THE ECONOMICS OF TRIP SCHEDULING, TRAVEL TIME VARIABILITY AND TRAFFIC INFORMATION
ISBN 978 90 361 0351 0

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 558 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.
THE ECONOMICS OF TRIP SCHEDULING, TRAVEL TIME VARIABILITY AND TRAFFIC INFORMATION

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Economische Wetenschappen en Bedrijfskunde
op maandag 22 april 2013 om 15.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Stefanie Peer

geboren te Ried im Innkreis, Oostenrijk
promotoren: prof.dr. E.T. Verhoef
prof.dr. C.C. Koopmans
A small dictionary of useful Austrian scheduling terminology

Es geht sich (alles) (immer) (irgendwie) aus.

This phrase is used to express the ability and willingness to solve a scheduling problem, however, usually without making a commitment to a specific timing. General terms such as alles (everything), immer (always), irgendwie (somehow) and combinations of them are often inserted to further enhance vagueness.

Das geht sich (jetzt) (gerade) nicht (mehr) (ganz) aus.

This phrase indicates the inability and/or reluctance of the speaker to solve a scheduling problem. Terms such as jetzt (now), gerade (just), mehr (more), ganz (quite), and combinations thereof are frequently added to emphasize that the problem could have almost been solved, and to conceal that the solution of the scheduling problem may remain impossible and/or undesired in the future, too.

Source: Experience
Preface

The first time I came across the subject of transport economics was in 2007, when I wrote my Master thesis in Innsbruck, Austria, about the location behavior of individuals and firms in Alpine areas. Noticing that this was a fairly complex combination, I decided to first study transport economics in a country where it would be easily justifiable to ignore mountains in the analyses: The Netherlands. While I faced many challenges doing the research that finally led to this dissertation, the modeling of Alpine structures was certainly not among them.

Many of these challenges have been overcome thanks to the support of my supervisors Carl Koopmans and Erik Verhoef. Carl, thank you for kindly introducing me to the world of Dutch transport policy making, and always reminding me of keeping my research policy relevant. Erik, it was a great pleasure to work with you over the last years. I have benefited immensely from your profound knowledge of economics and your detailed comments on my work. I would also like to express my gratitude to the members of my PhD committee – Bruno de Borger, Ben Immers, Gerard de Jong, John Polak, and Piet Rietveld – for taking the time to read my thesis and providing very valuable comments.

For a major part of my PhD period I was lucky to not only have one, but two workplaces, namely at the KiM (Kennisinstituut voor Mobiliteitsbeleid) in The Hague and at the Department of Spatial Economics at the VU University Amsterdam. In both places I was blessed with very supportive and enthusiastic colleagues. My colleagues at the VU with whom I conducted the research on the Spitsmijden experiment deserve a special word of thanks: Jasper, Paul and Yin-Yen, it was a very nice and instructive experience to investigate the scheduling behavior of Dutch commuters together with you. Paul, thank you also for your positive spirit in our less than successful attempts to derive social optima in stochastic bottleneck models.

I would like to thank my family, and in particular my mum and my dad, for always being there for me when I needed them, and for never questioning what I was doing during all those years in the Netherlands. I am grateful for having two wonderful sisters. Lisa, thanks to you, in summer 2008, I did not only start my career as a transport economist, but also as an auntie to my beloved nephew Luca. Katharina, I am very happy you share my interest in all kinds of spatial topics. Also your help with some of the maps contained in this thesis is truly appreciated.

1The latter does not exactly hold true for my grandfather. His monetary gifts always happened to be a bit larger for me, compared to his other grandchildren, because he could never imagine that someone would pay me a fair salary for being behind a computer all day, writing some dubious papers on even more dubious subjects.
Above all, I would like to thank my friends. Some of them I had already met during school in Ried, others while studying (or shall I rather say snowboarding?) in Innsbruck, cycling through Italy, following the Master program in Rotterdam, working in Amsterdam and The Hague, or during conference visits. And yet others I got to know because we were neighbors, shared flats, or because we had common (boy)friends. I am endlessly grateful to all of you for your friendship and all the precious moments we experienced together over these last few years.

The friends I spent most time with while working on this dissertation are most probably those from the VU. Thank you, Aart, Ceren, Christiaan, Jessie, Łukasz, Sergej, Yuval and many others for great conversations, shared laughter and moral support during the more difficult times. I was so lucky to share my office with Ceren since the start of my PhD. Thank you for being the kindest, most wonderful office mate I could ever have wished for. Unforgettable are all those late nights we spent at the office during the first two years as PhD students, because Aart kept providing us with constant entertainment throughout the day. Yuval, thanks for your great humor, for sharing interesting stories and for being my reliable climbing mate during the last two years we have been together in the same office. We have excelled many routes together, and I hope many more (overhangs) will follow. Christiaan, I think nobody has had more patience in listening to my Austrian-inspired Dutch than you. Ironically, despite your efforts in teaching me Dutch, it was still you who wrote the Dutch summary of this thesis. Thank you so much.

Last but not least, I would like to thank two very special ladies, Jola and Heidi, together with whom I did my Master program in Rotterdam, for their friendship and encouragement, for late-night-teas home and fun evenings out. We did not have a lot of contact during our Master program, but we definitely made up for this in the years that followed. We started off going climbing together, however, as the years passed, our focus shifted more and more towards the socializing part, including dark beers and soccer (PSV Eindhoven–SV Ried will never be forgotten!). This might well have been a subconscious attempt to integrate with the male dominated world of transport economists, and a preparation for our prospective transport economics joint-ventures. The location is still up in the air, however, I believe, all three of us agree that the main choice criterion will be the proximity of mountains.

The generous funding of my research by the Dutch Ministry of Infrastructure and Environment (IenM) is greatly appreciated.
Contents

Preface iii

Contents v

List of Figures vii

List of Tables viii

List of Abbreviations ix

1 Introduction 1
   1.1 Motivation and background . . . . . . . . . . . . . . . . . . . . . . . . 2
   1.2 Valuation of travel time, schedule delays and variability . . . . . . . 4
   1.3 Structure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
   1.4 Scope . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2 Prediction of travel time variability for cost-benefit analysis 11
   2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
   2.2 Data and variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
   2.3 Empirical analysis of travel time variability . . . . . . . . . . . . . . . 19
   2.4 Implications for CBA . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
   2.5 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

3 Door-to-door travel times in departure time choice models 31
   3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
   3.2 Related literature . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
   3.3 Geographically Weighted Regression . . . . . . . . . . . . . . . . . . . 38
   3.4 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
   3.5 Results: Geographically Weighted Regression . . . . . . . . . . . . . . 44
   3.6 Departure time choice models . . . . . . . . . . . . . . . . . . . . . . . 46
   3.7 Results: Departure time choice models . . . . . . . . . . . . . . . . . . 52
   3.8 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

4 Overstating vs. overreacting: Commuters’ perceptions of travel times 55
   4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
   4.2 Related literature . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
   4.3 Specification of the travel time ratio . . . . . . . . . . . . . . . . . . . . 62
# List of Figures

1 Outline of the thesis ........................................ 7
2 Illustration speed-flow curve ................................. 13
3 Network of highway links included in the analysis ........ 15
4 Relating mean delay and standard deviation ............... 20
5 Traffic regimes ............................................... 23
6 Difference of fitted and actual values of variability ...... 25
7 Variability predictions based on Model 2 .................. 27
8 Schematic map of the experimental setup ................. 33
9 Speed observations ......................................... 34
10 Optimal bandwidth choice .................................. 45
  10.1 Home–C1 ................................................. 45
  10.2 C2–Work ................................................ 45
11 Kernel weights ............................................. 46
12 Speed predictions ......................................... 47
  12.1 Map ..................................................... 47
  12.2 Predictions: C1–C2 speed = 50 km/h .................. 47
  12.3 Predictions: C1–C2 speed = 100 km/h .............. 47
13 Literature framework ....................................... 60
14 Histogram travel time ratio (home–work) .................. 65
15 Screenshot SP experiment .................................. 72
16 Diagrammatic exposition: SRPAT ≤ LRPAT ............... 89
17 Descriptives LRPAT and SRPAT ............................ 94
  17.1 Average weekday-specific travel times ................. 94
  17.2 Scatter LRPAT vs. SRPAT ............................. 94
  17.3 Histogram LRPAT .................................... 94
  17.4 Histogram SRPAT .................................... 94
18 LRPAT ± x minutes ........................................ 99
19 Standard bottleneck model .................................. 111
20 Unpriced equilibrium ....................................... 115
21 First-best tolls and schedule delay costs .................. 120
22 Tolls and schedule delays if only short-run tolls are available .... 121
23 What determines your expected travel time? ............. 143
24 How often do you inform yourself about actual traffic conditions? .... 144
25 Behavioral response to delays ............................. 145
26 Reported travel time distributions ......................... 146
27 Cumulative distribution of travel times .................... 147
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FI: explanatory variables</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>Results auxiliary regressions for FI</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>Regression results (non-traffic-regime-based model)</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>Regression results (traffic-regime-based model)</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Variability predictions (RI) based on Model 2</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>Variability predictions (FI) based on Model 2</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>Descriptives GPS</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Results OLS and GWR model</td>
<td>46</td>
</tr>
<tr>
<td>9</td>
<td>Descriptives of drivers considered in the choice model</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>Estimation results: departure time choice models</td>
<td>53</td>
</tr>
<tr>
<td>11</td>
<td>Reported vs. observed travel times (in minutes)</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>Descriptive statistics of driver- and link-specific variables</td>
<td>66</td>
</tr>
<tr>
<td>13</td>
<td>Regression results: Explanation of the travel time ratio</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>Descriptive statistics</td>
<td>73</td>
</tr>
<tr>
<td>15</td>
<td>MNL models</td>
<td>75</td>
</tr>
<tr>
<td>16</td>
<td>Panel latent class model</td>
<td>78</td>
</tr>
<tr>
<td>17</td>
<td>Descriptives datasets</td>
<td>96</td>
</tr>
<tr>
<td>18</td>
<td>Main estimation results</td>
<td>96</td>
</tr>
<tr>
<td>19</td>
<td>Robustness checks</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>Latent class estimation results</td>
<td>101</td>
</tr>
<tr>
<td>21</td>
<td>Overview datasets used in Chapters 3–5</td>
<td>142</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>C1, C2</td>
<td>Camera locations</td>
<td></td>
</tr>
<tr>
<td>CBA</td>
<td>Cost-benefit analysis</td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>Fixed effects</td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td>Fine information</td>
<td></td>
</tr>
<tr>
<td>GPS</td>
<td>Global positioning system</td>
<td></td>
</tr>
<tr>
<td>GWR</td>
<td>Geographically weighted regression</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>Long-run</td>
<td></td>
</tr>
<tr>
<td>LRPAT</td>
<td>Long-run preferred arrival time</td>
<td></td>
</tr>
<tr>
<td>MNL</td>
<td>Multinomial logit</td>
<td></td>
</tr>
<tr>
<td>MXL</td>
<td>Mixed logit</td>
<td></td>
</tr>
<tr>
<td>NTR</td>
<td>Non-traffic-regime-based</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary least squares</td>
<td></td>
</tr>
<tr>
<td>PAT</td>
<td>Preferred arrival time</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>Random effects</td>
<td></td>
</tr>
<tr>
<td>RI</td>
<td>Rough information</td>
<td></td>
</tr>
<tr>
<td>RP</td>
<td>Revealed preference</td>
<td></td>
</tr>
<tr>
<td>SDE</td>
<td>Schedule delay early</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>Schedule delay late</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>Stated preference</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>Short-run</td>
<td></td>
</tr>
<tr>
<td>SRPAT</td>
<td>Short-run preferred arrival time</td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>Traffic-regime-based</td>
<td></td>
</tr>
<tr>
<td>VOR</td>
<td>Value of (travel time) reliability</td>
<td></td>
</tr>
<tr>
<td>VOT</td>
<td>Value of (travel) time</td>
<td></td>
</tr>
<tr>
<td>VSDE</td>
<td>Value of schedule delay early</td>
<td></td>
</tr>
<tr>
<td>VSDL</td>
<td>Value of schedule delay late</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction
1.1 Motivation and background

