). In the Step Toll scenario the modal share is even characterized by a slight increase (+1%) of car trips.

### 6.5.1 Impacts on socio-demographic groups

When looking more specifically at travel behavior changes of inhabitants from different areas, it clearly emerges from the tables below (Table 6.9 and Table 6.10) that people living outside the cordon are characterized by higher elasticity and people living inside the cordon seem to be not very sensitive to the toll in terms of decision of crossing the cordon. On the other hand, people from Zone 1 seems to be more inclined to switch from car trips to pt trips, especially when compared with inhabitants from the Greater Zurich Area. This is an expectable results, since

---

<sup>7</sup> Note that payment refers the act of crossing the cordon and there in no reference to the amount of charge paid

people who are living inside the cordon might be somehow obliged to cross it to access activities (e.g. work) located outside. Then, switching mode might represent for this people the only available alternative to avoid the toll.

As concern differences among the schemes, the Flat Toll seems to produce highest reduction of car trips through the cordon thanks to a considerable amount of rerouted trips and new trips by PT. The Step Toll apparently determines higher rerouting rather than modal shift. Perhaps, the decrease of pt share might be related to the new peak of traffic arising late in the morning (see Sub-section 6.2.1).

Decrease/Increase of trips crossing the cordon	Flat Toll	Step Toll	Spread Toll
Zone 1 (%)	0.0	0.0	0.0
Zone 2 (%)	-3.06	-2.94	-0.98
Zone 3 (%)	-4.30	-3.81	-1.89

Table 6.9: Decrease/increase of trips crossing the cordon in the three identified zones

Decrease/Increase of pt trips	Flat Toll	Step Toll	Spread Toll
Zone 1 (%)	5.96	-3.17	2.20
Zone 2 (%)	4.74	-6.20	2.54
Zone 3 (%)	1.81	-5.46	1.79

Table 6.10: Decrease/increase of pt trips in the three identified zones

Table 6.11 and Table 6.12 provide a more detailed overview of the travel behavior changes deriving from the schemes at neighborhood level. From this perspective, the districts characterized by higher rerouted trips are those situated in the eastern part of the city (Kreis 7, Kreis 8, Kreis 12). The highest amount of new pt trips is produced in the districts whose the largest portion of their area is inside the cordon (Kreis 1, Kreis 4, Kreis 5).

Decrease/Increase of trips crossing the cordon <sup>8</sup>	Flat Toll	Step Toll	Spread Toll
kreis 3 (%)	-0.58	-0.63	-0.79
kreis 9 (%)	-2.28	-2.54	-0.66
kreis 10 (%)	-2.26	<b>-2.88</b>	-0.96
kreis 6 (%)	-2.49	<b>-2.66</b>	-0.71
kreis 7 (%)	<b>-2.87</b>	-1.67	<b>-1.67</b>
kreis 11 (%)	-1.63	-1.58	-0.74
kreis 12 (%)	<b>-2.73</b>	-1.24	-1.24
kreis 8 (%)	<b>-3.76</b>	<b>-2.87</b>	<b>-1.55</b>
kreis 1 (%)	0.93	0.70	-0.35
kreis 4 (%)	-0.19	0.14	-0.23
kreis 5 (%)	-1.40	-0.20	<b>-1.60</b>

<sup>8</sup> The bold values indicates the top three districts in terms of decreased trips crossing the cordon

Table 6.11: Decrease/increase of trips crossing the cordon in the districts of Zurich

Decrease/Increase of pt trips <sup>9</sup>	Flat Toll	Step Toll	Spread Toll
kreis 3 (%)	2.09	-5.62	-1.38
kreis 9 (%)	-0.11	-7.65	-0.43
kreis 10 (%)	3.55	-16.0	-1.01
kreis 6 (%)	6.47	-4.31	1.60
kreis 7 (%)	3.91	-5.26	1.79
kreis 11 (%)	-0.83	-5.71	-2.12
kreis 12 (%)	1.40	-8.49	4.89
kreis 8 (%)	2.94	-6.23	1.87
kreis 1 (%)	<b>7.18</b>	-3.17	<b>5.65</b>
kreis 4 (%)	<b>4.55</b>	-1.59	<b>3.81</b>
kreis 5 (%)	<b>8.82</b>	-0.41	<b>6.67</b>

Table 6.12: Decrease/increase of pt trips in the districts of Zurich

Trips directed to work activities are the ones characterized by higher increase of rerouting and modal shifts (Table 6.13 and Table 6.14). Trips directed to home are also characterized by a considerable decrease of “tolled” trips, followed by trips directed to educational places and last by trips to shopping and leisure activities. A possible explanation for this phenomenon can be found in the physiological nature of the tolling schemes that are implemented during the typical morning and evening commuting periods.

Decrease/Increase of trips crossing the cordon	Flat Toll	Step Toll	Spread Toll
home (%)	-8.89	-8.34	-8.07
work (%)	-8.28	-7.94	-5.30
education (%)	-4.84	-4.72	-3.74
shop&leisure (%)	-2.10	-1.92	-1.87

Table 6.13: Decrease/increase of trips crossing the cordon of different purposes of trips

Decrease/Increase of pt trips	Flat Toll	Step Toll	Spread Toll
home (%)	5,82	-1.49	2.73
work (%)	8.60	-0.55	5.28
education (%)	2.22	-1.16	1.42
shop&leisure (%)	-2.64	-6.75	-0.48

Table 6.14: Decrease/increase of pt trips of different purposes of trips

### 6.5.2 Influence of socio-demographic characteristics on travel behavior changes

In order to investigate the influence of socio-demographic characteristics on the travel choices of avoiding the cordon and replacing car with PT, a regression analysis is required. Given the dichotomous nature of the dependent variable (stop entering the cordon/switching to PT), a logistic regression is performed to identify the probability “given” by the independent variables (socio-demographic category). The dependent variable is set to 0/1 and the independent

<sup>9</sup> The bold values indicates the top three districts in terms of increased pt trips

variables are coded by means of dummy variables. A more detailed description of the logistic regression analysis and the interpretation of the outputs is provided in the Appendix II.

Overall, the typology of activity and living location do not seem to play a major role in the decision to avoid the toll by means of rerouting and modal shift. Precisely, both the Nagelkerke  $R^2$  (close to 0) the  $\chi^2$  (from the Hosmer-Lemeshow test) aimed at expressing the goodness of fit, demonstrate that the variables did not appreciably improve the null model. However, the table below (Table 6.15) highlights that purpose of the trip (activity) has an appreciable influence on the decision of rerouting when Flat Toll and Step Toll are applied (highest adjusted R square values). As it has been mentioned already, this fact may be determined by the nature of the tolling scheme itself that operates during periods of the day while mainly trips directed from home to work and vice versa occur.

R square	Flat Toll		Step Toll		Spread Toll	
	Living place	activity	Living place	activity	Living place	activity
Stop entering	0.0019	0.421	0.014	0.356	0.006	0.076
switching mode	0.058	0.04	0.001	0.015	0.001	0.008

Table 6.15: Overview of influence of socio-economic variables on travel choices expressed by the R square indicator

The two tables below show the regression coefficients indicating the effect of activity variables on the decision to stop entering the cordon. In this case all the coefficients are statistically significant as their p-values<sup>10</sup> indicate (lower than 0.05). Higher values of Beta represent higher probability of stop crossing the cordon (dependent variable equal to 1). As it possible to see from the parameter  $Exp(B)$  in the Table 6.16, when a Flat Toll is applied, trips directed to home and work have the highest probabilities to be rerouted, i.e. the probability of rerouting of one of these trips is three and seven times higher than education and shopping-leisure related trips. The same interpretation applies for Table 6.17 representing the output of the logistic regression for the Step Toll.

Flat toll	Beta	p value	Exp(B)
home	-0.295	0,000	0.744
work	-0.999	0.000	0.368
education	-2.064	0.000	0.127
shop&leisure	-1.84	0.000	0.161

Table 6.16: Regression analysis of influence trip purposes on the choice of stop entering the cordon determined by the Flat Toll

<sup>10</sup> The p-value tests the hypothesis that each coefficient is different from 0.

Step toll	Beta	p value	Exp(B)
home	-1.211	0.000	0.298
work	-3.468	0.000	0.031
education	-4.000	0.000	0.018
shop&leisure	-3.902	0.000	0.020

Table 6.17: Regression analysis of influence trip purposes on the choice of stop entering the cordon determined by the Step Toll

### 6.5.3 Economic impacts of travel behavior changes

At this point it would be interesting to investigate whether such a behavioral response would imply any particular trend in the economic impacts. In order to do that, a t-test is performed to compare the average increase/decrease of utility of the agents who enter the cordon regardless of the toll and of those who stop entering.

As it possible to see from Table 6.18 and Table 6.19, comparing the two averages in case of Flat Toll and Step Toll, an opposite trend characterizes the two categories. While people who continue entering the cordon seem to benefit from the introduction of the toll, those who decide to avoid are worsened. The Step Toll seems to produce higher benefits for both the categories compared with the Flat Toll that seems to penalize especially users who stop to enter the cordon. Given the high values of standard deviation (way larger than the mean), additional analyses of the distributions: skewness and kurtosis, are carried out and reported in the Appendix III. The outcomes highlight in all the cases, a presence of fat-tailed distribution, characterized by acute peaks around the mean and fatter tails. These results ultimately imply that the effects (positive and negative) of the tolls among these groups may vary considerably.

Flat toll	mean	Std dev
continue entering	0.51	32.03
stop entering	-2.4	30.96

Mean difference: 2.9; standard error difference: ; p-value: 0.00005;

Table 6.18: t-test analysis of groups of users who continue entering the cordon and those who stop in case of Flat Toll

Step toll	mean	Std dev
continue entering	1.76	32.74
stop entering	-0.33	31.59

Mean difference: 2.10; standard error difference: ; p-value: 0.001;

Table 6.19: t-test analysis of groups of users who continue entering the cordon and those who stop in case of Step Toll

To sum up, from the previous analyses focused on travel behavior changes it emerged that, in case of Flat Toll and Step Toll, the purpose of the trip might play a role in the decision of avoiding the charge by rerouting. Indeed, trips directed to home and work seem to have higher

probabilities to be rerouted in order to avoid the toll. Furthermore, a preliminary investigation aimed at identifying the economic impacts of such a response, has underlined how people who decide to stop entering the cordon have (on average) lower utility. However, it is not possible to conclude from this basic analysis that the reason behind different economic impacts consists of such a travel behavior choice, as several other factors might affect the final utility gains/losses of agents.

## 6.6 Economic efficiency and distributional impacts of the schemes

Changes of average utility of the whole population and of specific socio-demographic categories represent a straightforward indicator of the efficiency and equity effects determined by the toll. Then, a thorough investigation of trends due to the different schemes will be carried out in the following sections with the support of statistical t-tests to verify the significance of these changes. Finally, a more specific analysis consisting of multiple linear regression will be performed in order to identify specific relations between socio-demographic characteristics and the economic impacts.

### 6.6.1 Comparison of overall impacts deriving from the schemes

A comparison of the average utility of agents for the whole population provides a first indication of the economic efficiency of tolling schemes, intended as ability to maximize the overall benefits of people. From this perspective, the Step Toll seems to be the most efficient scheme, followed by the Flat Toll and the Spread Toll that even determines negative economic impacts (on average). As Although the standard deviation is rather high, it does not compromise the significance of the t-test reported in Table 6.20, aimed at verifying the change between average utilities. The presence of high values of standard deviation indicates that the impacts are rather spread out among the population.

	mean	Std dev	Std error mean	Significant variation
<b>Flat toll</b>	0.592	32.254	0.163	yes
<b>Step toll</b>	2.007	32.038	0.162	yes
<b>Spread toll</b>	-0.49	31.788	0.161	yes

Table 6.20: overall comparison of impacts of tolling schemes in terms of average utility changes

### 6.6.2 Impacts of tolling schemes on different socio-demographic categories

An overview of the economic impacts expressed as average increase-decrease of utility across the different groups is presented in Figure 6.14, Figure 6.15, Figure 6.16 and further discussed in the Appendix IV. In order to reduce the impacts of outliers, the two tails of the probability distribution corresponding to the highest 2.5% increases and lowest 2.5% decreases are excluded from the calculations.