An efficient and reliable transport system is crucial for the functioning of modern economies, as it gives people as well as firms the possibility to engage in different activities at different locations. Passenger transport, for instance, enables people to choose residential locations that are different from their work locations. The large extent to which people make use of this possibility gives rise to the well-known phenomenon of traffic congestion.

This dissertation tries to achieve a better understanding of scheduling choices individuals make in the face of congestion. From disaggregate data on departure time and travel routine choices, the travelers’ willingness to pay for reducing recurrent and non-recurrent congestion is estimated. Understanding travel behavior at the level of the individual, rather than at a more aggregate (e.g. network) level facilitates more precise predictions of travelers’ responses to transport policies, and thus more accurate policy appraisals. Throughout the dissertation the role of traffic information is emphasized. So, it affects the distinction between recurrent and non-recurrent congestion, the travelers’ valuation of travel time and trip timing aspects, as well as optimal road pricing schemes.

Recurrence congestion tends to be concentrated in time as well as in space, with the most prominent examples being the morning and the evening peak, caused by commuting traffic. In the Netherlands, for instance, more than 75% of the kilometers travelled by private persons between 7:00 and 8:00 a.m. are due to commuting, while outside peak hours the corresponding share ranges between 10 and 20% (Ministerie van Verkeer en Waterstaat, 2011). The temporal concentration of commuting can be attributed to a variety of factors, such as the natural day-night rhythm and school starting hours. Also positive temporal agglomeration forces at work might be relevant. These imply that productivity increases in the number of people present at the workplace, for instance due to better possibilities for collaboration and communication among employees (e.g. Hall, 1989; Henderson, 1981; Weiss, 1996).

Costs induced by recurrent congestion do not only result from travel time losses, but also when travelers, in the face of congestion, choose not to arrive at their planned destination at their preferred time of the day: Since travelers are able to anticipate the recurrent congestion pattern, they can decide to travel earlier or later in order to reduce travel time. In turn, they incur so-called schedule delay costs from not arriving at their preferred arrival time. In equilibrium, commuters who arrive close to their preferred arrival time face the longest travel times, but the lowest schedule delay costs. Vickrey (1969) was the first to model these trade-offs between costs of travel time and costs of schedule delays. His model of a road bottleneck explicitly takes into account the dynamic nature of congestion as a result of endogenous departure time decisions. Chapter 6 of this thesis extends the standard bottleneck model by adding endogenous choices of travel routines.

Non-recurrent congestion implies stochasticity of travel times at a given moment in time. The latter can be caused by both demand-side as well as supply-side fluctuations (Bates et al., 2001). A typical example that underlines
the unexpected nature of non-recurrent congestion are traffic incidents. It has been shown both theoretically as well as empirically that the extent of non-recurrent congestion tends to increase during periods in which recurrent congestion is growing, and to decrease during periods in which recurrent congestion is ceasing (e.g. Fosgerau, 2010; Fosgerau et al., 2008, as well as Chapter 2 of this dissertation). While recurrent congestion causes schedule delays with respect to the preferred moment of arrival, non-recurrent congestion causes schedule delays with respect to the expected moment of arrival. In addition to the schedule delay costs, non-recurrent congestion may provoke additional disutility due to the anxiety and stress related to not knowing the exact arrival time, and therefore not being able to make precise plans. Some authors refer to these costs as planning costs (e.g. Noland et al., 1998).

Different indicators are used to describe the time-of-day specific probability distribution of non-recurrent delays, which is often referred to as travel time variability, or its counterpart, reliability. The indicators used most frequently are the standard deviation (e.g. Senna, 1994) and the variance (e.g. Jackson and Jucker, 1982). In addition to these measures of variability that are defined from the user’s perspective, the engineering literature often takes the viewpoint of the network operator, emphasizing the role of network performance. A commonly used concept is network robustness, which is typically defined as the ability of a network to cope with serious disruptions (e.g. Immers et al., 2004).

The distinction between recurrent and non-recurrent delays can be fairly ambiguous, since the boundary between these two types of delays is largely determined by the (traffic) information available to drivers. For instance, a delay that has been expected and is considered recurrent by one driver, may be unexpected and hence considered non-recurrent by another driver who is less well-informed. Information comprises two elements here: First, the knowledge about the factors that tend to affect travel times such as weather conditions, weekdays or seasonal patterns; and second, the knowledge on how these factors affect travel times. Undoubtedly, when drivers are well-informed about the attributes of their choice alternatives, they can more competently optimize their travel choices. Also the timing of the information matters. The earlier drivers obtain the traffic information, the better they are able to adapt their choices (e.g. Polak and Jones, 1993). Technical advances during recent years have greatly improved the availability as well as the quality of available information. However, despite of these obvious benefits of information for individual drivers, Arnott et al. (1999) prove the existence of equilibria where more refined information may lead to lower overall efficiency.

Uncertain outcomes are not only inherent to travel decisions but to a wide range of (economic) decisions. As a consequence, chapters dedicated to uncertainty and information can be found in almost all standard economics textbooks (e.g., 2 Other studies use inter-percentile ranges, such as the differences between the 90th and the 50th percentile of the travel time distribution (e.g. Lam and Small, 2001; Small et al., 2005), or buffer time and tardy trip indices. See Van Lint et al. (2008) for an overview. 3 In few studies, robustness is defined from the users’ perspective, usually referring to extreme delays, for instance as result of incidents (e.g. Korteweg and Rienstra, 2010).
Chapter 1. Introduction

Mas-Colell et al., 1995; Pindyck and Rubinfeld, 2001; Varian, 1992). Consumers, for instance, may encounter uncertainty related to the quality of a product they consider buying, while firms may face uncertainties with respect to their future demand and cost levels, and – in an oligopolistic market – also with respect to the costs and strategic actions of competitors (e.g. Tirole, 1988). In many of these choice situations, uncertainty can be reduced through the acquisition of information, inducing economists to study questions concerning the value and the optimal amount of information (e.g. Hirshleifer and Riley, 1992).

1.2 Valuation of travel time, schedule delays and variability

A large share of the benefits of transport policies is usually due to reductions in recurrent and non-recurrent congestion (Hensher, 2001; Small and Verhoef, 2007). For instance, in the appraisal of the congestion charging scheme in London, benefits from reductions in travel times account for 77% and benefits from improvements in reliability for 10% of overall benefits (Evans, 2007). While the benefits due to travel time savings have been included in cost-benefit analyses of transport policies for many decades, the inclusion of benefits related to non-recurrent congestion has a much shorter tradition (e.g. OECD & International Transport Forum, 2010). Up to now, most cost-benefit analyses still ignore benefits that result from reductions in scheduling costs caused by recurrent congestion. In a study focused on the Dutch highway network, Koopmans and Kroes (2004) argue that these should be approximately equal in size to the benefits from travel time reductions.

The benefits related to a specific transport policy depend strongly on the assumptions made regarding the travelers’ willingness to pay for reducing recurrent and non-recurrent congestion. The unit value attached to a decrease in travel time is generally referred to as ‘value of time’ (VOT). Concerning the unit values attached to reductions in (both recurrent and non-recurrent) schedule delays, a distinction is commonly made between the ‘value of schedule delay early’ (VSDE) for earliness and the ‘value of schedule delay late’ (VSDL) for lateness. Furthermore, the ‘value of reliability’ (VOR) refers to the value attached to an improvement in reliability (or conversely, a decrease in travel time variability).

Different utility functions have been specified for the analysis of scheduling behavior. Small (1982), building on the work of Becker (1965) and DeSerpa (1971) on the allocation of time on non-work activities, was the first to estimate the values of time and schedule delays in a departure time choice model. He takes into account recurrent congestion, but ignores non-recurrent congestion. Contrariwise, early studies that do take into account non-recurrent congestion usually ignore recurrent congestion. Some of these focus on the scheduling implications of non-recurrent congestion (e.g. Gaver, 1968; Knight, 1974; Polak, 1987), while others assume that the utility function is (linearly) dependent on travel time and a measure of travel time dispersion (e.g. Black and Towriss, 1993; Jackson and Jucker, 1982). The latter are frequently labelled ‘mean-variance’ approaches, even though not all of them use the variance of travel times as an indicator of non-recurrent congestion. Noland
and Small (1995) were the first to conjoin these models, taking into account both recurrent as well as non-recurrent congestion. Following their work and the study undertaken by Bates et al. (2001), Fosgerau and Karlström (2010) show analytically that the optimal departure time depends linearly on the mean and the standard deviation of the travel time distribution, given that the standardized travel time distribution is constant over the time of the day. Under this assumption, the ‘mean-variance’-type representation of non-recurrent congestion is thus equivalent to its representation in terms of schedule delays.