As concern the factor of proximity to the cordon, the tolling schemes seem to produce different impacts (Figure 6.14). The Step Toll determines considerable gains for all the groups, especially for people living in the Greater Zurich Area. The Flat Toll seems to benefit everyone in equal way regardless of the distance from the cordon. Conversely the Spread Toll determines a rather uniform decrease of utility. A t-test analysis of averages of different groups identifies the improvement in Zone 3 determined by the Step Toll, and the worsening in Zone 2 and Zone 3, as the only significant relative changes from the Flat Toll (see Appendix IV). The standard deviation of the utility change (between 18 and 21 depending on the scheme), the skewness and kurtosis values highlight fat-tailed distributions for the three different groups, which indicate the impacts still vary considerably even for the same category.

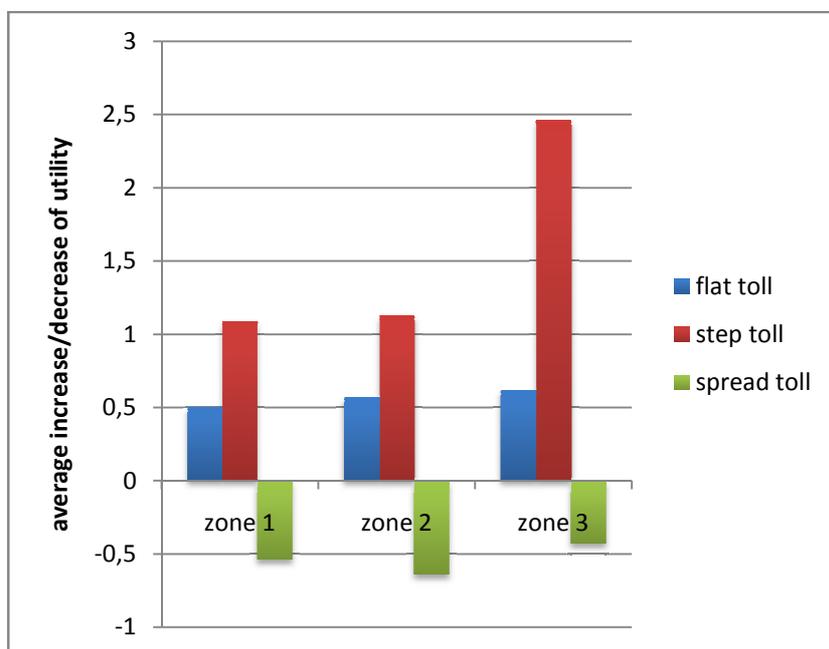


Figure 6 14: average increase/decrease of utility in the three identifies zones

A more detailed analysis of the impacts at neighborhood level (Figure 6.15) provides a better insight into the effects of tolling schemes. This investigation provides a clear picture of the most “critical” districts: Kreis 4 and Kreis 5, that are worsened especially by the Flat Toll and Spread Toll. These two neighborhoods are situated in the central area of the city and a large portion of their area is within the cordon. For this reason the Flat Toll, characterized by lower flexibility, seems to affect them more than the rest of the city. The Spread Toll, does not produce enough improvement of traffic conditions (especially during the evening) to compensate the burden of the charge, resulting in an overall decrease of utility.

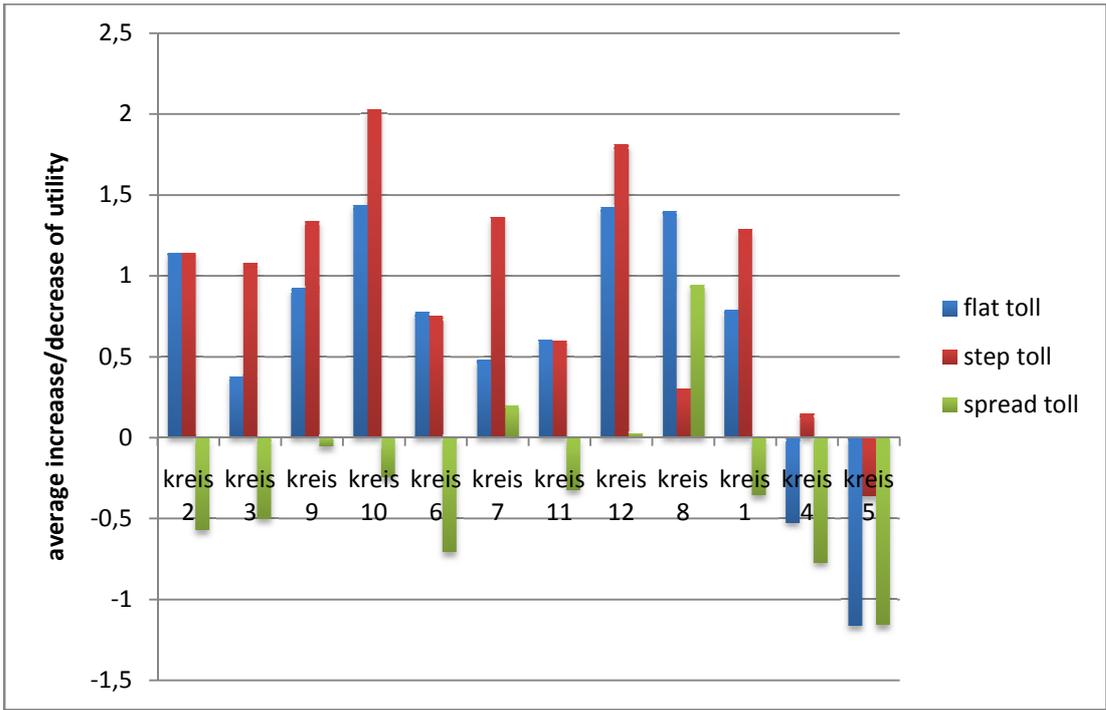


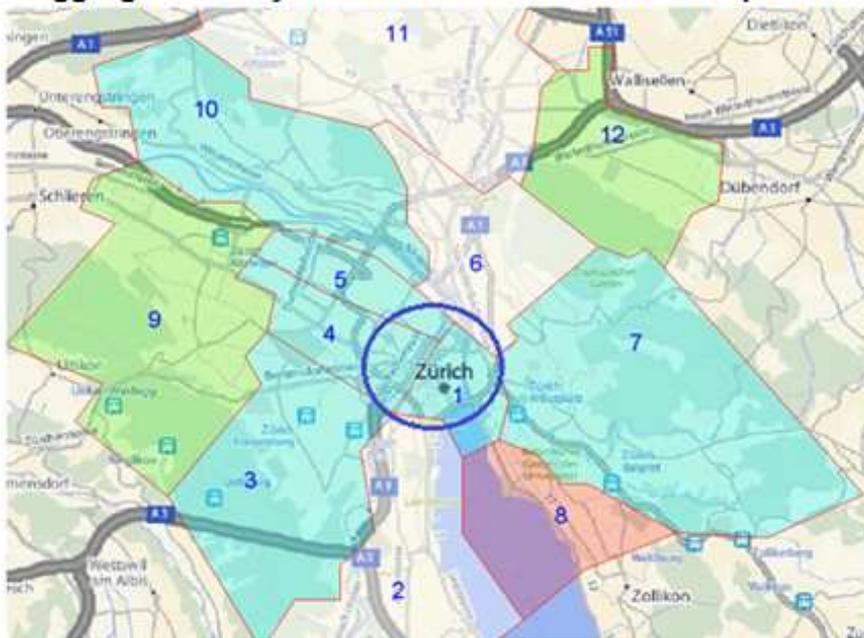
Figure 6.15: average increase/decrease of utility across the districts of Zurich

Figure 6.16 and Figure 6.17 represent respectively the significant relative changes across the different districts determined by the Step Toll and Spread Toll derived from a t-test analysis. The different colors express the relative in/decrease of utility of the two time-varying schemes compared with the Flat Toll. As concern the variance of results within the same district, all the neighborhoods are still characterized by relatively high variance among users.

It is worth mentioning that inhabitants of districts partly inside the cordon or in the very proximity such as Kreis 3, Kreis 4, Kreis 5 and Kreis 7 are the ones who benefit the most from the introduction of a Step Toll (blue zones is Figure 6.16). Another possible factor of influence might be the presence of important road connectors between the South-West and North-East part of Zurich (provincial roads 1, 3, 4). Indeed, traffic conditions in the proximity might become more fluid with a more flexible toll because of a smaller amount of diverted traffic from the cordon (rat running phenomenon).

Figure 6.17 representing the decrease of average changes of utility for the districts also highlights some interesting outcomes. Indeed, Kreis 2, Kreis 1, Kreis 6 and Kreis 12, that create a sort of corridor on the South-West and North-East axis, are the most worsened by the Spread Toll in comparison to the Flat Toll. That means that a large part of traffic coming from the two major motorways (A1, A3) still pass through the city centre of the city.

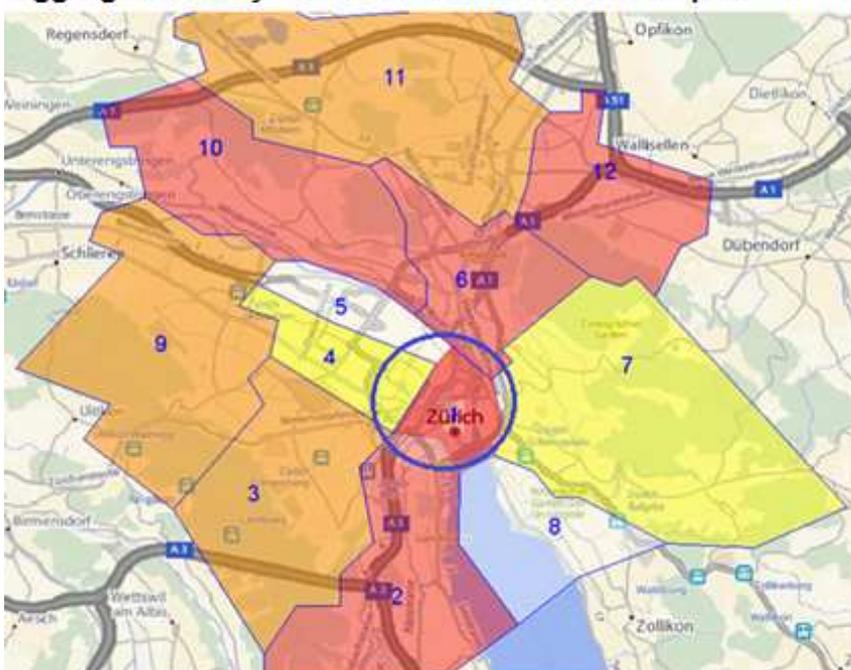
### Aggregated utility increases/decreases of the Step Toll



Legend: in light-blue average increases of score between 0.00-0.50; in green average increases of score between 0.50-1.00; in red decreases of score higher than -1.00; in grey not significant changes.

Figure 6.16: average utility changes (relative to the Flat Toll) determined by the Step Toll across the districts of Zurich

### Aggregated utility increases/decreases of the Spread Toll



Legend: in yellow average decreases of score between 0.00-0.50; in orange average decreases of score between 0.50-1.00; in red decreases of score higher than -1.00; in grey not significant changes

Figure 6.17: average utility changes (relative to the Flat Toll) determined by the Spread Toll across the districts of Zurich

As to the impacts on different typologies of trip, the three schemes produce considerably different results (Figure 6.18). While the Step Toll seems to be beneficial to all the categories of trip, the Flat Toll has a negative impacts on those trips directed to working activities. The Spread Toll seems to be again disadvantageous for all the trips.

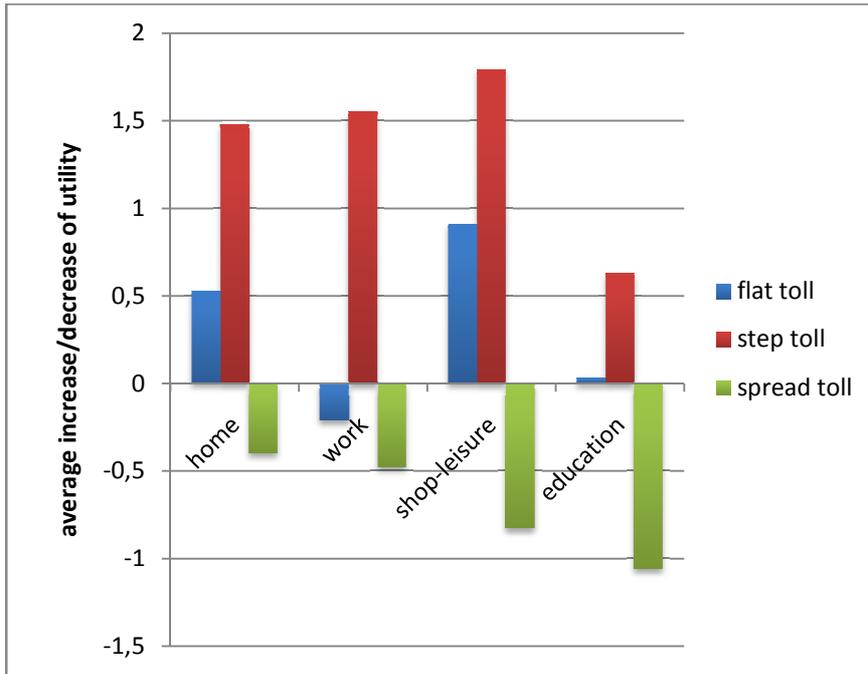


Figure 6.18: average increase-decrease of utility for different trip purposes

The results of the t-test (see Appendix IV) confirm that the two time-varying tolls determine significant changes of impacts compared to the Flat Toll. The Step Toll improves the scores for all the activity-related trips and in particular for the ones directed to work activities. On the other hand, the Spread Toll determines a considerable decrease of the average score for trips directed to home and to shopping and leisure activities.

### 6.6.3 Relationship between socio-demographic characteristics and economic impacts

In order to verify whether belonging to any specific socio-demographic groups influences the economic impacts of the toll, a multiple linear regression is performed. The implicit assumption of a linear relation between the dependent variables (socio-demographic features) and the independent one (change of utility) has been made in order to allow a more straightforward comparison of the groups. This type of analysis helps to identify possible categories of winners and losers expressed by means of a coefficient *Beta*. The higher the coefficient, the higher increase will be (likely) obtained by the agent belonging that specific category as it shown by equation xx :

$$\Delta Utility = \beta_o + \beta_i \cdot x_i$$

Where  $\beta_o$  represents a constant,  $\beta_i$  are the regression coefficients of  $x_i$  dependent variables expressed by means of dummy variables (0/1). The standardized regression coefficients<sup>11</sup> (*Beta*

<sup>11</sup> Standardized regression coefficients re-express the variation of the dependent variable as effect of one standard deviation change of the independent variable (holding all the others constant)

*stand*) are also estimated in the analyses in order to allow an easier comparison of the influence of different independent variables. The variance inflation indicator (*VIF*) for every variable is also estimated in order to identify any possible multicollinearity (and ultimately latent variables). As a rule of thumb when the *VIF* is higher than 10, the variable may merit further investigations.

More sophisticated methods might have been performed for this typology of investigations (by means of non-linear models). However, since the goal is to highlight different trends across the groups rather than determine a predictive function based on the characteristics, the above mentioned approach has been preferred for its relative simplicity.

The influence of these variables seem to be overall quite marginal and not always significant. Living location seems to affect the score of the agents in case of Flat Toll and Step Toll, while the activity performed has little influence in all the scenarios (Table 6.21).

	Flat toll		Step toll		Spread toll	
	living place	activity	living place	activity	living place	activity
<b>significant</b>	yes	yes	yes	yes	no	yes

Table 6.21: significant relationships between socio-demographic characteristics and economic impacts

Table 6.22 represents the linear regression for activity variables when the Flat Toll is applied. The results confirm that work activities are more penalized compared to home (constant). Shopping-leisure activities have no impact on the variation of scores. Table 6.23 certifies that two districts: Kreis 4 and Kreis 5, are particularly worsened by the application of the Flat Toll. Indeed, their regression coefficients are not only negative, but they also have higher absolute value than the constant.

Flat toll	Beta	Beta stand	t-value	p-value	VIF
<b>Constant (home)</b>	1,539	-	4,333	0,000	-
<b>Work</b>	-1,689	-0,026	-4,435	0,000	1,32
<b>Education</b>	-1,323	-0,012	-2,237	0,025	1,15
<b>Shop&amp;leisure</b>	-0,180	-0,003	-0,482	0,630	1,28

Table 6.22: Linear regression. Relationship between trip purpose and utility changes (Flat Toll)

Flat toll	Beta	Beta stand	t-value	p-value	VIF
<b>Constant(Kreis 12)</b>	1,412	-	3,499	0,000	-
<b>Kreis 4</b>	-1,940	-0,023	-3,153	0,002	1,58
<b>Kreis 5</b>	-2,580	-0,015	-2,441	0,015	1,14

Table 6.23: Linear regression. Relationship between residential area and utility changes (Flat Toll)

In case of Step Toll trips directed to work are slightly penalized compared to home and to the same extent of educational and shopping-leisure related trips (all significant, Table 6.24). As concern the living location, Kreis 4 and Kreis 5 seems to be again the most disadvantaged neighborhoods, but to lesser extent than in the Flat Toll (Table 6.25). For example, people living in Kreis 4 is not anymore worsened by the tolling scheme (the net utility change is positive). These results are somehow expected since a more flexible scheme such as the Step Toll should advantage those groups that are more affected by the tolls for reasons of time (commuters) and space (residents inside or in the proximity of the cordon).

Step toll	Beta	Beta stand	t-value	p-value	VIF
Constant (home)	3,577	-	10,305	0,000	-
work	-1,658	-0,026	-4,389	0,000	1,32
education	-2,520	-0,023	-4,305	0,000	1,15
Shop&leisure	-1,290	-0,02	-3,489	0,000	1,28

Table 6 24: Linear regression. Relationship between trip purpose and utility changes (Step Toll)

Step toll	Beta	Beta stand	t-value	p-value	VIF
Constant (Kreis12)	1,812	-	3,579	0,000	-
Kreis 4	-1,661	-0,020	-2,422	0,015	1,98
Kreis 5	-2,717	-0,013	-1,98	0,048	1,24

Table 6 25: Linear regression. Relationship between residential area and utility changes (Step Toll)

The Spread Toll does not produce any considerable imbalance of impacts among the activity related trips. Work, educational and shop-leisure trips are slightly worsened compared to home (Table 6.26). As concern the geographic distribution of effects, the linear regression analysis have not produced any significant output. This finding implies that apparently no specific living location is either favored or disadvantaged by the scheme. Then, from an “equity perspective” this aspect can be interpreted as a positive feature of the scheme.

Spread toll	Beta	Beta stand	t-value	p-value	VIF
Constant (home)	0,62	-	1,801	0,072	-
work	-1,012	-0,016	-2,72	0,007	1,23
edu	-1,353	-0,013	-2,332	0,02	1,54
shop&leisure	-1,179	-0,018	-3,201	0,001	1,36

Table 6 26: Linear regression. Relationship between trip purpose and utility changes (Spread Toll)

## 6.7 The profile of winners and losers

Once all the statistical tests have been conducted, it is finally possible to identify the typical profile of winners and losers deriving from the tolling schemes. Figure 6.19 provides a synthetic outline of the distributional effects characterizing the different tolls.

	losers	winners
<b>Flat Toll</b>	-Inhabitants of kreis 4 and kreis 5 -Trips directed to work	-Inhabitants of kreis 2, kreis 10, kreis 12, kreis 8 -Trips directed to home and shopping-leisure activities
<b>Step Toll</b>	-Inhabitants of kreis 5 (but to a lesser extent)	All the users seem to benefit from this scheme, but in particular: -Inhabitants from kreis 1, kreis 4, kreis 10, kreis 12 -Trips directed to work
<b>Spread Toll</b>	No real distributional effects. All the identified socio-demographic categories seem to be worsened, but in particular way: -Inhabitants of kreis 4, kreis 5, kreis1, kreis 2, kreis 6 -Trips directed to home and shopping-leisure activities	No real winners apart from inhabitants of kreis 7 and kreis 8

Figure 6 19: Overview of winners and losers

The most disadvantaged users by the Flat Toll correspond to agents living in the central districts Kreis 4 and Kreis 5. A possible reason behind that might lie in the fact that part of the inhabitants from these districts are likely charged as they live inside the tolled area in combination with the increase of traffic due to the diversion of part of the trips from the cordon (rat running phenomenon). People travelling to work are also slightly worsened by the introduction of the Flat Toll since it affects mainly people travelling during the morning peak hours.

The Step Toll determines an overall improvement for all the categories of users. The application of a time-varying toll seems to be beneficial especially for residents of Kreis 4, who are affected by a positive increase of score, and Kreis 5, whose negative impact is considerably reduced. Also the trips directed to work are characterized by a significant increase of gain. All these improvements might be explained by the flexibility of a time-varying scheme such as the Step Toll that reduces the negative impacts of the fares without compromising its beneficial effects from traffic performance perspective.

The Spread Toll has a slight negative impact on all the identified socio-demographic categories. Particularly, people living in Kreis 4 and Kreis 5 (districts directly affected by the toll) are characterized by the highest loss. Apparently, the improvement of traffic conditions for people living in this area is not enough to counterbalance the negative effect of the charge. Also people

living in the proximity of the cordon like Kreis 2, Kreis 6 are worsened by the introduction of the Spread Toll.

## **6.8 Public acceptability of tolling schemes**

Identifying the acceptability of a complex policy measure such as a congestion pricing is a difficult task given the various impacts, the spin-off effects and the several stakeholders involved (road users, road and public transport authorities, policy-makers). Investigating the response of people to the alternative tolling schemes goes beyond the scope of this research. However, with the support of the previous analyses and some additional ones of the amount of winners and gains, it is possible to make some preliminary objective considerations on the achievable levels of acceptability.

In this study, we will adopt the conception of acceptability proposed by Mayeres and Proost (2002), according to which the utility change reflects the potential acceptability of a reform. If the utility increases for an individual, he will probably accept the policy measure. Then, a pricing reform will gain the necessary public support only if it shows clear welfare gains for the majority of the population. The larger number of winners is achieved, the higher probability the pricing proposal will be approved in hypothetical elections. In this evaluation, the extent of the gain and losses will also be considered, since a considerable loss of a minority might represent a potential obstacle to the application of the scheme. Similarly, modest gains for a large part of the population might result indifferent to people.

In order to allow some preliminary considerations about the potential acceptability (subsection 6.8.4) of the pricing schemes, a deeper investigation of their effects in terms of amount winners and losers, their average gain and loss, and the total net gains will be performed. The same geographic division in districts used before has been employed again not only to confront more accurately with this issue, but also to provide an approach closer to real policy-making cases. As sections 2.1 and 2.2 have highlighted that even relatively small groups of opponents might be able to stop the introduction of congestion pricing schemes.