The valuations of travel time, schedule delays and variability are generally computed from disaggregate revealed preference (RP) or stated preference (SP) data on travel and in particular scheduling choices. The defining characteristic of RP data is that the corresponding choices are observed in real choice situations, while SP data are collected from choice experiments that take place in a laboratory setting. The choice sets are usually defined such that the possible choice alternatives are discrete rather than continuous. The primary transport-related examples of choice sets are choices between departure time intervals, routes and modes. Random-utility discrete choice models (McFadden, 1974) can then be used to estimate the trade-offs between the chosen and unchosen alternatives (see Chapters 3–5 for applications of this method).

Most empirical travel choice models use SP data. One reason for the predominance of SP data is that their collection is more convenient and less costly than the collection of RP data. So, RP experiments that allow for the estimation of monetary valuations require a cost component (a toll or reward) that varies by time of the day, route, or mode. Moreover, travel times for both the chosen as well as the unchosen scheduling choices must be known to the analyst (see Chapter 3 of this dissertation for a more detailed discussion on this topic). And finally, the RP setting must be such that the attribute values of the choice alternatives are not overly correlated with each other; otherwise, it becomes impossible to determine the according coefficient estimates.

While SP data are usually easier to gather than RP data, they suffer from the drawback that respondents may behave differently in a laboratory setting than in a real-life setting. Reasons for this so-called hypothetical bias may for instance be strategical interests in affecting the future implementation of policies, time inconsistency in actual but not in hypothetical behavior, or difficulties to understand the choice task (e.g. Brownstone and Small, 2005; Hensher, 2010; Louviere and Hensher, 2001; Wardman, 2001). Also the format of the SP questions may affect the results (see for instance Tseng et al. (2009) for a discussion on different possibilities to represent variability in an SP departure time choice experiment). More recently, various studies use datasets containing both SP and RP data with the intention that the combination corrects for deficiencies each data source has when used separately (e.g. Börjesson, 2008; Brownstone and Small, 2005; Small et al., 2005, as well as Chapter 4 of this dissertation).
The estimated values of travel time, schedule delays and variability differ widely across studies. The valuations are found to depend, among others, on the trip motive as well as the travel mode. Recent literature reviews regarding the value of time can for instance be found in Shires and De Jong (2009), Wardman (2001), and Zamparini and Reggiani (2007), while the following studies contain meta-analyses regarding scheduling and reliability valuations: Brownstone and Small (2005); Carrion and Levinson (2012); Li et al. (2010); Tseng et al. (2005). The VOT is generally found to be between 20 and 90% of the corresponding gross wage rate (Small and Verhoef, 2007). It is a common result that the VSDE is lower than the VOT, and the VSDL is higher than the VOT (e.g. Asensio and Matas, 2008; Hollander, 2006).

Most studies suggest that the VOR (when reliability is expressed as the standard deviation of travel times) is between 0.5 to 1.6 times as high as the VOT (e.g. Li et al., 2010). However, the estimate obtained for the VOR depends on the way recurrent delays are included in the utility function of the travelers. The VOR can either be estimated directly, if a measure of travel time dispersion is included in the utility function (‘mean-variance’ method), or it can be computed from the VSDE and the VSDL using the analytical results obtained by Fosgerau and Karlström (2010). The first approach is widely used in practice as it is very easy to apply. While it ignores the dynamic nature of scheduling choices, it may still provide for an indirect estimate of scheduling preferences. For this to hold, the standardized travel time distribution needs to be constant, and the resulting VOR must only be applied to travel time distributions of the same type (Fosgerau and Karlström, 2010).

The second approach of defining the VOR as a function of scheduling preferences is largely favored in the academic literature. The main reason is that scheduling models provide a better behavioral underpinning by explicitly taking into account scheduling decisions and the consequences thereof: namely, to arrive early or late. This argument is supported by empirical evidence showing that the value attached to variability per se tends to be fairly low once scheduling variables are included in the model (e.g. Asensio and Matas, 2008; Noland and Polak, 2002; Noland et al., 1998; Small, 1999). Another advantage of the second approach is that for a given set of scheduling valuations, the VOR can be computed for any travel time distribution. Fosgerau and Karlström (2010) suggest that the VSDE and VSDL can be derived from SP experiments that include deterministic delays as a choice attribute, avoiding the difficulties of representing non-recurrent congestion in an SP setting. However, Chapter 5 of this dissertation, challenges the underlying assumption that recurrent and non-recurrent delays are valued equally.

Due to the advantages of the scheduling approach, most chapters of this dissertation represent non-recurrent congestion in terms of schedule delays rather than measures of dispersion. Chapter 2 is an exception: Due to its focus on the relation between recurrent and non-recurrent congestion at the aggregate (network link) level, it is natural to express non-recurrent congestion by an indicator of dispersion (in this case, the standard deviation).

---

6The VOR/VOT ratio is often referred to as reliability ratio.
1.3 Structure

This dissertation contains various studies that aim at achieving a better understanding of the scheduling decisions of travelers, specifically in situations when recurrent and non-recurrent congestion are present and information on travel times is available. An overview of its outline is provided in Figure 1.

Figure 1: Outline of the thesis

Chapter 2 starts by developing simple rules of thumb to compute the costs of non-recurrent congestion for the case that drivers have either ‘rough’ or ‘fine’
information on travel times. In contrast to Chapters 3–6, the analysis in Chapter 2 is not conducted at the level of the individual travelers, but at the more aggregate level, using travel time observations along highway links. The studies conducted at the level of the individual traveler comprise both empirical (Chapters 3–5) as well as theoretical (Chapter 6) analyses.

Chapters 3–5 have in common that they use RP data from a real-life peak avoidance experiment, which is described in detail in Appendix A of this thesis. In all three chapters, monetary valuations of travel times and schedule delays are derived, using discrete choice models. The latter requires that travel times are known for both chosen as well as unchosen departure time alternatives. Chapter 3 shows the relevance of using door-to-door travel times for this purpose, and develops a method to derive these for the case that continuous travel time measurements are only available for the main road network and sparse GPS measurements are available for the secondary network. The resulting door-to-door travel times are then used as an input for the RP-based departure time choice models presented in Chapters 4 and 5. Chapter 4 studies travel time (mis-) perceptions of travelers, and how these may affect the valuations of travel time and schedule delay obtained in both RP- and SP-based scheduling models. In Chapters 5 and 6, a distinction between long-run choices of travel routines and short-run choices of departure times is introduced. While Chapter 5 focuses on the empirical implications of this distinction, Chapter 6 shows how the distinction can be incorporated in a theoretical bottleneck model.

Chapter 2 investigates the relation between recurrent and non-recurrent congestion for a number of highway links, with the goal to derive rules of thumb that can be used to compute the costs of non-recurrent congestion in cost-benefit analyses. Two different indicators of non-recurrent congestion are defined, both of which are based on the standard deviation. The first measure assumes that, for a given road link and given time of day, expected travel times are constant across all working days (‘rough information’), while the second one assumes that expected travel times depend on day-specific factors such as weather conditions or weekdays (‘fine information’). Costs of non-recurrent congestion can then be estimated for both measures as a function of recurrent congestion as well as other variables such as the length of the link, traffic management measures and the relative predominance of the three main traffic regimes (free-flow, congested, hyper-congested).

Chapter 3 develops a method to approximate driver-, day- and time-of-day-specific door-to-door travel times for situations where continuous travel time measurements are only undertaken on specific parts of the road network, whereas sparse GPS observations are available for the remaining parts of the network. The resulting travel times are then used to compute the attributes of the departure time alternatives, based on which travel time and scheduling valuations are estimated, employing the random utility framework. In order to apply this framework, it is crucial that the attributes of the (discrete) departure time alternatives are known for both chosen as well as unchosen departure time alternatives. This chapter shows that biases in the valuations may result from not using door-to-door travel times or imprecise measures thereof.
Chapter 4 discusses travel time perceptions. Participants of the peak avoidance experiment have been asked to provide estimates of their average travel times for various parts of their home-work commute. These are compared to the observed travel times, revealing that on average they overreport travel times by a factor of approximately 1.5. Chapter 4 investigates whether this finding simply reflects a reporting error, or whether it also translates to their actual choice behavior in RP and SP settings. Specifically, the hypothesis brought forward by Brownstone and Small (2005) is tested: If drivers overestimate delays in reality, it is likely that in an SP setting travelers react to stated delays as if they were overestimated by an extent similar to the one they apply themselves, resulting in a relatively low value of time for the SP data. For example, if a person perceives a true 10-minute delay as a 15-minute delay, he might react to a 15-minute delay in an SP setting in the same way as he would react to a 10-minute delay in reality.

Chapter 5 and 6 introduce a distinction between long-run choices of travel routines and short-run choices of departure times. Theoretical and empirical studies on scheduling behavior have so far widely ignored that travelers have more flexibility to adjust their schedule in the longer run compared to the shorter run. The preference structure assumed in these two chapters reflects that a difference exists between the ‘long-run preferred arrival time’, which would be relevant if travelers were unconstrained in their scheduling decisions, and the ‘short-run preferred arrival time’, which is the routine arrival time that drivers choose in the face of recurrent congestion. Also differences between long-run and short-run travel time expectations are considered, which occur because more information on travel time realizations becomes available in the short run. While in Chapter 5 short-run and long-run values of time and schedule delay are estimated, Chapter 6 comprises a theoretical bottleneck model in the spirit of Vickrey (1969), which allows for an endogenous choice of (long-run) travel routines and (short-run) departure times. The private equilibrium, the social optimum as well as second-best optima are characterized.

1.4 Scope

This dissertation covers mainly passenger transport, with a particular focus on commuting trips (rather than other trip motives) and its epiphenomenon of peak-hour congestion. One reason for omitting freight transport is that it plays a fairly minor role during peak hours. For instance, Adviesdienst Verkeer en Vervoer (AVV) (2005) states that on Dutch highways, freight transport accounts for only 11-12% of overall road transport during the morning and the evening peak (in contrast to 17% during the rest of the day). Another reason is that scheduling decisions in freight transport differ substantially from scheduling decisions of commuters. One main difference is that logistics processes usually involve multiple parties such as shippers, hauliers, and third-party logistics operators, all of which may attach different values to a reduction in recurrent and non-recurrent congestion (e.g. Fowkes et al., 2004).
Furthermore, the studies contained in this thesis only consider car travel, thus disregarding public transport or non-motorized modes. Again, this follows from the research interest in peak-hour congestion and its effects on scheduling decisions. For no other mode, recurrent congestion is as evident as for car travel. On more pragmatic grounds, cars are still the dominant mode of transport. For instance, in the Netherlands, almost half of all trips are taken by car. In terms of distances travelled, car travel even accounts for 74% of the overall number of kilometers traveled in the Netherlands (Kennisinstituut voor Mobiliteitsbeleid (KiM), 2011).

Route choices are not considered in this dissertation either. In the context of peak-hour commuting, scheduling choices tend to be more important than route choices, since recurrent delays are likely to affect the entire network. As a consequence, the possibilities to gain from from switching to alternative routes are limited (e.g. Chorus et al., 2006; Jou and Mahmassani, 1996).

The focus of this dissertation on scheduling decisions also affects the way how traffic information is modeled. The analyses contained in this thesis concentrate on information concerning (road) supply and demand that drivers are typically aware of in advance of traveling (e.g. weather conditions, day of the week, season). Relatively little emphasis is given to the modeling of information that becomes available only on the day of travel, or during the travel itself (i.e. real-time information), as the flexibility for rescheduling is limited in this case.

Even though Chapters 5 and 6 refer to the choice of routine travel behavior as long-run behavior, this label should be understood in a relative sense, namely as opposed to short-run departure time decisions. This dissertation does not cover ‘actual’ long-run decisions concerning for instance the place of residence, or the type of work to engage in (e.g. Clark et al., 2003; Kim et al., 2005; Van Ommeren et al., 1999; White, 1977).

Finally, the behavioral models used in this dissertation assume that scheduling choices are the result of (expected) utility maximization. Alternative theories of behavior such as those based on the concept of bounded rationality (Simon, 1955) are not explicitly considered here.
CHAPTER 2

Prediction of travel time variability for cost-benefit analysis
2.1 Introduction

Travel times on roads are usually not stable over time. Variations occur in demand for travel as well as in road capacity, and these cause travel times to vary. Parts of these fluctuations are known to travelers. Especially frequent travelers are aware of recurrent congestion patterns during peak hours, or the adverse effect of heavy rain on travel times. This study focuses on variations in travel times that are not expected by drivers. They cause drivers to arrive at their destinations earlier or later than expected. In most cases, such delays come at a cost to drivers. They might face waiting times, or the need to reschedule activities. With most people being risk-averse, the uncertainty of arrival time might be accompanied by feelings of stress and anxiety.

To include the costs associated with unreliable travel times in cost-benefit analysis (CBA), both the drivers’ (monetary) valuation of unreliable travel times and the extent of unreliability with and without the project need to be known. While the values that drivers attach to travel time variability and schedule delays have been derived in various stated and revealed preference experiments (e.g. Hensher, 2001; Lam and Small, 2001; Small, 1982), only little research has been done on explaining and forecasting the extent of travel time variability. Most of the research in this area is based on simulation studies. The outcomes of these studies are usually not easily transferable to CBA, as they focus on stylized cases of traffic networks without the use of empirical data (e.g. Li et al., 2009; Nagel and Rasmussen, 1995; Nicholson and Du, 1997). Studies that explicitly predict the variability of travel times based on empirical data, and calculate the resulting costs, were done by Eliasson (2006) and Fosgerau et al. (2008). They both point out the importance of including variability-related costs into CBA, suggesting that 10–15% of the costs associated with changes in travel times can be attributed to changes in variability. Tu (2008) focuses on the relation between travel time variability and flow, however, without computing the resulting costs of variability.

A common result of previous research is that the main explanatory factor of travel time variability is mean travel time. Fosgerau (2010) proves that, under the assumptions of Vickrey’s bottleneck model, if mean travel times are increasing between two time periods, the standard deviation of travel times increases as well. However, he also shows that the mean and the standard deviation vary differently from each other during the peak. Kouwenhoven et al. (2005) show empirically that a higher mean travel time usually implies a higher level of variability. Eliasson (2006) derives similar results but he also shows that if congestion is very severe, variability can become a decreasing function of travel time.

Also in the current study, we find a non-linear relationship between travel time variability and mean travel time. We investigate whether the non-linearity can be attributed to different ‘traffic regimes’. The regimes can be defined according to their location in the speed-flow relationship of traffic, which follows from the so-called fundamental diagram of traffic that shows how speeds fall with traffic density (Greenshields, 1935; Haight, 1963; Hall, 2002; Hall et al., 1992; Kerner, 2004).

---

7This chapter is based on Peer et al. (2012).
In this study, we distinguish between the free-flow, congested and hyper-congested regime. The free-flow regime refers to travel speeds close to free-flow speed, and flow rates considerably lower than the maximum capacity of a road segment. The congested regime describes conditions where travel times are significantly higher than free-flow travel times, and flow is relatively close to maximum capacity. Finally, hyper-congestion occurs if the traffic density that implies a maximum flow (capacity) is exceeded and queuing takes place, leading to a below-maximum flow and considerable delays. In terms of the speed-flow curve plotting traffic flow on the horizontal and speed on the vertical axis, the free-flow and congested regimes are located at the upper (downward bending) segment of the curve, whereas the hyper-congested regime corresponds with the lower (upward bending) section of the curve. An illustration of a speed-flow curve is shown in Figure 2.

Only few papers have looked explicitly at the influence of flow on travel time variability. Tu et al. (2007) investigate the correlation between (in)flow variability and travel time variability. They find that travel time variability is hardly related to the variability of flow in the free-flow and hyper-congested regime, whereas it is positively correlated with flow variability in the congested regime.\(^8\) We focus on the impact of the relative shares of the traffic regimes on variability.

The current study also introduces two measures of travel time variability. These differ in the assumption on the extent of information available to drivers. The first measure is based on the assumption that for a given road link, and a given time of the day, expected travel times are constant over the entire year. We will call this ‘rough information’ (RI). The second measure assumes that drivers adjust expectations on travel times according to (public) information on weather conditions, weekday, season and demand patterns, and will be referred to as ‘fine information’ (FI).

---

\(^8\)Tu et al. (2007) define three traffic regimes: fluent traffic, transition traffic and capacity traffic. The definitions correspond closely to the ones used here (free-flow, congested and hyper-congested traffic).
Chapter 2. Prediction of variability

This chapter thus contributes to the existing literature on predicting travel time variability by explaining the linkage of travel time variability and mean delay for different traffic regimes. The results give a clear indication that the relative shares of the traffic regimes have an impact on the relation between mean delays and standard deviation. Unlike earlier literature, this study provides two measures of variability, and has a strong focus on the applicability of the results in CBA.

The chapter is organized as follows. Section 2.2 provides an overview of the travel time data and the variables used in the analysis. Section 2.3 contains the empirical analyses. A non-traffic-regime based model (NTR) as well as a traffic-regime based model (TR) are derived, and results are analyzed. Section 2.4 proceeds by discussing the implications of the results for cost-benefit analysis. Finally, Section 2.5 provides concluding remarks and suggestions for further research.

2.2 Data and variables

The goal of the research is to identify factors that determine travel time variability and to estimate their influence on variability. The focus is on variables that lend themselves for inclusion in CBA. That is, variables should be likely to be readily available to cost-benefit analysts (for instance as output of traffic models). Besides this, variables should not be defined at a too detailed level. CBA are usually used as an instrument to assess the costs and benefits of relatively large-scaled projects. Therefore, cost calculations based on very detailed variables are usually neither useful nor feasible.

Travel time data

For the statistical analysis we use travel time distributions of 145 (one-directional) highway links in the Netherlands. Since travel time data are not available for door-to-door trips, we focus on the highway network, which is very dense in the Netherlands. For this reason, almost 50% of all driving takes place on highways.\(^9\)

Figure 3 shows a map of the Netherlands, indicating all road stretches included in the analysis. Most of them are concentrated in the West of the country, which is one of the most densely populated areas in Europe.

Links have been defined in a way such that each of them starts and ends after a highway intersection. The link length varies between 2.2 km and 37.1 km, with an average of 13.2 km. To define road stretches in this way is partly motivated by the idea that for drivers these intersection points are natural points of reference. Frequently they are bottlenecks and in many cases they are also distinct landmarks. Not choosing all road links to be of equal length also has the additional advantage that it allows to investigate the effect of length (uninterrupted by intersections) on variability, and thus renders the research applicable to a larger variety of roads.

\(^9\)This number refers to non-freight transport. The percentage is based on the number of km driven on Dutch highways in 2006 (around 50 billion km, assuming a 20% rate of freight transport) divided by the number of overall km driven (around 95 billion km). (Ministerie van Verkeer en Waterstaat, 2006, 2008)
2.2. Data and variables

The dataset comprises travel time data of all working days in 2008 (255 days). Weekends and public holidays have been excluded as demand patterns are usually quite different on such days. In addition, we only consider the time span between 6:00 and 20:15 (57 quarters).\textsuperscript{10} Thus, for each link, 57 travel time distributions are defined, each comprising the travel time observations on the 255 working-days of 2008 for a specific quarter of an hour.