### **6.8.1 Overall efficiency of tolling schemes**

As it is possible to see from Figure 6.20 representing the percentage of people gaining from the congestion pricing schemes, the average amount of “winners” varies between 54% and 63% across the districts according to the different scheme. Such a percentage is estimated as a net value in terms of utility before any revenue distribution scheme is applied. The districts with lower percentages of winners are those situated totally and partly inside the cordon for all the schemes (Kreis 1, Kreis 4, Kreis 5). The Step Toll produces a slightly higher amount (one or two additional points) for each district, with the exception of Kreis 8 where the percentage of

winners decreases of about 2%. It is interesting to see that the Spread Toll, although its lowest benefits described in the previous sections (Section 6.2 and 6.6), is characterized by higher levels of winners in every district, particularly in Kreis 4 and Kreis 5 where the increase exceeds 5%.

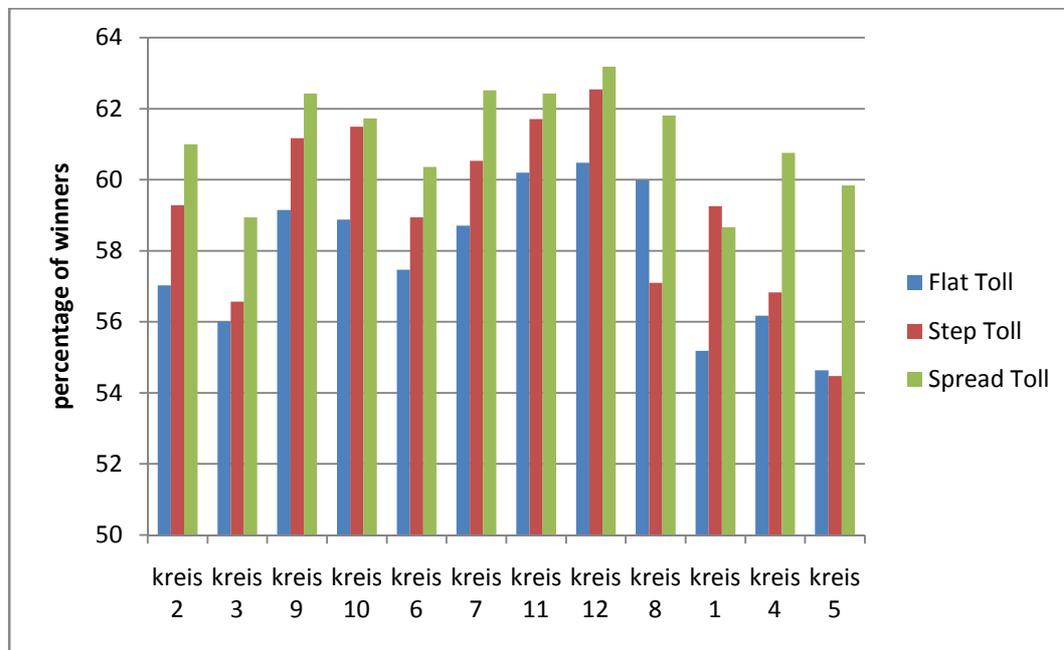


Figure 6.20: amount of winners (with positive utility change) determined by the different schemes

Conversely, the average gain and average loss deriving from the tolling schemes depict the situation from a different angle (Appendix V). Indeed, the Spread Toll is not only characterized by the lowest average gain, but also the highest average loss. This aspect implies that such scheme, although it positively affects more users than the other ones, it produces lower gains. On the other hand, those who are worsened, they are subject to a more severe damage. The Flat Toll and the Step Toll determine very similar results in terms of average gains and losses, even though the latter yields to higher gains for Kreis 4 and Kreis 5 (this is in line with the previous section). It is interesting to see also that Kreis 4 and Kreis 5 are characterized by lower gains and losses in all circumstances. This finding suggests that the effects of the toll in these districts are milder regardless of the typology of schemes.

The cost-benefit ratio of each district reported in Table 6.26 provide a further indication of the potential acceptability of the schemes. From this perspective, the Flat Toll and Step Toll result efficient for the majority of districts except for Kreis 4 and Kreis 5. On the contrary the Spread Toll performs rather bad in every neighborhood. In this study the indicator is estimated as a ratio between the total benefits expressed by utility gains and the total costs expressed by the

sum of utility losses and revenues. If the benefit-cost ratio exceeds one, that the measure is overall beneficial and in theory a good candidate for acceptance.

	k2	k3	k9	k10	k6	k7	k11	k12	k8	k1	k4	k5
<b>flat toll</b>	1.2	1.0	1.1	1.2	1.1	0.9	1.0	1.2	1.2	1.1	0.7	0.6
<b>step toll</b>	1.3	1.1	1.3	1.3	1.0	1.2	0.9	1.3	1.0	1.2	0.9	0.7
<b>spread toll</b>	0.8	0.7	0.8	0.7	0.7	0.8	0.7	0.8	1.0	0.8	0.6	0.5

Table 6 27: benefit-costs ratio for different districts

### 6.8.2 Redistribution schemes

An estimation of the total amounts of losses determined by the schemes allow some further reasoning on the possibility to reach higher levels of (potential) public acceptability by means of refunds from the revenues. Indeed, if total revenues are larger than the total losses, in theory, compensation would be always possible and a situation of complete acceptability could be potentially achieved.

The amount of total losses for the Flat Toll, Step Toll and Spread Toll corresponds respectively to -66.407 CHF, -62.618 CHF and -76.135 CHF. The amount of total revenues for the Flat Toll, Step Toll and Spread Toll corresponds respectively to 18.111 CHF, 14.196 CHF and 2.473 CHF. Clearly, the earnings are not sufficient to refund all the losers and reach a situation of total gain.

The application of two basic refund schemes: a lump sum to everyone and a lump sum to agents living in the districts inside the cordon show how higher levels of acceptance can be reached when charges are paid back to users. For example, Figure 6.21 shows that, when revenues are equally redistributed to everyone, the Step Toll and the Flat Toll can become progressive for about 70% of the agents in all the districts. It is worth mentioning that the Spread Toll, when revenue rebate schemes are applied, is not affected by considerable improvements, as the collected amount is rather low.

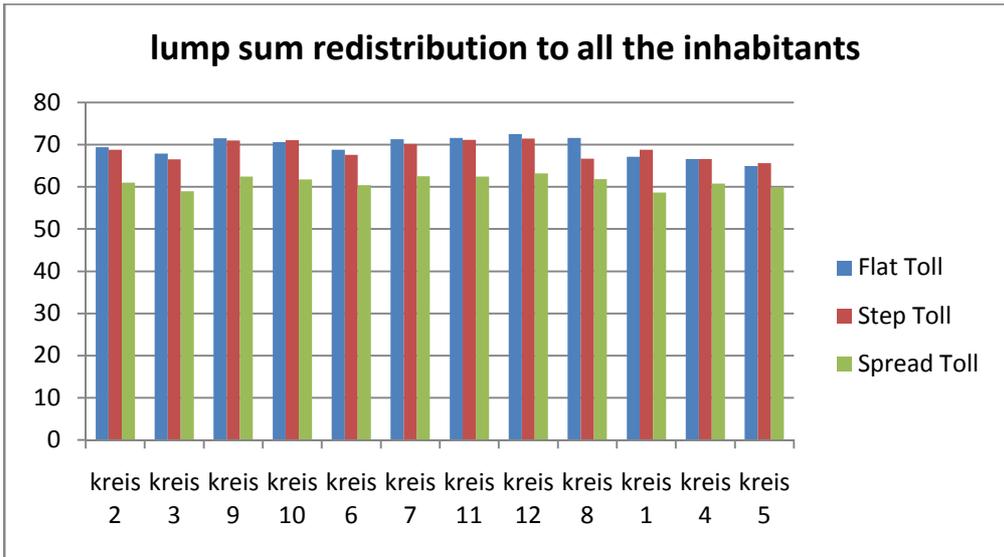


Figure 6 21: amount of winners in the different districts after a lump-sum redistribution to all of them

A possible alternative refunding schemes might consist of a lump sum redistribution to all the residents inside the cordon (Kreis 4, Kreis 5, Kreis 1). The main rationale behind this strategy (similar to London Congestion Charge discounts) is to pay back those people who will be more likely affected by the toll. Then, as Figure 6.22 shows, that share of winners in Kreis 1, Kreis 4 and Kreis 5 would reach almost 90% in case of Step Toll and Flat Toll.

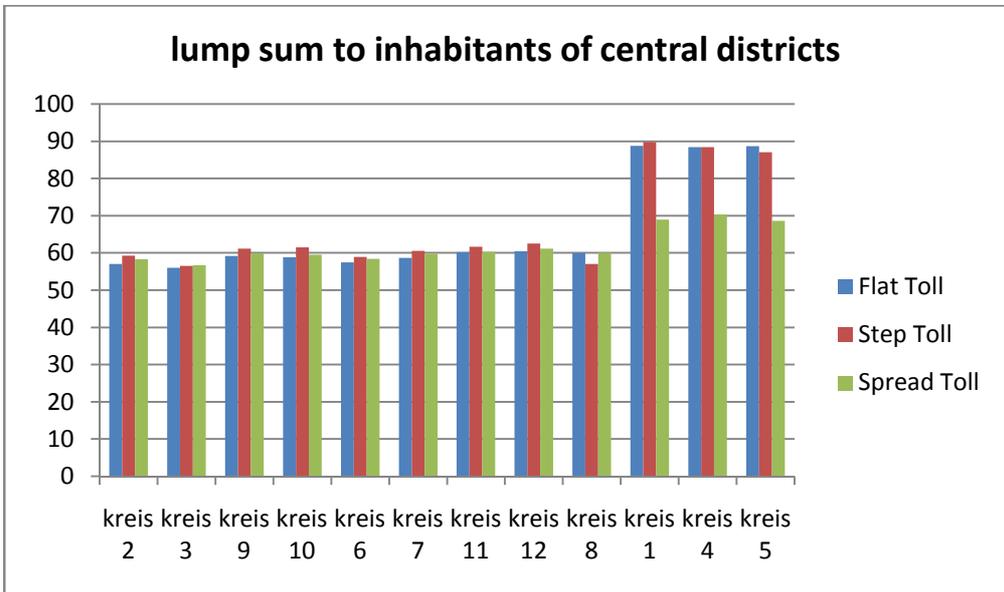


Figure 6 22: amount of winners in the different districts after a lump-sum redistribution to inhabitants of central districts

Another possible solution in terms of rebate scheme may consist of a reduction of transit fares. This strategy would eventually give an additional boost to the modal shift and create a win-win

situation in combination with the tolling schemes<sup>12</sup>. In order to evaluate the levels of acceptability achievable with this approach, a lump sum is redistributed to pt users travelling to the cordon. As it possible to see from Figure 6.23, winners among pt users (directed to the tolled area), who are already the majority without any refund, would reach considerable levels (around 70%) for all the schemes. Like in the previous cases, the largest improvements are achieved by the Flat Toll and the Step Toll (+10%).

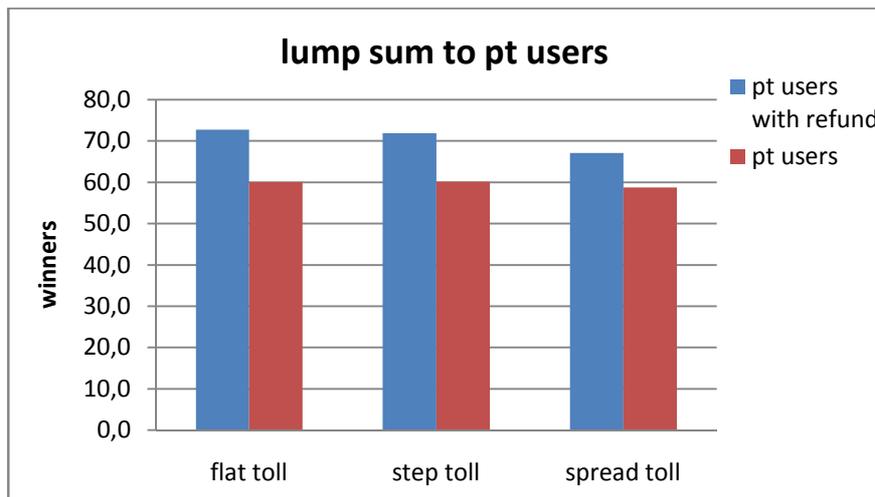


Figure 6 23: amount of winners among the pt users before and after a lump sum redistribution to pt users

### 6.8.3 Final considerations about the public acceptability issue

The findings from the analyses performed in this section provide an additional insight of the potential public acceptability of the alternative tolling schemes and they allow some preliminary considerations from the policy-making perspective. The geographical focus of the investigations on the economic impacts has been adopted in order to identify the most problematic districts and potential opponents already during the design stage of the schemes. This approach might help to find the most suitable solutions to achieve higher levels of acceptability.