\textsuperscript{10}One reason for excluding night-time observations is that relatively short delays during night hours are usually not caused by congestion but rather by day-to-day differences in the composition of traffic flow. For instance, an above-average number of freight vehicles might lead to travel time observations that are higher than average. This type of variability, however, does not result in variability-induced costs to the drivers and is therefore not useful to include in cost-benefit analyses. Moreover, longer delays during night time may be caused road works. However, these are usually announced upfront and a large share of drivers is expected to be aware of them.
Travel time variability

As in many previous studies, we use the standard deviation (SD) of travel times to represent variability. The standard deviation can be computed easily and is a natural way of expressing the dispersion of distributions. Moreover, it can be easily integrated in traffic assignment models (e.g. Significance et al., 2012). We verified that our main results remain valid when alternative measures of variability are used. An example are percentile based indicators, such as the difference between the 90th percentile and the median of the travel time distribution, as suggested by Lam and Small (2001)\(^{11}\). For an overview of alternative indicators of dispersion we refer to Van Lint et al. (2008).

In the transport economics literature, the valuation of unreliability of travel times is often related to the valuation of implied expected schedule delays, rather than being directly related to dispersion indicators. This so-called scheduling model allows for the distinction between the valuation of schedule delays early and schedule delays late (with respect to a preferred arrival time) (Hollander, 2006). We do not use this approach here, as we do not observe the preferred arrival times of drivers in our sample. Recent literature (Fosgerau and Karlström, 2010; Li et al., 2009) has shown that under certain conditions, knowledge of the preferred arrival times is in fact not necessary. In case of a constant (standardized) travel time distribution, the variability induced expected costs for establishing a valuation of unreliability consistent with the scheduling model of arriving early or late can be expressed in terms of the standard deviation of travel times. In these cases, a direct correspondence exists between the approach used in this study and the approach that involves scheduling costs.

We propose to distinguish between two different measures of variability. The distinction stems from the fact that travel time variability can either be defined relative to expectations based on what we call ‘rough information’ (RI) or ‘fine information’ (FI). In the RI case, we assume that expected travel times for a given link and a given quarter are constant over successive days, and thus only vary between links and quarters of an hour. In contrast, the FI case assumes that travelers adjust their expectations on travel times also according to weather conditions, season, weekdays and network-wide delays. Measures for travel time variability can be expected to be smaller for FI than for RI. Whether predictions of variability based on RI or FI are used in CBA should therefore depend on the context within which the value of reliability has been derived. The corresponding formulas for travel time variability based on RI and FI are shown below:

\[
SD_{RI}(q, l) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (tt_{i,q,l} - \mu_{q,l})^2} \quad (2.1)
\]

\[
SD_{FI}(q, l) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (tt_{i,q,l} - \mu_{i,q,l})^2}, \quad (2.2)
\]

\(^{11}\)These results are available from the authors upon request.
where the index \( q \) indicates a specific quarter of an hour, \( l \) a specific road link, and \( i \) stands for a specific day of the year. \( N \) is the number of days (255), \( t_t \) the travel time, and \( \mu \) the expected travel time. In the case of RI, \( \mu \) is constant for a certain quarter and link. For FI, \( \mu \) depends on day-specific characteristics.

Day-specific expected travel times are determined using regression analysis. Separate regressions are run for each link and each quarter, as it is likely that the effects of the day-specific factors differ across road links and times of the day. The dependent variable in these regression are the travel times \( t_{t_i,q,l} \) during workdays of 2008. The explanatory variables are listed in Table 1. The weather-related variables are aggregated over the day. This is consistent with drivers usually not having precise information on weather conditions. Network-wide delays are used as an indicator of average congestion on all road links during a specific day. They are based on the sum of delays occurring between 6:00 and 20:15 on all road links during a specific day. The aggregate delays are weighted by the relevance of the link in the network, expressed by annual flow levels. These network-wide delays are included to reflect more predictable daily demand shocks (for instance due to school holidays), causing travel time variations in addition to those induced by the supply factors already identified. Including this variable implies that we assume that travelers will have the information that explains the network-wide delays on a specific day.

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Category</th>
<th>Indicator</th>
<th>Type of Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekdays</td>
<td>Tuesday</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wednesday</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thursday</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friday</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td>Season</td>
<td>July/August</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>December/January</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>maximum temp. &lt; 5°</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>maximum temp. &gt; 25°</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>duration rainfall &gt; 5 h</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>amount rainfall &gt; 10mm</td>
<td>dummy</td>
<td></td>
</tr>
<tr>
<td>Network</td>
<td>Network-wide Delays</td>
<td>continuous</td>
<td></td>
</tr>
</tbody>
</table>

The regressions carried out for each link and each quarter, using the explanatory variables of Table 1, yield an average R-squared of 0.16 (median: 0.12). Table 2 shows that the R-squared varies considerably across regressions. So, the 10th percentile value corresponds to an R-squared of only 0.03, whereas the 90th percentile value equals 0.34.

For the coefficient medians as well as the 10th and 90th percentiles we again refer to Table 2. This table also shows in how many percent of regressions a certain variable has shown to be significant at the 1%, 5% and 10% level, respectively. While all coefficients are significant at the 10% level in at least 10 % of all cases, this

12 The mean of the indicator is standardized to 1.
Table 2: Results auxiliary regressions for FI

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>Coefficient</th>
<th>Percentage of Coefficients significant at</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>10th Perc.</td>
</tr>
<tr>
<td>Constant</td>
<td>6.44</td>
<td>2.46</td>
</tr>
<tr>
<td>Tuesday</td>
<td>-0.02</td>
<td>-1.85</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-0.01</td>
<td>-1.81</td>
</tr>
<tr>
<td>Thursday</td>
<td>0.01</td>
<td>-1.43</td>
</tr>
<tr>
<td>Friday</td>
<td>-0.01</td>
<td>-1.41</td>
</tr>
<tr>
<td>July/August</td>
<td>-0.01</td>
<td>-0.96</td>
</tr>
<tr>
<td>December/January</td>
<td>-0.04</td>
<td>-0.81</td>
</tr>
<tr>
<td>max. temp. &lt; 5°</td>
<td>-0.04</td>
<td>-0.89</td>
</tr>
<tr>
<td>max. temp. &gt; 25°</td>
<td>-0.03</td>
<td>-0.84</td>
</tr>
<tr>
<td>duration rain &gt; 5 h</td>
<td>0.01</td>
<td>-0.78</td>
</tr>
<tr>
<td>amount rain &gt; 10mm</td>
<td>-0.03</td>
<td>-1.20</td>
</tr>
<tr>
<td>Network-wide Delays</td>
<td>0.61</td>
<td>0.04</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

percentage is considerably higher for weekday dummies (42–51%) and network-wide delays (72%).

Also coefficient values are quite heterogeneous across regressions. Only for network-wide delays, coefficients tend to have the same sign (positive). For the other coefficients, medians are fairly close to 0, and coefficients are positive or negative in a similar number of regressions. This results implies counterintuitive results such as decreasing travel times during adverse weather conditions in some regressions. It can be shown, that this is, to a major extent, the consequence of including the variable ‘network-wide delays’, as this variable partly also captures weather-related effects. Using the coefficients derived in the regressions, expected travel times can be calculated for each day. Based on these day-specific expected travel times (FI), the standard deviation is derived according to Eq. 2.2.

Mean delay

To include travel times in our analysis, the values first need to be made comparable between links that differ in lengths and free-flow times. We found that good results can be obtained when travel times are simply expressed in terms of delays (with respect to free-flow travel time).13 In order to calculate delays, link-specific free-flow times must be determined. We apply the numerical optimization method developed by Van Aerde and Rakha (1995).14 It fits a speed-flow curve to the traffic observations minimizing the (squared) errors with respect to speed, flow as well

13 Alternative procedures to standardize travel times have been investigated as well, however, produced less powerful results. One alternative is to scale all road stretches to a specific length (e.g. Kouwenhoven et al., 2005). A second alternative is to use average speeds instead of average travel times. And third, a ratio of travel times divided by free-flow time can be used to make travel times on different links comparable (Eliasson, 2006).

14 The according software (SPD-CAL) can be found under: http://filebox.vt.edu/users/hrakha/Software.htm#SpdCal
as density. The idea that a 5-minute delay on a road link of 5 km is different from a 5-minute delay on a road link of 30 km is controlled for by including the length of the road link (and various interaction terms with link length) as an explanatory variable in the regression analysis. We do not specify a model with time dynamics (e.g. by including temporally lagged terms of mean delay), since our model should be usable in policy analysis where data on these dynamics are usually not available. Finally, note that mean delay is calculated in the same way for both information regimes, FI and RI.

Link-specific variables

Besides mean travel time, we also include the number of lanes, the length of a link, free-flow speed and speed-at-capacity\(^{15}\) as covariates to explain travel time variability. These variables can be considered exogenous with respect to travel time variability. Speed-limits, or other traffic management measures, in contrast, exhibit a higher degree of endogeneity as they are frequently used to improve travel time variability. Therefore, they do not allow for an unbiased identification of a causal relationship regarding variability, which is essential for CBA. Other link-specific variables were tested but did not prove to be significant. In particular, we tested for the effect of the percentage of freight transport, the number of ramps per kilometer, whether a link is located in the West of the Netherlands, and whether the link contains a potential physical bottleneck (bridges, tunnels).

Average traffic flow is not included in the equation as it is double-valued with respect to travel time for a conventional backward-bending speed-flow relation. Given that flow is the product of density and speed, low flows can either occur under conditions close to free-flow conditions (high speeds and low densities), or under severely congested conditions (low speeds and high densities); see also Small and Verhoef (2007).

2.3 Empirical analysis of travel time variability

The empirical analysis consists of two parts. In the first part, models are estimated that explain variability by mean delay as well as certain road characteristics, but not by traffic regime; thus only using variables that are likely to be available to cost-benefit analysts. These models are named ‘non-traffic-regime based’ (NTR). From these regressions, we find that the relation between mean delay and variability is positive but non-linear. In the second part, we show that these non-linearities can be largely attributed to differences in the relative shares of the traffic regimes (free-flow, congested and hyper-congested) over the year. The corresponding models are summarized under the heading of ‘traffic-regime based’ (TR).

\(^{15}\)Speed-at-capacity is the speed at maximum capacity. Like free-flow speed, also speed-at-capacity is derived using the methodology developed by Van Aerde and Rakha (1995).
Chapter 2. Prediction of variability

**Non-traffic-regime-based model (NTR)**

Figure 4 plots the mean delay and standard deviation for all working days, for each quarter of an hour for each link.\(^{16}\) The graphs shows that the relation between mean delay and the standard deviation is positive. This intuitively makes sense. Deviations from mean travel times due to fluctuations in road supply and demand are likely to occur more often during times at which delays take place regularly (like morning and evening peaks). In addition, they tend to have a higher impact. Under free-flow conditions, moderate fluctuations in supply and demand conditions have a more limited effect on travel times, since there is still unused capacity.

Figure 4 also demonstrates the difference between RI and FI based variability. In the rough information (RI) case, standard deviations are considerably higher than in the fine information (FI) case. This does not come as a surprise, since part of the travel time variability that is due to weather, season, weekday and demand conditions has been removed in the FI case, resulting in lower variability. Furthermore, it is evident from the graphs that the relation between mean delay and travel time variability is not linear. We will examine this further in the regressions below.

![Figure 4: Relating mean delay and standard deviation](image)

The data points displayed in Figure 4 are used in the regression analyses. Travel time variability is the dependent variable, and mean delay (together with other variables) shows up as explanatory variable. The panel structure of this dataset, with the road links as cross-sectional dimension and the quarters as temporal dimension, is taken into account in the form of random effect models. Using a panel structure in the estimations leads to a better fit of the models. We do not use a fixed effects model, since it does not allow for the estimation of time-invariant variables. It was verified that the estimates of the (time-variant) variables in the fixed effects model, and in the random effects model are very close to each other.

\(^{16}\)The figure contains 8265 data points (145 links times 57 quarters).
Individual-level robust standard errors are computed. They are consistent when heteroskedasticity and autocorrelation are present.

In the first two regressions in Table 3, Model 1, we assume a linear relationship between standard deviation and mean delay. The RI-based model yields an R-squared of 0.75, and the FI-based model an R-squared of 0.67. The worse fit of the FI-based model probably results from using fitted instead of observed travel time data for the calculation of the FI-based variability. It can be verified that including more day-specific explanatory variables in the auxiliary regressions (in addition to the ones shown in Table 1) improves the fit of the FI-based models. As expected, we find a significantly lower coefficient for mean delay when the standard deviation is based on FI compared to RI.

The third and fourth regression in Table 3, Model 2, assume a non-linear relationship between standard deviation and mean delay. In addition to mean delay, also average speed, link-specific variables, and various interaction and exponential terms are added as regressors.

Table 3: Regression results (non-traffic-regime-based model)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RI</td>
<td>FI</td>
<td>RI</td>
<td>FI</td>
</tr>
<tr>
<td>Mean Delay (MD)</td>
<td>0.764***</td>
<td>0.578***</td>
<td>1.319***</td>
<td>1.191***</td>
</tr>
<tr>
<td>MD^2</td>
<td></td>
<td></td>
<td>−0.040***</td>
<td>−0.048***</td>
</tr>
<tr>
<td>MD^3</td>
<td></td>
<td></td>
<td>6.51 * 10^{-4}***</td>
<td>9.47 * 10^{-4}***</td>
</tr>
<tr>
<td>Mean Speed (MS)</td>
<td></td>
<td></td>
<td>0.187***</td>
<td>0.183***</td>
</tr>
<tr>
<td>MS^2</td>
<td></td>
<td></td>
<td>−1.28 * 10^{-3}***</td>
<td>−1.21 * 10^{-3}***</td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
<td>0.152***</td>
<td>0.140***</td>
</tr>
<tr>
<td>Length^2</td>
<td></td>
<td></td>
<td>−3.20 * 10^{-3}***</td>
<td>−2.84 * 10^{-3}***</td>
</tr>
<tr>
<td>MD*Length</td>
<td></td>
<td></td>
<td>−1.47 * 10^{-3}**</td>
<td>−4.12 * 10^{-3}**</td>
</tr>
<tr>
<td>Lanes</td>
<td>0.172*</td>
<td></td>
<td>0.147*</td>
<td></td>
</tr>
<tr>
<td>MD*Lanes</td>
<td>−0.053</td>
<td></td>
<td>−0.026</td>
<td></td>
</tr>
<tr>
<td>Free-Flow Speed</td>
<td>0.021**</td>
<td></td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>Speed-at-Capacity</td>
<td>0.018***</td>
<td></td>
<td>0.015***</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.451***</td>
<td>1.455***</td>
<td>−10.260***</td>
<td>−9.312***</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.751</td>
<td>0.667</td>
<td>0.856</td>
<td>0.810</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.126</td>
<td>1.052</td>
<td>0.889</td>
<td>0.824</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1

We find that the fit of Model 2 is considerably better compared to Model 1 for both RI and FI. The RMSE drops from 1.13 to 0.89 for RI and from 1.05 to 0.82 for FI, respectively. For both the RI and the FI case, we find that the derivative of the standard deviation with respect to mean delay decreases. The findings with respect to the other covariates are largely consistent between the FI and the RI regime. For both regimes, a shorter link is on average associated with lower variability. This result is reasonable, as variations in demand and road capacity tend to cause smaller deviations from average travel times on shorter roads compared to longer roads. Also the number of incidents is likely to be higher on longer roads. Nevertheless, variability increases less than proportionally with link length. This can be explained by ‘averaging-out’ effects. On longer road links, delays on one
part of the link can be compensated by free-flow conditions on another part of the link.

The marginal effect of the number of lanes on variability is a function of mean delay. For smaller delays, variability is positively correlated with the number of lanes. This may be caused by a larger number of accidents induced by more frequent lane changes. For longer delays, however, variability is negatively correlated with the number of lanes. It seems like another effect might become more important; if an incident happens, a higher number of lanes gives more possibilities for traffic to still pass, decreasing variability. Both free-flow speed as well as speed-at-capacity are correlated positively with variability.\textsuperscript{17} It can be shown that a positive relation exists between free-flow speed and speed-at-capacity. Roads with both low free-flow-speed and low speed-at-capacity tend to be roads that are easily congestible. Already at relatively high speeds, congestion and hyper-congestion, respectively, appear, implying also higher levels of variability.

**Traffic-regime based model (TR)**

Model 2 showed that the derivative of travel time variability with respect to mean delay decreases. Figure 5 illustrates that this pattern is closely related to traffic regimes. We distinguish between three traffic regimes: free-flow, congested and hyper-congested. Free-flow conditions are defined, somewhat arbitrarily, by a speed that is higher than a fraction 0.9 of the speed associated with free-flow travel time. Congested conditions are defined for lower speeds, but higher than speed-at-capacity. Hyper-congested conditions prevail if the speed drops below speed-at-capacity. Therefore, flow congestion is present in the free-flow and the congested regime, and (downstream) bottleneck congestion in the hyper-congested regime.\textsuperscript{18} In Figure 5, each data point shown in Figure 4 is attributed to the dominant traffic regime in terms of relative shares.

We find that the slope is steeper in the free-flow regime compared to the congested regime, and that in turn the slope of the congested regime is steeper than the slope in the hyper-congested regime. This gives a possible explanation for the non-linear distribution of data points found in Figure 4, where no traffic regimes are taken into account. Since the average number of observations with hyper-congestion as dominant regime is increasing in mean delay, the slope becomes less steep at higher levels of mean delay. Furthermore, the slopes in in Figure 5 are much closer to linear compared to the graphs in Figure 4.

Again we use (random effects) regressions to test these notions. In the first two estimations in Table 4, Model 3, mean delay is interacted with the three regimes. In Model 4, the additional regressors from Model 2 are added. Model 3 includes mean delay interacted with the relative shares of the three regimes among the 255 observations for each link-daytime combination. Comparing the fit of Model 3 to

\textsuperscript{17}Note that free-flow speed does not only enter the equation directly but also indirectly, via the calculation of mean speed.

\textsuperscript{18}Although results do not change qualitatively, the distinction into three rather than two regimes (congested vs. hyper-congested) leads to a substantial improvement in the fit of the models.
2.3. Empirical analysis of travel time variability

Figure 5: Traffic regimes

- RI (Free-flow, Congested, Hyper-Congested)
- FI (Free-flow, Congested, Hyper-Congested)

The NTR models, we find that it does not only have a better fit compared to Model 1, but also compared to Model 2. Compared to Model 2, the RMSE drops from 0.89 (0.82) to 0.78 (0.78) for the RI (FI) case. This is an indication of the power of the traffic regimes to actually explain and predict travel time variability. Consistent with Figure 5 we find that the lower the regime is ranked in terms of speed, the lower the derivative of the standard deviation with respect to mean delay.

A possible explanation of this finding is that for quarters during which on most days no congestion occurs, mean delays are strongly associated with (non-recurrent) fluctuations in demand and supply. In contrast, during time intervals in which hyper-congested conditions are dominant, the observed mean delays are likely to be strongly associated with recurrent demand patterns. Fluctuations in demand and supply are associated with relatively more variability compared to recurrent congestion patterns.

Clearly, fluctuations in demand and supply also occur under circumstances where hyper-congestion is the usual traffic condition. However, since we define the road links over several kilometers, from one highway intersection to another, and allow for highway entries and exits on the links, the links can include one or
more bottlenecks. Traffic conditions are therefore likely to be heterogeneous across a link. Upstream of the bottleneck(s), hyper-congested traffic conditions can be found, whereas downstream of the bottleneck, the congested and free-flow regimes prevail. As a consequence, delays can be expected to average out over a link. If an incident causes one of the upstream bottlenecks to have unusually long queues, the smaller outflow may imply shorter queues at downstream bottlenecks. Thus, demand and supply fluctuations are suggested to have a relatively smaller impact on travel times and travel time variability on a, normally, hyper-congested link than on a link that is usually subject to free-flow conditions. This gives a possible explanation for the comparatively lower coefficient of mean delay interacted with the share of hyper-congested observations.

In Model 4, the additional explanatory variables from Model 2 are added to Model 3. The coefficients in Table 4 give an indication that the traffic regimes capture the non-linearities in the relationship of the standard deviation and mean delay. Adding the explanatory variables of Model 2 to the regimes interacted with delay, leads to only minor improvements in terms of RMSE (compared to Model 3). For the RI-based models the RMSE decreases by 0.02, and for the FI-based models by 0.01. The relative improvement from adding additional regressors were considerably higher for the NTR models, where the relative declines were equal to 0.23 for both the RI and the FI case.