The previous analyses have highlighted that all the schemes are overall beneficial for the majority of inhabitants (about 60%) of the twelve districts of Zurich even without any rebate scheme is implemented. Under these circumstances, the Spread Toll achieves the largest share of winners, but on the other hand, its deriving benefits are on average lower than the other two schemes, and the losses higher. This issue raises the crucial strategic question whether it is preferable to implement a pricing scheme that determines profits to a higher number of users or that produces greater benefits to fewer. However, in this specific case study, the considerable

<sup>12</sup> A similar approach has been adopted in Stockholm where part of the revenues has been invested in enhancement of public transport services

difference between total and average benefits of the different schemes seems to be the most important factor to be considered in the decision.

This analysis has also stressed the importance of the application of revenue redistribution schemes. Indeed, when a basic rebate is paid back as a lump sum to users, the Flat Toll and the Step Toll present significant improvements and seem to be more progressive than the Spread Toll. As concern potential seeds of resistance, the analyses suggest the most problematic districts could be Kreis 4 and Kreis 5, as they are characterized by the lowest percentages of winners and lowest average gains. However, the experimental implementation of a rebate specifically targeted to these neighborhoods, has demonstrated that with proper compensation it is possible to achieve very high shares of winners (up to 90%). The possibility to refund (subsidize) pt users (directed to the centre) by means of reduced fares seems to determine positive results for all the schemes. Then, it also seems a valuable option to increase public acceptance and incentivize public transport usage at the same time. Alternative rebate schemes directed to other specific categories like car users and different incomes might have been applied, however such investigation goes beyond the scope of this research.

Although several additional factors such as the complexity of the scheme, the time factor, the political context, and the commitment of political parties, affect public acceptance of tolling schemes, these preliminary analyses form an objective basis for broader and deeper discussion.

## 7. Conclusions

In this study of congestion pricing different schemes based on the macroscopic characteristics of traffic have been developed within an activity-based transport model. Alternative tolls have been implemented in order to combine fundamental features from the classic economic theory, such as the marginal cost pricing, with some of the most recent findings in traffic flow theory like the (g)MFD. Then, a comprehensive appraisal including several engineering and economic aspects has been carried out in order to identify the strengths and weaknesses of these schemes from a social welfare perspective intended as a combination of efficiency and equity. Finally, in light of the results, some preliminary considerations of their potential acceptability have been made with the support of some additional analyses.

The previous chapters (Chapter 2, 3 and 4) have described the major issues and theoretical challenges related to the theoretical formulation and operational application of congestion pricing models. Above all, we brought to light the necessity of dynamic models consistent with the physics of traffic and the importance of public acceptability, which can be achieved by means of high efficiency and a “fair” distribution of gains and losses. The design of new tolling schemes (Chapter 5) has offered the opportunity to gain a further insight into the field of macroscopic modeling of traffic and shed light on a series of macroscopic phenomena and characteristics of congestion. The implementation of the schemes controlled by the (g)MFD within a multi-agent transport model (MATSim) represents an important achievement, given the relative originality of the approach. The developed framework for the evaluation (Chapter 6) has not only allowed a thorough comparison of the schemes, but also enhanced the current knowledge about the efficiency and the distributional impacts of this typology of schemes. A similar methodology might be applied by policy-makers as a starting point for a discussion about a possible strategy to achieve public acceptance of congestion pricing schemes.

In this final chapter, the main findings deriving from this study will be presented in Section 7.1. The implications deriving from these ones together with some specific issues of the traffic flow theory arisen during the design of tolling schemes will be discussed in the conclusions of Section 7.2. The main practical implications of this study, especially concerning the adopted pricing model and appraisal framework will be made in Section 7.3. Finally, Section 7.4 will give recommendations about future research to improve the current study and explore related topics.

### 7.1 Main findings

This study has brought to light several facts that range from the traffic flow theory to the transport economics discipline. The investigations of macroscopic traffic characteristics of the

network have shed light on important issues of the traffic flow theory (Subsection 7.1.1). The answers of the research questions have led to findings mainly related to the questions of efficiency, distributional effects and public acceptance of the proposed schemes (Subsection 7.1.2).

### **7.1.1 Results from the traffic flow theory perspective**

The investigations of the property of spatial distribution of congestion in the network, expressed by the spatial spread of density, have revealed an additional complexity of the nature of the aggregated traffic relationship. This characteristic, which can be modeled as additional dimension in a generalized macroscopic fundamental diagram (gMFD), plays a major role in the decrease of efficiency of the system. Indeed, increases of spread corresponding to clustering of congestion might determine considerable drops of production, even for moderate levels of demand. As a consequence, the network might exhibit traffic states way below the theoretical upper bound and too scattered to delineate a sharp MFD. These findings are in line with the homogeneity conditions to derive a “well-defined” MFD expressed by Daganzo, Geroliminis and colleagues.

This paper provides additional empirical support of the existence of frequent hysteresis phenomena in the MFD plane. Hysteresis consisting of clockwise loops in the accumulation-production plane shows that flows at the onset of congestion are higher than at the offset. Furthermore, hysteresis itself might be considered as a reason of decrease of performance of the system (Step Toll case), since frequent cycles of loading-unloading of the network can deteriorate its production.

### **7.1.2 Research questions**

The answers to the research questions have led to the following main findings.

#### **1. Do tolling schemes differently controlled by the MFD determine comparable improvements from the traffic efficiency perspective? If not, which is the most and least efficient?**

From a traffic performance perspective all the schemes produce appreciable improvements of traffic conditions. The Flat Toll and Step Toll, which have similar performance, seem to be overall more beneficial than the Spread Toll. The main reason behind that can be found in the higher fares that determine an higher reduction of demand. A detailed analysis of the improvements expressed by several performance indicators has shown how these schemes might not be efficient in addressing congestion issues when the distribution of traffic (inside the cordon) is not even. Particularly, the impacts are different during the day: improvements in terms of decrease of delay and heaviness of congestion during evening peak are equal or even

lower than in the morning peak, despite the higher reduction of traffic. The main reason for that probably lies in the different nature of decrease of performance: while in the morning the drops occurs because of a considerable increase of accumulation, during the evening the spread of congestion plays a crucial role.

## **2. Are tolling schemes differently controlled by the MFD characterized by similar trends in travel behavior changes?**

Both the Flat Toll and the Step Toll determine a decrease of demand comparable to real experiences of congestion pricing (Stockholm, London) in terms of trips crossing the cordon during the whole day. It is worth mentioning that the Flat Toll generates an higher reduction of overall demand, while the Step Toll determines higher rescheduling of trips. As expected, the Spread Toll does not determine relevant changes because of the lower fare. People living outside the cordon are more inclined to reroute whereas people living inside are more likely to switch to PT. Probably because of the nature of the tolling schemes that operates during the typical periods of commuting, the activities that appear most affected are work and home.

## **3. Do tolling schemes differently controlled by the MFD determine comparable improvements from the economic welfare perspective? If not, which is the most/least efficient?**

As to the economic effects deriving from different schemes, the analyses have shown that alternative tolls produce different economic impacts. The economic appraisal based on the utility allowed a comprehensive evaluation of the effects of congestion pricing including the higher travel costs, changed travel behavior and different travel times. The Step Toll performs better than the other schemes probably thanks to its larger flexibility as the investigations of Sections 6.5 and 6.6 imply. Especially the possibility to regulate traffic through the rescheduling of trips rather than by rerouting or changing mode seems to be the main benefit of this scheme. The Spread Toll is characterized by an average decrease of utility and low benefit-cost ratio. The fact that the standard deviation of utility change is rather high in all the schemes implies that some agents might substantially benefit or lose from the schemes. However, by means of statistical tests it was possible to demonstrate that the difference among average changes of utility is statistically significant for all the schemes. These results suggest that, although the Spread Toll applies a low fare, this is not enough to determine substantial benefits. On the contrary, the other two schemes more than compensate the costs of higher charges thanks to the appreciable improvement of traffic conditions.

**4. Are tolling schemes differently controlled by the MFD characterized by similar distributional effects? If not, is there any socio-demographic economic category that results particularly advantaged or disadvantaged by a specific tolling scheme?**

This study has shown that different distributional effects arise in accordance with varying implementation schemes. From a spatial equity perspective some districts result more advantaged than others according to the typology of toll. The Flat Toll and Step Toll have similar trends, although the latter produces higher benefits for residents of those districts inside the cordon. The Spread Toll does not determine significant differences across the neighborhoods with the exception of those inside the cordon that are particularly worsened. Also from a vertical equity perspective, schemes result regressive for certain categories of trip and progressive for other ones. The Flat Toll is in particular disadvantageous for trips directed to working activities whereas the Step Toll determines a significant improvement for them. The Spread Toll seems to be rather regressive for all the categories and it worsens all of them to a similar extent.

**5. Do the different schemes achieve the same level of potential acceptability?**

The potential acceptability of tolling schemes has been estimated objectively by looking at the share of winners, the average gains and losses and the benefit-cost ratio in the twelve districts of Zurich. When considering the amount of winners, all the schemes resulted progressive with similar results across the neighborhoods. From this perspective, the Spread Toll achieves a slightly higher share than the other schemes (up to five points of percentage in the central districts). Vice versa, the Flat Toll and Step Toll are characterized by higher average gains and lower average losses. This situation raises the crucial question whether it is preferable to benefit a larger portion of users with lower gains or produce greater benefits to fewer. In light of the previous discussions (Section 2) and the results of Section 6, the Step Toll and the Flat Toll seem the most convenient as they determine considerably higher benefits than the Spread Toll in the face of a slightly lower share of winners. Furthermore, the implementation of three basic revenue redistribution strategies has highlighted how rebates might increase the benefits and the levels of potential acceptance of the Flat Toll and Step Toll. Then, if revenues are considered as a resource, these stricter schemes result more beneficial than the more moderate Spread Toll.

**How do cordon-based pricing schemes governed by the macroscopic fundamental diagram address social welfare?**

Social welfare in this study has considered as a combination of efficiency and equity, where the latter is intended as the (balanced) distribution of gains and losses across different socio-demographic groups. The implementation of tolling schemes differently controlled by the gMFD has brought to light differences in the way social welfare is addressed.

As to the traffic and economic efficiency, the Step Toll and the Flat Toll are overall beneficial, whereas the Spread Toll determines very small traffic improvements and has a general negative economic impact on the population. The Step Toll is in the best performing schemes probably due to its flexibility. On the other hand, the investigations of distributional effects highlight that the Spread Toll is “fairer” than the other schemes since it does not advantage any specific category of users (but it equally worsens everyone). This is somehow an expected outcome, given the lower levels of fares. A comparison between the Step Toll and Flat Toll has shown that the first one produces “better” results in terms of distributional impacts across the socio-demographic groups (almost no group is disadvantaged).

Hence, all in all the time-varying toll that penalizes users differently according to the performance of the network (Step Toll) is more beneficial from a social welfare perspective than an uniform toll (Flat Toll) that maintains the total demand below a critical threshold for the entire peak. The “milder” version of the time-varying toll (Spread Toll) that applies lower fares accounting for the factor of spatial distribution of traffic misses the social welfare objective because of its poor efficiency.

## **7.2 Conclusions**

The implementation of tolling schemes based on macroscopic traffic relationships and the broad evaluation based on the agent-based modeling framework allow some preliminary general considerations that might be confirmed by additional research studies. The main implications deriving from the analyses and design of tolls controlled by the gMFD are discussed in Subsection 7.2.1. General conclusions from the evaluation of the schemes are drawn in Subsection 7.2.2. Finally, Subsection 7.2.3 makes some considerations about the consistency of the implemented model with traditional congestion pricing models.

### **7.2.1 Design of schemes based on the gMFD**

The design of tolling schemes based on aggregated traffic relationships has offered the opportunity to gain a deeper insight into phenomena related to the spatio-temporal congestion patterns in large urban networks. The most important one consists in the spatial spread of density that determines the efficiency of cordon-based schemes controlled by the gMFD. The results of the analyses have pointed out the limitations of traditional cordon-based pricing schemes in addressing complexities such as the clustering of congestion. Indeed, tolls based on

charging drivers entering a certain area are able to limit the demand, but not to affect the route choice (inside the cordon). Factors related to the urban morphology like the road network and land use of a city might affect the distribution of traffic and ultimately the effectiveness of this demand management measure. For example, urban areas characterized by a very regular structure like US cities, thanks to the redundancy of their network might “suffer” less from clustering of congestion than very hierarchical-structured cities like the European ones. On the other hand, concentration of functions (widely adopted practice in American cities) does not favor balanced distributions (in space and time) of travel patterns. The influence of fluctuations of demand that are responsible for hysteresis loops and decreases of performance suggests that a smooth control of traffic demand might be more beneficial from a traffic flow perspective.

### **7.2.2 Efficiency, distributional impacts and acceptability of different schemes**

Results in terms of economic impacts of the schemes imply that time-varying tolls might be more beneficial than uniform tolls thanks to their flexibility. Indeed, the Step Toll is a closer approximation to first-best congestion pricing. This finding might have useful implications from an operational point of view as it demonstrates that a smooth control of traffic demand is a valuable alternative to more invasive strategies. Although the levels of demand and fares would differ according to the specific studies, it is reasonable to think that the mechanism and the effects will be similar in other case studies. The failure of the Spread Toll also from an economic perspective shows a consistency between traffic management and economic approach.

Interestingly, results in terms of distributional effects of tolls seem to confirm the dualism of efficiency and equity in congestion pricing. Indeed, the Spread Toll that has been deliberately designed to be “fairer” is in the end more detrimental for everyone. On the other hand, the Step Toll is not only more beneficial, but it also determines fewer inequities than the Flat Toll. Again, the main reason for that is probably its flexibility. Anyway, it is difficult to generalize the results because of crucial factors like the distribution of traffic and several complexities behind travel patterns in large metropolitan areas.

Similarly, it is hard to derive general conclusions in terms of advantaged and disadvantaged groups given the unique conditions of each city. Nevertheless, the fact that congestion pricing schemes (usually) operate in a limited central area of the city and during specific times of day suggests that people commuting from home to work and vice versa, and residents of districts inside the cordon might be the most vulnerable groups. Hence, characteristics of the urban areas affected by cordon-based tolls could influence the intensity of the distributional effects. For example, if the cordon circumscribes a relatively small area, people living inside might be forced to cross it to reach most of their everyday activities. On the contrary, cordon-based tolls

applied to very large and dense of economic activities areas might harm commuters from the outer neighborhoods.

As to the question of public acceptance, this study shows how the share of winners and the extent of gains and losses deriving from the schemes might determine conflicting views. The scheme that produces the higher share of winners is not necessarily the one that determines the highest gains (and lowest costs). Hence, deciding which scheme has higher chances to be potentially accepted is not straightforward even when it is based on objective results.

### **7.2.3 Consistency with traditional congestion pricing models**

Making a comparison between the results derived from the pricing model based on (g)MFD and those achievable with the classic bottleneck model is not trivial task given the considerable differences between them. First, the bottleneck model is characterized by inelastic demand and typically only departure time choice is allowed. In our model demand is sensitive to price and travel behavior includes mode and route choices. Second, in the bottleneck model trip-preferences are exogenous (normally they all have the same desired arrival time) whereas in our model scheduling preferences are the result of a complex equilibrium of several activities (home, work, etc.) and they all differ from each other. Third, in our model, the spatial dimension plays a key role in the design of charges and in the efficiency of the schemes, particularly in the Spread Toll. The original bottleneck model, instead, considers a single road or it has been extended to networks by making certain assumptions (Subsection 3.3.3). Even in the most recent approaches presented in Section 3.5 no explicit consideration on the spatial distribution of congestion inside the cordon has been made.

Ideally, in the bottleneck model a continuously varying toll (fine toll) would replace completely queuing costs by revenues (first-best pricing). In this case, total travel costs of each user are unvaried, since travelers reschedule their departures to minimize their costs. A uniform toll, constant throughout the peak, would likely affect the total demand rather the time departure. By increasing the price, the toll lowers the amount of users and consumer surplus and it becomes thus, comparable to the static MCP, characterized by much lower gains than the first-best bottleneck toll (Van den Berg, 2012). The Step Toll is in between the fine toll and uniform toll, since it concurrently raises the price and changes the time departure pattern. The higher the number of steps, the closer the schemes get to first-best pricing and the higher the benefits.

The results in terms of fares and welfare gains of the Flat Toll and Step Toll seem to be consistent with similar studies performed by means of the bottleneck model. Both of them are able to decrease the costs of users determined by the waste of congestion (hypercongestion). Similarly to other studies based on the bottleneck model (Arnott *et al.*, 1993; Van den Berg,

2012) and on queue-based models (de Palma *et al.*, 2005), the time varying toll results more beneficial than the uniform toll. The analysis of travel behavior changes confirmed that the Step Toll determines a higher rescheduling of trips in place of a lower decrease of demand compared to the Flat Toll. As concern the Spread Toll, it is not possible to check its consistency with the bottleneck model, since the issue of uneven distribution of traffic is not really considered. It seems very hard to include this aspect in the bottleneck model by adopting different capacities according to the level of spatial spread.

### **7.3 Implications for practice**

In addition to the aforementioned conclusions, this study has led to a series of speculations of practical relevance. Some implications are more related to the direct findings of the experiments like those concerning the effectiveness of cordon-based tolls (Subsection 7.3.1), whereas other ones derive from the adopted methodology (Subsection 7.3.2 and 7.3.3).

#### **7.3.1 Effectiveness of cordon-based tolls and combination with traffic management**

The possibility of modeling congestion at macroscopic level has proved to be not only a functional approach to measure the state of the system, but also an useful means to investigate its fundamental characteristics and control them. Understanding and considering the phenomenon of spatial spread of density becomes a priority when traffic and mobility management policies are implemented. When strategies aimed at spreading users more evenly over the network come into play, traffic management measures like gating, traffic signal control, variable-message signs and GPS might achieve larger improvements. If the pricing solution is pursued anyway, then an alternative tolling scheme based on differentiated pricing might be a valuable alternative solution. In this case, the network could be divided in two classes of roads: the “primary” ones, characterized by high performance and with the highest priority; and the “secondary” ones, characterized by lower performances (but, perhaps higher critical accumulation) that could function as “storage”. Fares could be set in order to regulate the optimal usage of the two sub-networks. Ideally, the pricing mechanism could be used in combination with traffic management measures aimed at controlling the inflow in vulnerable locations of the network to avoid triggering of congestion. For example, gating systems, whose main goal is to guarantee fluency of traffic in targeted areas by means of traffic signal control, might be used to improve the performance of network. Therefore, it would be possible to avoid drops of congestion due to the spread and reach the theoretical upper bound of the MFD. At this point, major decreases of production could be simply avoided by regulating the overall demand through the pricing mechanism.

### **7.3.2 Tolling schemes controlled by the gMFD for complex case studies**

The model adopted in this study consisting of tolling schemes controlled by the gMFD within agent-based transport model has shown substantial benefits compared to the classic analytical models. First, it is possible to reach higher levels of realism in terms of heterogeneity of users, sensitivity to the toll and scale of the study. The major advantage of the gMFD approach consists in the reliability and consistency with the traffic engineering perspective. Furthermore, the highly disaggregated nature of the model (single agents) allows a series of detailed analyses of the distributional impacts and public acceptability, not easily achievable with the traditional analytic models. The efficiency of the second-best pricing schemes implemented in this study represents an additional bonus and encouraging result for further research aimed at improving the model and supporting practical applications. As Small and Verhoef (2007) observe, when larger networks are investigated and increasing levels of realism are required, numerical methods should substitute analytical methods. The main downside of these models consists in their lower tractability, since numerical solutions are more difficult to be interpreted. The model proposed in this study represents somehow an attempt to provide an easily interpretable strategy to determine the levels of tolls in complex simulations. Then, when approaching case studies characterized by large networks and high degree of complexity, the proposed model seems to be the more suitable than traditional models. Anyway, the bottleneck model could still be used as a benchmark for preliminary assessments of case studies under specific conditions: traffic mainly directed to the centre; even distribution of traffic inside the cordon; presence of a single peak.

### **7.3.3 A comprehensive appraisal framework as a tool for policy-makers**

Nowadays, considering the full range of issues deriving from congestion pricing is becoming of crucial importance already during the design phase. Evaluations focused on the optimal usage of road capacity might overlook important effects like costs and benefits of individual travel behavior changes. Traditional economic assessments, on the other hand, might neglect important aspects regarding the generation and propagation of congestion. The agent-based modeling framework proposed in this study allows the appraisal of alternative tolling schemes from a broad perspective. Indeed, the possibility to simulate trip chains over the day in large urban scenarios with several thousand agents with decent reliability in terms of traffic outputs offers the opportunity to tie the engineering and economic approaches towards congestion pricing more closely. The evaluation of benefits and costs at highly disaggregated level allows a series of speculations about the distributional effects and the potential levels of acceptability of the schemes. The consideration of these issues at an early stage could make the difference in finally determining the success or failure of this policy measure since the most critical categories of users could be identified for each tolling scheme. Furthermore, the comparison of

individual utility provides a basic, but clear indication of the potentiality of the scheme from the public acceptability perspective. All these aspects together represent a good starting point for further policy considerations on the redistribution of revenues and the introduction of discounts for specific groups.

## **7.4 Implications for research**

During the research, several issues and questions emerged, so that further research would be recommended. The current work can be improved along different directions like the agent-based model, the design of tolls and the evaluation framework in order to produce more reliable and complete results (Subsection 7.5.1). Furthermore, given the number of issues arisen during the experiments, a series of topics of research has been identified for future research in different fields (Subsection 7.5.2).

### **7.4.1 Recommendations for improved research**

Additional complexity could be introduced in the agent-based model used for the simulation, by including additional heterogeneity of users in terms of incomes in order to investigate more in depth the distributional impacts of the schemes. Especially, as mentioned in Chapter 2, income of people and ultimately their value-of-time can play an important role in the way people evaluate and react to congestion pricing schemes. In theory, this aspect could be introduced by assigning a different marginal utility for each work activity based on the income or by developing income-dependent (dis)utilities of travelling like Kickhofer *et al.* (2010).

The design of tolling schemes might be improved by means of a more dynamic estimation-implementation process to derive real time tolling schemes within the same simulation. This approach would probably enhance higher efficiency of the time-varying schemes, as they would capture the variability of traffic conditions better than the flat ones, which are conceptually closer to fixed averages. Such a responsive strategy, similar to traffic management strategies like the ramp metering, could operate better in case of non-recurrent congestion phenomena (e.g. accidents, adverse weather conditions). However, this approach would entail a shift towards the issue of advanced information as travel behavior choices like the departure time and mode choice are pre-trip decisions based on expected rather than actual toll and queue patterns (Small and Verhoef, 2007). As a consequence, instantaneous toll adjustments would probably affect only the route choice and frustrate drivers that might face sudden increases of charge and, or unexpected queues.