The coefficients of the relative shares of the traffic regimes interacted with mean delay are quite similar in Models 3 and 4, especially for the free-flow and the congested regime. Compared to Model 2, the additional regressors in Model 4 show, on average, a lower significance. Their signs, however, are in most cases equal to the ones observed in Model 2.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD*Share(free-flow)</td>
<td>2.291***</td>
<td>1.983***</td>
</tr>
<tr>
<td>MD*Share(congested)</td>
<td>1.365***</td>
<td>0.998***</td>
</tr>
<tr>
<td>MD*Share(hyper-cong.)</td>
<td>0.414***</td>
<td>0.287***</td>
</tr>
<tr>
<td>MD^2</td>
<td>0.034***</td>
<td>5.46 * 10^-3*</td>
</tr>
<tr>
<td>MD^3</td>
<td>-6.92 * 10^-4***</td>
<td>-7.44 * 10^-5*</td>
</tr>
<tr>
<td>Mean Speed (MS)</td>
<td>0.035***</td>
<td>0.073***</td>
</tr>
<tr>
<td>MS^2</td>
<td>-3.40 * 10^-4***</td>
<td>-5.42 * 10^-4***</td>
</tr>
<tr>
<td>Length</td>
<td>0.125***</td>
<td>0.120***</td>
</tr>
<tr>
<td>Length^2</td>
<td>-3.14 * 10^-3***</td>
<td>-2.85 * 10^-3***</td>
</tr>
<tr>
<td>MD*Length</td>
<td>-0.013***</td>
<td>-0.011***</td>
</tr>
<tr>
<td>Lanes</td>
<td>0.038</td>
<td>0.047</td>
</tr>
<tr>
<td>MD*Lanes</td>
<td>0.010</td>
<td>0.013*</td>
</tr>
<tr>
<td>Free-Flow Speed</td>
<td>-8.33 * 10^-3</td>
<td>-5.76 * 10^-3</td>
</tr>
<tr>
<td>Speed-at-Capacity</td>
<td>0.025***</td>
<td>0.018***</td>
</tr>
<tr>
<td>Constant</td>
<td>0.480***</td>
<td>0.589***</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.864</td>
<td>0.808</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.781</td>
<td>0.775</td>
</tr>
</tbody>
</table>

*** p < 0.01, ** p < 0.05, * p < 0.1
2.4 Implications for CBA

Model Choice

From the results in the previous section, we find that the TR models yield better predictions of variability than the NTR models, which renders them also the preferred models to use in CBA. Employing NTR models for predicting variability in CBA requires the cost-benefit analyst to know how the relative shares of the traffic regimes will change due to these projects. For instance, if a project involves a relative increase of downstream capacity, hyper-congestion is likely to occur less often after the implementation of the project.

However, in most cases, the changes in relative shares of the traffic regimes are difficult to predict. Then, Model 2 is the preferred alternative. It shows a significantly better fit than Model 1, which assumes a linear relation between mean delay and variability. And it includes variables which are usually available to cost-benefit analysts. Moreover, Figure 6 verifies that there is no structural bias in the estimation of Model 2. For most observations, the difference between fitted and actual values of variability is close to 0.

![Figure 6: Difference of fitted and actual values of variability](image)

An example based on Model 2

In this section, a numerical example based on Model 2 and its implications for CBA are discussed. Based on the coefficients derived in Table 3, predictions of variability are presented as a function of mean delay as well as road link length for both RI and FI. Road link length is added since it shows to have considerable impact on variability. For the other variables, values close to their sample means are used (number of lanes=2.5; free-flow speed=105 km/h; speed-at-capacity=80 km/h). Figure 7 shows 3D graphs of the variability predictions for the RI- and FI-based models. Tables 5 and 6 summarize the underlying numerical data of the predictions, as well as of the first and the second derivative of variability with respect to mean delay for three different link lengths (5, 10 and 20 km).
Table 5: Variability predictions (RI) based on Model 2

<table>
<thead>
<tr>
<th>Delay (in minutes)</th>
<th>5 km</th>
<th>10 km</th>
<th>20 km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( f )</td>
<td>( f' )</td>
<td>( f'' )</td>
</tr>
<tr>
<td>0.5</td>
<td>1.57</td>
<td>2.25</td>
<td>-2.55</td>
</tr>
<tr>
<td>1</td>
<td>2.44</td>
<td>1.34</td>
<td>-1.24</td>
</tr>
<tr>
<td>2</td>
<td>3.37</td>
<td>0.66</td>
<td>-0.34</td>
</tr>
<tr>
<td>4</td>
<td>4.34</td>
<td>0.41</td>
<td>-0.03</td>
</tr>
<tr>
<td>8</td>
<td>5.88</td>
<td>0.37</td>
<td>-0.01</td>
</tr>
<tr>
<td>16</td>
<td>8.39</td>
<td>0.27</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

\( f \): Absolute Prediction of the Standard Deviation (SD), \( f' \): First Derivative (\( dSD/dMD \)), \( f'' \): Second Derivative (\( d^2 SD/dMD^2 \))

Assumptions: Nr. of Lanes=2.5; Free-flow speed=105 km/h, Speed-at-capacity=80 km/h

Table 6: Variability predictions (FI) based on Model 2

<table>
<thead>
<tr>
<th>Delay (in minutes)</th>
<th>5 km</th>
<th>10 km</th>
<th>20 km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( f )</td>
<td>( f' )</td>
<td>( f'' )</td>
</tr>
<tr>
<td>0.5</td>
<td>1.45</td>
<td>1.94</td>
<td>-2.34</td>
</tr>
<tr>
<td>1</td>
<td>2.19</td>
<td>1.12</td>
<td>-1.13</td>
</tr>
<tr>
<td>2</td>
<td>2.93</td>
<td>0.50</td>
<td>-0.30</td>
</tr>
<tr>
<td>4</td>
<td>3.62</td>
<td>0.27</td>
<td>-0.03</td>
</tr>
<tr>
<td>8</td>
<td>4.60</td>
<td>0.22</td>
<td>-0.01</td>
</tr>
<tr>
<td>16</td>
<td>6.06</td>
<td>0.17</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

\( f \): Absolute Prediction of the Standard Deviation (SD), \( f' \): First Derivative (\( dSD/dMD \)), \( f'' \): Second Derivative (\( d^2 SD/dMD^2 \))

Assumptions: Nr. of Lanes=2.5; Free-flow speed=105 km/h, Speed-at-capacity=80 km/h

It is confirmed that the first derivative is a decreasing function of mean delay.\(^{19}\) Also it can be clearly shown that variability is larger for longer road links (however, less than proportionally). Comparing the RI and the FI case, we observe that the gap between the corresponding variability predictions increases in mean delay, both in absolute and relative terms.

As for CBA the absolute levels of travel time variability matter less than changes in travel time variability, the non-linear shape of the relationship between mean delay and standard deviation has an intriguing consequence: If delay is high, this leads to a smaller absolute change in travel time variability than if the initial delay had been smaller. Therefore the finding that the relation between mean delay and standard deviation is close to linear for somewhat longer delays, is especially interesting with respect to CBA. A linear relationship between the two variables is straightforward to use in CBA, as the rate of change in variability with respect to mean delay does not depend on mean delay. Tables 5 and 6 show that the second derivative becomes close to 0 at delays of around 4 to 8 minutes (depending on the length of the road link). Therefore, for longer delays it is valid to use a constant

\(^{19}\)The first derivative increases again slightly for longer mean delays. This can be considered a result from very few observations of these delays. Only 0.7% of the mean delays in the sample exceed 15 minutes.
2.4. Implications for CBA

Figure 7: Variability predictions based on Model 2

20Note that these numbers are quite close to the according value of 0.25 suggested by the CPB Netherlands Bureau for Applied Policy Analysis in their guidelines for cost-benefit analysis (Besseling et al., 2004)

ratio that directly connects costs of variability with costs of delays. An illustrative calculation can show the simplicity of this approach. In doing so, we assume a reliability ratio (=value of reliability divided by value of time) of 0.8, which is consistent with what has been found in previous research (e.g. Hamer et al., 2005). This ratio is then multiplied by the first derivatives given in Tables 5 and 6. For instance, for a delay of 8 minutes on a link of 10 km, the increase in the costs of reliability if costs of travel time change by one is equal to $0.8 \times 0.41 = 0.32$ under the assumption of RI, and $0.8 \times 0.25 = 0.20$ under the assumption of FI. In other words, the costs associated with travel time variability are 0.32 times as high as the costs associated with delays in RI case, and 0.20 in the FI case.

The use of FI- and RI-based variability

It remains to be discussed whether in CBA the extent of unreliability should be measured based on RI or FI. Ultimately, CBA is focused on calculating the costs related to variability. For this purpose it requires both predictions of the future development of variability and monetary values that drivers attach to changes in variability. These two factors should correspond to each other as closely as possible in order to obtain realistic results. The monetary values attached to variability are either derived from revealed preference (RP) or stated preference (SP) studies. RP valuations are derived from actual, observed choices of travelers in respect to departure time, route and/or mode. SP studies, in contrast, are based on hypothetical choice experiments, in which participants are asked to indicate their preferred travel alternative out of a given choice set. These differences in the underlying data source also imply differences with respect to the information that is assumed available to drivers.

Drivers participating in RP studies are likely to make their travel choices taking into account some information on day-specific factors, suggesting FI might be
more appropriate. However, in order to measure the trade-offs that drivers face, variability on the roads frequented by these drivers needs to be known. In most cases the variability is measured on basis of observed traffic variability, without taking into account day-specific factors. The same idea is reflected in the RI based value measure, which also includes variabilities that drivers are actually aware of. This makes it more adequate to use RI rather than FI in such cases, even though real behavior is based on FI.