The appraisal framework could be improved by including the decrease of variability of travel time as an indicator of the efficiency of the schemes. As several scholars highlight (OECD, 2007;

Fosgerau and van Dender, 2010), variability and predictability of travel times are highly valued by road users who base their travel decisions based on this factor too. Since congestion pricing schemes might provide additional benefits by increasing the reliability of travel times in the network, it would be a good practice to include this aspect in the appraisal. However, dealing with travel time distributions is no trivial task in agent-based models like MATSim that simulate only one day of traffic.

As to the feasibility of the proposed tolling schemes, some practical issues about the collection and processing of data, and estimation of the toll need to be overcome. The main barrier consists in the collection of the data necessary to build a (g)MFD, since monitoring resources are often scarce and many cities do not have access to the large amount of data required to build it. A recent study by Ortigosa *et al.* (2013) has investigated the possibility to derive an accurate MFD from a limited fraction of links of the network and the results seem to be encouraging.

#### **7.4.2 Recommendations for future research directions**

Several topics of interest that could be addressed by different research areas stemmed from this study as well. For example, traffic engineers might be interested in pursuing deeper and broader investigations of the effects of the spatial distribution of congestion and hysteresis loops on traffic performance. The concept of (g)MFD can be further developed and improved in order to make it more theoretically sound and exploit its full potentialities. Finally, as mentioned in Section 7.3, it might be interesting to relate these issues with the efficiency of different traffic management and control solutions.

From a policy-making and economic perspective, it would be interesting to use a similar agent-based approach to research the distributional effects and acceptability of congestion pricing by including investigations of different revenue redistribution schemes. For example, discounts on the car-ownership taxes, infrastructure investments and improvement of public transport services could be analyzed to seek to most equitable and attractive solution. This approach, which is characterized by a long-term view, might also consider aspects like long-term travel behavior changes (decision of making a trip) or the relocation of commercial and residential activities.

Finally, another idea of research that combines the interests of different areas (traffic flow theory, transport economics and transport policy) would be the development of policies based on the (g)MFD that include parking and public transport pricing. Indeed, it has been demonstrated that parking represents a significant portion of travel costs in terms of money and time. Furthermore, people looking for a parking spot may slow down other users and

ultimately decrease the traffic performance of the network. Then, it would be useful to identify the effects of the parking research process and define its marginal cost in the disutility of travelling. Then, different pricing policies controlled by the (g)MFD might be adopted to improve the performance of the system. For example, tariffs could vary in time depending on the availability and traffic conditions of each area of the network. On the same page, the performance of public transport services could be modeled by means of aggregated traffic relations and the fares could be set to achieve the optimal usage of the network. A similar approach has been adopted by Zheng and Geroliminis (2013) who extended the single mode MFD to multimodal cases and combined it with tolls in order to derive the optimal allocation of road space between car and bus lanes.

## References

- Adler, Jeffrey L., and Mecit Cetin. 2001. A direct redistribution model of congestion pricing. *Transportation Research Part B: Methodological* 35 (5): 447-60.
- Anas, Alex, and Robin Lindsey. 2011. Reducing urban road transportation externalities: Road pricing in theory and in practice. *Review of Environmental Economics and Policy* 5 (1): 66-88.
- Arnott, Richard. 2011. A Bathtub Model of Traffic Congestion.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1990. Departure time and route choice for the morning commute. *Transportation Research Part B: Methodological* 24 (3): 209-228.
- Arnott, Richard, Andre De Palma, and Robin Lindsey. 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*: 161-79.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1994. The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy*: 139-61.
- Balmer, M., K. Meister, M. Rieser, K. Nagel, and Kay W. Axhausen. 2008. Agent-based simulation of travel demand: Structure and computational performance of MATSim-TETH, Eidgenössische Technische Hochschule Zürich, IVT Institut für Verkehrsplanung und Transportsysteme.
- Brilon, Werner, Justin Geistefeldt, and Matthias Regler. 2005. Reliability of freeway traffic flow: a stochastic concept of capacity. *Proceedings of the 16th International symposium on transportation and traffic theory*. Vol. 125143.
- Buisson, Christine, and Cyril Ladier. 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board* 2124 (1): 127-36.
- Chakirov, Artem, Alexander Erath, Alexander Erath, and Alexander Erath. 2012. Overcoming challenges in road pricing design with an agent-based transport simulation Eidgenössische Technische Hochschule Zürich, IVT, Institute for Transport Planning and Systems.
- Daganzo, Carlos F. 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41 (1): 49-62.
- Daganzo, Carlos F., Vikash V. Gayah, and Eric J. Gonzales. 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B: Methodological* 45 (1): 278-88.
- Daganzo, Carlos F., and Nikolas Geroliminis. 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9): 771-81.
- Daganzo, Carlos F. 1985. The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck." *Transportation science* 19 (1): 29-37.

- . 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9): 771-81.
- De Palma, André, and Mogens Fosgerau. 2010. Dynamic and static congestion models: A review. ————. 2010. Dynamic and static congestion models: A review.
- De Palma, André, Moez Kilani, and Robin Lindsey. 2005. Congestion pricing on a road network: A study using the dynamic equilibrium simulator METROPOLIS. *Transportation Research Part A: Policy and Practice* 39 (7): 588-611.
- de Palma, André, and Robin Lindsey. 2011. Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies* 19 (6): 1377-99.
- De Palma, André, and Robin Lindsey. 2004. Congestion pricing with heterogeneous travelers: A general-equilibrium welfare analysis. *Networks and Spatial Economics* 4 (2): 135-60.
- Ecola, Liisa, and Thomas Light. 2009. Equity and congestion pricing. Rand Corporation: 1-45.
- Eliasson, J. 2010. Essential insights for decision-makers considering congestion pricing. Paper presented at European Transport Conference, 2010, .
- Eliasson, Jonas, and Lina Jonsson. 2011. The unexpected “yes”: Explanatory factors behind the positive attitudes to congestion charges in stockholm. *Transport Policy* 18 (4): 636-47.
- Eliasson, Jonas, and Lars-Göran Mattsson. 2006. Equity effects of congestion pricing: Quantitative methodology and a case study for stockholm. *Transportation Research Part A: Policy and Practice* 40 (7): 602-20.
- . 2006. Equity effects of congestion pricing: Quantitative methodology and a case study for stockholm. *Transportation Research Part A: Policy and Practice* 40 (7): 602-20.
- Fosgerau, Mogens, and André De Palma. 2012. Congestion in a city with a central bottleneck. *Journal of Urban Economics* 71 (3): 269-77.
- Fosgerau, Mogens, and van Dender, Kurt. 2013. Congestion pricing with complications. *Transportation* 40 (3) : 479-503.
- . 2012. Congestion in a city with a central bottleneck. *Journal of Urban Economics* 71 (3): 269-77.
- Franklin, Joel P. 2006. The distributional effects of transportation policies: The case of a bridge toll for Seattle. University of Washington.
- Gaunt, Martin, Tom Rye, and Simon Allen. 2007. Public acceptability of road user charging: The case of edinburgh and the 2005 referendum. *Transport Reviews* 27 (1): 85-102.
- Gayah, Vikash V., and Carlos F. Daganzo. 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research Part B: Methodological* 45 (4): 643-55.

- Geroliminis, Nikolas, and Carlos F. Daganzo. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42 (9): 759-70.
- . 2007. Macroscopic modeling of traffic in cities. Paper presented at Transportation Research Board 86th Annual Meeting, .
- Geroliminis, Nikolas, and David M. Levinson. 2009. Cordon pricing consistent with the physics of overcrowding. In *Transportation and traffic theory 2009: Golden jubilee.*, 219-240Springer.
- Geroliminis, Nikolas, and Jie Sun. 2011. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. *Procedia-Social and Behavioral Sciences* 17 : 213-28.
- . 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological* 45 (3): 605-17.
- Goodwin, Phil B. 1989. The «rule of three»: A possible solution to the political problem of competing objectives for road pricing. *Traffic Engineering & Control* 30 (10): 495-7.
- Grether, Dominik, Benjamin Kickhöfer, and Kai Nagel. 2010. Policy evaluation in multiagent transport simulations. *Transportation Research Record: Journal of the Transportation Research Board* 2175 (1): 10-8.
- Harrington, Winston, Alan J. Krupnick, and Anna Alberini. 2001. Overcoming public aversion to congestion pricing. *Transportation Research Part A: Policy and Practice* 35 (2): 87-105.
- Jaensirisak, S., M. Wardman, and AD May. 2005. Explaining variations in public acceptability of road pricing schemes. *Journal of Transport Economics and Policy*: 127-53.
- Jones, Peter. 1998. Urban road pricing: Public acceptability and barriers to implementation.
- Jones, Peter M. 1995. Road pricing: The public viewpoint. In *Road pricing: Theory, empirical assessment and policy.*, 159-179Springer.
- Kickhöfer, Benjamin, Michael Zilske, and Kai Nagel. 2010. Income dependent economic evaluation and public acceptance of road user charging. Paper presented at Proceedings of the Kuhmo Nectar Conference on Transportation Economics, .
- . 2010. Income dependent economic evaluation and public acceptance of road user pricing. Paper presented at Proceedings of the Kuhmo Nectar Conference on Transport Economics, .
- Knight, Frank H. 1924. Some fallacies in the interpretation of social cost. *The Quarterly Journal of Economics* 38 (4): 582-606.
- Knoop, Victor L., Serge Hoogendoorn, and JWC Van Lint. 2013. Impact of traffic dynamics on macroscopic fundamental diagram. Paper presented at Transportation Research Board 92nd Annual Meeting, .
- Knoop, VL, and SP Hoogendoorn. 2013. Empirics of a generalized macroscopic fundamental diagram for urban freeways. Paper presented at Transportation Research Board 92nd Annual Meeting, .

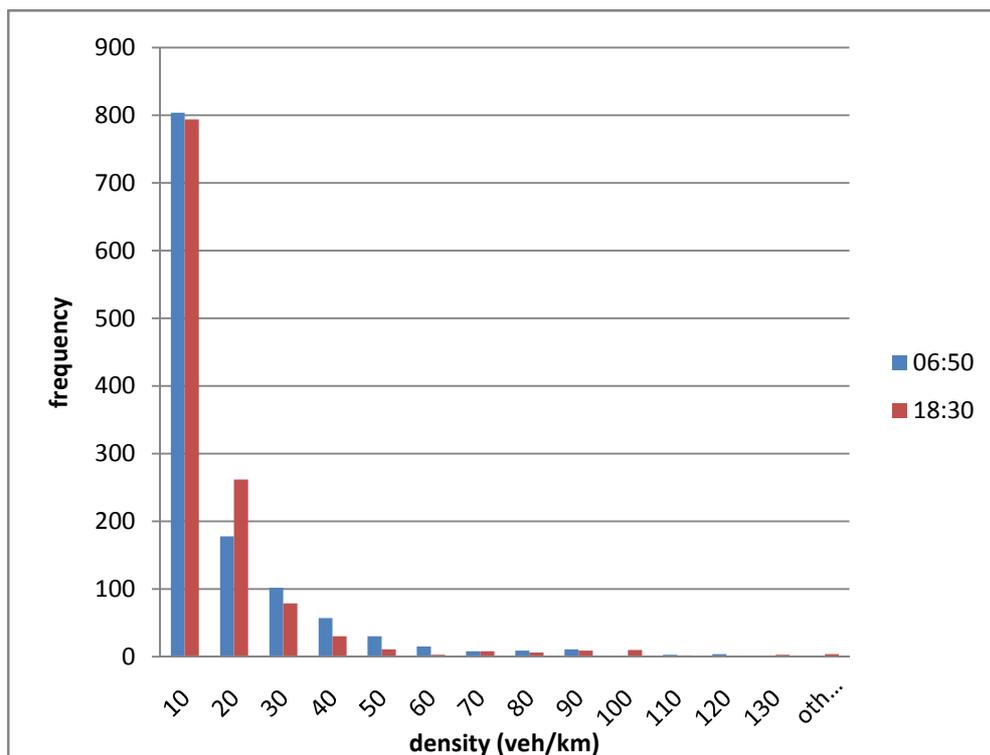
- Laih, Chen-Hsiu. 2004. Effects of the optimal step toll scheme on equilibrium commuter behaviour. *Applied Economics* 36 (1): 59-81.
- Laird, James, Chris Nash, and Simon Shepherd. 2007. Cordon charges and the use of revenue: A case study of Edinburgh. *Research in Transportation Economics* 17 : 161-87.
- Levinson, David. 2010. Equity effects of road pricing: A review. *Transport Reviews* 30 (1): 33-57.
- Litman, Todd. 2003. Measuring transportation: Traffic, mobility, and accessibility. *ITE Journal* 73 (10).
- Mahmassani, Hani S., and Saberi, Meead. 2013. Urban network gridlock: Theory, characteristics, and dynamics. *Procedia-Social and Behavioral Sciences* 80: 79-98.
- Manville, Michael, and David King. 2013. Credible commitment and congestion pricing. *Transportation* 40 (2): 229-49.
- Maruyama, Takuya, and Agachai Sumalee. 2007. Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model. *Transportation Research Part A: Policy and Practice* 41 (7): 655-71.
- May, Anthony D., Rongfang Liu, Simon P. Shepherd, and Agachai Sumalee. 2002. The impact of cordon design on the performance of road pricing schemes. *Transport Policy* 9 (3): 209-20.
- Mayeres, Inge, and Stef Proost. 2002. Reforming transport pricing: An economist's perspective on equity, efficiency and acceptability. Center for Economic Studies-KU Leuven.
- Mazlounian, Amin, Nikolas Geroliminis, and Dirk Helbing. 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (1928): 4627-47.
- . 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (1928): 4627-47.
- Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M., Waraich, R. A., & Axhausen, K. W. 2010. Large-scale agent-based travel demand optimization applied to Switzerland, including mode choice, paper presented at the 12th World Conference on Transportation Research.
- Nagel, Kai, and Gunnar Flötteröd. 2009. Agent-based traffic assignment: going from trips to behavioral travelers. 12th International Conference on Travel Behaviour Research (IATBR), *Jaipur*.
- Newell, Gordon F. 1987. The morning commute for nonidentical travelers. *Transportation Science* 21 (2): 74-88.
- Ortigosa, Javier, Monica Menendez, and Hector Tapia. 2013. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. *EURO Journal on Transportation and Logistics*: 1-22.
- Pigou, Arthur Cecil. 1924. *The economics of welfare* Transaction Publishers.

- Pridmore, Alison, and Apollonia Miola. 2011. Public acceptability of sustainable transport measures: A review of the literature.
- Ramjerdi, Farideh. 2006. Equity measures and their performance in transportation. *Transportation Research Record: Journal of the Transportation Research Board* 1983 (1): 67-74.
- Rawls, John. 1971. *A theory of justice* (Cambridge, Mass.: Harvard University).
- Rieser, Marcel, and Nagel, Kay. 2008. Network breakdown “at the edge of chaos” in multi-agent traffic simulations. *The European Physical Journal B*, 63(3): 321-327.
- Saberi, Meead, and Hani S. Mahmassani. 2012. Exploring properties of networkwide flow-density relations in a freeway network. *Transportation Research Record: Journal of the Transportation Research Board* 2315 (1): 153-63.
- Saifuzzaman, Mohammad, André De Palma, and Kiarash Motamedi. 2012. Calibration of METROPOLIS for ile-de-france.
- Santos, Georgina, and Laurent Rojey. 2004. Distributional impacts of road pricing: The truth behind the myth. *Transportation* 31 (1): 21-42.
- Schaller, Bruce. 2010. New York City’s congestion pricing experience and implications for road pricing acceptance in the United States. *Transport Policy* 17 (4): 266-73.
- Small, Kenneth A. 1992. Using the revenues from congestion pricing. *Transportation* 19 (4): 359-81.
- . 1983. The incidence of congestion tolls on urban highways. *Journal of Urban Economics* 13 (1): 90-111.
- Small, Kenneth A., and Xuehao Chu. 2003. Hypercongestion. *Journal of Transport Economics and Policy*: 319-52.
- Small, Kenneth A., and Erik T. Verhoef. 2007. *The economics of urban transportation* Routledge.
- Small, Kenneth A., Clifford Winston, and Jia Yan. 2005. Uncovering the distribution of motorists’ preferences for travel time and reliability. *Econometrica* 73 (4): 1367-82.
- Small, Kenneth A., and Jia Yan. 2001. The value of “value pricing” of roads: Second-best pricing and product differentiation. *Journal of Urban Economics* 49 (2): 310-36.
- Ubbels, Barry, and Erik Verhoef. 2006. Acceptability of road pricing and revenue use in the Netherlands.
- van den Berg, Vincent AC. 2012. Step-tolling with price-sensitive demand: Why more steps in the toll make the consumer better off. *Transportation Research Part A: Policy and Practice* 46 (10) : 1608-1622.
- Varian, Hal R., and Jack Repcheck. 2006. *Intermediate microeconomics: a modern approach*.

- Verhoef, Erik T. 2002. Second-best congestion pricing in general static transportation networks with elastic demands. *Regional Science and Urban Economics* 32 (3): 281-310.
- Verhoef, Erik T., Peter Nijkamp, and Piet Rietveld. 1997. The social feasibility of road pricing: A case study for the randstad area. *Journal of Transport Economics and Policy*: 255-76.
- Verhoef, Erik, Peter Nijkamp, and Piet Rietveld. 1996. The trade-off between efficiency, effectiveness, and social feasibility of regulating road transport externalities. *Transportation Planning and Technology* 19 (3-4): 247-63.
- Vickrey, William S. 1969. Congestion theory and transport investment. *The American Economic Review* 59 (2): 251-60.
- . 1963. Pricing in urban and suburban transport. *The American Economic Review* 53 (2): 452-65.
- Viegas, Jose M. 2001. Making urban road pricing acceptable and effective: Searching for quality and equity in urban mobility. *Transport Policy* 8 (4): 289-94.
- Walters, Alan A. 1961. The theory and measurement of private and social cost of highway congestion. *Econometrica: Journal of the Econometric Society*: 676-99.
- Waraich, Rashid A., and Axhausen, Kay. W. 2012. Agent-Based Parking Choice Model. *Transportation Research Record: Journal of the Transportation Research Board*, 2319(1): 39-46.
- Wardrop, John Glen. 1952. Road paper. some theoretical aspects of road traffic research. Paper presented at ICE Proceedings: Engineering Divisions, .
- Wee, Bert, and Karst Geurs. 2011. Discussing equity and social exclusion in accessibility evaluations. *European Journal of Transport and Infrastructure Research* 11 (4).
- Yang, Hai, and Qiang Meng. 1998. Departure time, route choice and congestion toll in a queuing network with elastic demand. *Transportation Research Part B: Methodological* 32 (4): 247-60.
- Zheng, Nan, and Nikolas Geroliminis. 2013. On the distribution of urban road space for multimodal congested networks. *Procedia-Social and Behavioral Sciences* 80: 119-138.
- Zheng, Nan, Rashid A. Waraich, Kay W. Axhausen, and Nikolas Geroliminis. 2012. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice* 46 (8): 1291-303.

## Appendix I: Distribution of densities

The figure below shows the frequency distribution of the link densities for traffic conditions characterized by similar accumulation (about 13,5 veh/km/lane) during the morning and evening peak (6:50 and 18:30). The two distributions are slightly different from each other: while distribution of densities in the morning decrease more gradually, in the evening such a decrease is not so smooth. This is confirmed by parameters such as the kurtosis and skewness, which describe the symmetry and shape of the distribution. In the case of the evening distribution, the particularly higher values kurtosis suggests the presence of high tails and the higher value (compared to the morning) of skewness indicates higher left asymmetry. In particular, it is worth mentioning that the number of “very congested” links (above 100 veh/km) at 18:30 is twice the amount at 06:50.



## Appendix II: Logistic regression

Logistic regression is probably the most commonly used procedure when analyzing data with a binary (success/failure) of a target variable  $y$ . In this model, the success probability  $\pi$  is assumed to satisfy (Siminoff, 1998):

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \quad (1)$$

Where  $\{x_1, \dots, x_p\}$  is the set of predicting variables and  $\pi(x)$  the probability of success given  $x$ . In our study case, we modeled the choice of “stop crossing the cordon” or “switch to pt” as a binary dependent variable 1 or 0. The predicting variables express the socio-demographic characteristics as  $\{x_1, \dots, x_p\}$ . If, for example, the influence of trip purpose on the choice to stop entering is considered than eq. 1 would become:

$$\ln\left(\frac{stop(x)}{1-stop(x)}\right) = \beta_0 + \beta_{home}home + \beta_{work}work + \beta_{edu}edu + \beta_{shop}shop$$

The estimates of the coefficient *Beta* express the relationship between the independent variables and the dependent variable on the logit scale. These estimates describe how much the increase or decrease (if the sign of the coefficient is negative) of the predicted log odds of 1 is given by 1 unit increase (or decrease) in the predictor (holding all other predictors constant)<sup>13</sup>.

Since the coefficients are in log-odds units, in order to simplify their interpretation, they are often converted into odd ratios and expressed by means of *Exp(Beta)*. So that, for example:

$$stop(x) = e^{\beta_{home}}$$

This strategy allows an easier comparison among the influence of the different predictors.

---

<sup>13</sup> from <http://www.ats.ucla.edu/stat/sas/notes2/> (accessed October 18, 2013)

### Appendix III: skewness and kurtosis

The two tables below report the skewness and kurtosis values of two different user groups after the implementation of the toll in order to provide additional information about the shape of their distributions. The first index describe if and how the distribution “leans” to the one side of the mean. The second one measures the width of the peak and tails of the distribution. All the groups are characterized by an high index of kurtosis that implies the presence of an acute peak around the mean and two fat tails (leptokurtic shape). The skewness is rather close to zero in all the groups with the exception of the people who stop entering after the Flat Toll is implemented. These results imply a rather symmetrical distribution around the mean. The exception implies that the left tail is longer than the right tail.

<b>Flat Toll</b>	<b>skewness</b>	<b>kurtosis</b>
<b>Continue entering</b>	-0.15	22.30
<b>Stop entering</b>	-1.46	21.36

<b>Step Toll</b>	<b>skewness</b>	<b>kurtosis</b>
<b>Continue entering</b>	0.57	22.12
<b>Stop entering</b>	0.24	18.94

## Appendix IV: t-tests

The tables below report the results of t-tests applied to different socio-demographic categories to compare the effects of the Step Toll and Spread Toll relative to the Flat Toll.

Mean difference	Step toll	Spread toll
Zone 1	Not significant	Not significant
Zone 2	Not significant	-1.22
Zone 3	+1.8	-1.05

Mean difference	Step toll	Spread toll
kreis 1	0,50	-1,15
kreis 2	not significant	-1,72
kreis 3	0,70	-0,88
kreis 4	0,68	-0,25
kreis 5	0,80	not significant
kreis 6	not significant	-1,49
kreis 7	0,88	-0,28
kreis 8	-1,10	not significant
kreis 9	0,42	-0,98
kreis 10	0,59	-1,68
kreis 11	not significant	-0,93
kreis 12	0,39	-1,40

Mean difference	Step toll	Spread toll
Home	+0.95	-0.94
Work	+1.78	Not significant
Education	Not significant	Not significant
Shop&Leisure	+0.88	-1.75

## Appendix V: average gains/losses for different socio-demographic groups

The two figures below reports the average gain and average loss of individuals across the different districts (kreis).