SP studies usually state that travel times in the choice experiments are purely random. For instance, Tseng et al. (2009) compare travel time outcomes to the outcome of rolling a dice. If respondents follow this interpretation of pure randomness, it is more appropriate to measure variability in terms of FI rather than RI, as the FI-based variability widely eliminates variability due to factors usually known to drivers. Using the RI-based variability instead, results in an overestimation of the costs associated with variability. It is likely that in reality CBA tend to overestimate the costs of reliability due to this mismatch between the underlying assumptions of the valuation and the volume of variability. However, the extent of overestimation might be smaller than what would be expected at first sight. Namely, respondents might interpret the travel time distributions presented in SP studies as some distribution over all relevant (working) days, so that in reality they would have more information than just this distribution. This results in less uncertainty and, therefore, a lower estimated value of reliability. If this effect is large enough, the costs of variability might be represented more truthfully by an RI rather than an FI-based measure of variability.

If and to which extent this effect exists, needs to be judged on a case-by-case basis. It is likely that the information that respondents obtain during the SP experiment on the underlying assumptions of the travel time distribution plays a major role in this respect. The quality of the SP studies and their usefulness for CBA largely depends on whether a concise and understandable explanation on the background of travel time variability is given.

Finally note that FI is not uniquely defined. Here, FI is based on the factors cited in Table 1. These factors have been chosen as a representation of the information that most drivers have. In reality, the extent and quality of information does not only differ widely across drivers, but is also likely to be different for different times of the day or different road links. The definition of FI based variability can be adjusted for these aspects, if required.

2.5 Conclusions

We showed that travel time variability is well predictable using statistical models. Variability is correlated positively with delays, and the first derivative of travel time variability with respect to mean delay is decreasing. For longer delays the relation between these two variables is close to linear. The non-linear relation between mean delay and variability can be attributed to the distribution of traffic regimes. A one minute increase in mean delay is associated with a smaller increase in variability if the relative share of hyper-congested traffic observations is high.
during that time of the day. It is associated with a higher increase in variability, however, if free-flow observations are dominant.

Eliasson (2006) as well as Fosgerau et al. (2008) conclude that the costs related to travel time variability are around 10-15% of the costs related to changes in mean travel times. For longer delays we find a value of 19% for RI-based variability and 32% for FI-based variability (for a road link of 10 km). For shorter delays, costs tend to be higher. One possible explanation for the divergence with prior studies is that this study refers to highways, whereas Eliasson (2006) and Fosgerau et al. (2008) analyze urban roads. In urban networks travel time variability is already at a high level for no or very small delays (e.g. due to traffic lights). As a consequence, an increase in mean delay might increase travel time variability relatively less.

Our analysis does not take into account that travel times might be correlated between links. Although the presence of a spatial correlation structure in the model would be desirable, its setup is not trivial, and requires additional data. A spatial weight matrix would have to account for asymmetries of road network interruptions with respect to upstream and downstream links. Also possible substitution patterns among highway links, but also between highway links and the underlying road network should be included in some way. Finally, the extent of correlation between two links is probably not only determined by the distance between links but also by demand patterns. How to incorporate these elements in a satisfactory way is an interesting research topic by itself, but is beyond the scope of the current study.

Due to the different characteristics of the road links used in the analysis, the results obtained can be used to predict travel times on other Dutch highways and, with some caution, perhaps also on highways in other countries. The results should not be applied to roads that are considerably different from highways. However, the analysis done in this study can be conducted in a similar way based on travel time data gathered from other countries or road types.

One should also keep in mind that for the analysis in this chapter, road links were defined between two intersection points. This makes the point estimates considerably less applicable to stretches defined in other ways (e.g. for fixed lengths).

The analysis conducted here emphasizes the aggregate relationship between mean delay and travel time variability. However, the effect of traffic management measures on variability might differ considerably from their effect on travel times in terms of size and/or direction. Here, these effects are not distinguished. Future research may therefore focus on the impact of such traffic management measures on variability, not only because of scientific relevance, but also for practical reasons.
CHAPTER 3

Door-to-door travel times in departure time choice models: An approximation method using GPS data
3.1 Introduction

Congestion tends to be strongly related to the time of the day, mainly because commuters start and end work at similar hours. Therefore, in order to evaluate policies that mitigate congestion, models are needed that take these time of day related patterns into account. The dominant modeling approach is to assume that travelers optimize their departure time choice, trading off the costs of travel time and the costs of arriving earlier or later than their preferred arrival time. The underlying idea of this so-called scheduling model goes back to the early work of Vickrey (1969) and Small (1982). Empirical estimates of the drivers’ willingness to pay for reductions in travel times and schedule delays can be derived from data on hypothetical or actual scheduling choices, which are generally referred to as stated preference (SP) and revealed preference (RP) data, respectively. Scheduling models are usually estimated using discrete choice models, and therefore re-formulating the continuous departure time choice problem as a discrete problem with a finite number of alternative departure times.

While SP data have the advantage that the researcher is in control of the choice set and the attributes of the alternatives, the results may depend strongly on the design of the experiment, and be subject to hypothetical biases (e.g. Hensher, 2010). RP data, in contrast, are based on observed behavior, and therefore by definition do not suffer from hypothetical biases. Nevertheless, they are used in relatively few studies, mainly due to the requirements for detailed and high-quality data. For instance, there are only few real-life situations that incorporate time-of-day varying cost or reward components, which in turn are required for identifying the monetary valuations of time and schedule delays. Moreover, to allow for a precise parameter estimation, the monetary component must not be overly correlated with the other attributes of the utility function (e.g. travel time).

This study focuses yet on another data requirement evident in RP-based departure time choice models. It concerns travel time data, and more specifically door-to-door travel times. Clearly, the latter are the relevant travel times based on which travelers take their scheduling decisions. However, so far, many studies that estimate departure time choice models from RP data do not take into account door-to-door travel times, or only very rough measures thereof. This is mainly due to the poor availability of travel time data, especially outside the main road network. The problem of lacking travel time data is augmented by the fact that travel times need to be known not only for the chosen but also for the unchosen departure time alternatives. To our knowledge, no research has been undertaken that explicitly and systematically surveys the effects of using imprecise measures of (expected) travel times on the valuations of time and schedule delay.

We develop a model that allows for the (ex-post) approximation of driver-, day- and time-of-day-specific door-to-door travel times combining loop detector and GPS data. The travel times resulting from this model are then used to determine the

---

21 This chapter is based on Peer, Knockaert, et al. (2011).
22 Overviews of empirical research can for instance be found in Brownstone and Small (2005), Tseng et al. (2005) and Li et al. (2010).
attributes of the departure time alternatives. We show that less precise travel time definitions that ignore the spatial and temporal variation of speeds on some parts of the door-to-door trip may lead to biases in the value of time. A straightforward reason is that ignoring the spatial variation of speeds typically leads an imprecise measurement of travel times, which in turn affects the travel time coefficient to be biased downwards. However, there is a second, more subtle reason as well: Not accounting for the temporal variation of speeds means that the tendency of speeds to be correlated across links is ignored. In most situations a positive correlation is found, for instance, because during the morning and evening peaks, speeds are relatively low on most parts of the network (e.g. De Fabritiis et al., 2008; Hackney et al., 2007). If these correlations are ignored, the differences between peak and off-peak travel times is underestimated, and as a result, the (absolute) time coefficient, and as a consequence also the value of time will be overestimated.

The RP data used in this study were gathered during a real-life peak avoidance experiment (Spitsmijden in Dutch). Participants were eligible for a monetary reward if they avoided traveling on a specific highway link (which we refer to as ‘C1–C2’ as it is confined by two cameras, C1 and C2) during the morning peak. In a first phase of the experiment, 230 participants got equipped with GPS devices yielding door-to-door travel time measurements. The second phase of the experiment included approximately 2000 participants and was conducted over a period of multiple months. During that second phase only travel times along the C1–C2 link were measured. For the choice models estimated in this chapter, we only consider the observed scheduling behavior of (a subgroup of) participants of the second phase. However, we will use the GPS data gathered during the first phase to approximate door-to-door travel times (potentially) faced by participants of the second phase. Figure 8 shows a schematic map of the experimental setup.

Figure 8: Schematic map of the experimental setup
Chapter 3. Door-to-door travel times

In order to approximate driver-, day and time-of-day-specific door-to-door travel times, we use geographically weighted regression (GWR). Our model explicitly takes into account travel time correlations across trip parts, in particular between the links for which only infrequent GPS observations are available (home-C1, C2–work) and the link along which travel times are observed continuously by means of loop detectors (C1–C2). GWR allows for spatial heterogeneity in the correlation pattern. For a given speed on the C1-C2 link, predicted speeds along the home–C1 (C2–work) links are more similar between home (work) locations that are located closer to each other, but can differ more strongly between locations that are further apart. Figure 9 underlines the relevance of this approach. It plots the starting (home) and end (work) locations across space. The color scale indicates the average speed measured from the plotted home locations to C1, and from C2 to the work locations. Even without making a distinction between different times of the day, and therefore recurrent congestion levels, the spatial variation of speeds is substantial. A clearly visible pattern is that home locations closer to highways face relatively high speeds. The latter can most likely be explained by the fact that drivers with home (work) locations situated close to the highway cover a relatively higher share of their commute on the highway compared to drivers with a home (work) location further away from the highway. Even though highways may be congested during peak hours, they allow for relatively high travel speeds during off-peak hours.

Figure 9: Speed observations

While we decided to use a combination of loop-detector and GPS data to approximate door-to-door travel times, also reported travel times or travel times

Note that many of these ‘home’ and ‘work’ locations are located along the highway (the East-West connection passing through C1 and C2, along which many of the data points can be found). This finding may be attributable to drivers who only switched on their GPS device after departing from home, or switched it off before arriving at work. Another explanation is that in order to warrant the quality of the GPS data, we define various criteria that need to be fulfilled for recognizing specific GPS observations as being part of a trip through C1 and C2. As a consequence, we may conservatively assume a trip to start at a location closer to C1, or to end at a location closer to C2 than it actually did in reality, resulting in the possibility that the starting and/or end location of a trip are located along the highway.
generated from network models could have been used for this purpose. However, both data sources have in common that they are usually not available at a sufficiently fine temporal scale required to represent the trade-offs that drivers face between departure time choice alternatives. Furthermore, reported and network travel times are usually not day-specific. Generic travel times, however, do not allow for modeling travel time variability or day-specific travel time expectations, whereas both of these features are present and show to be relevant in the models estimated in this study.

The structure of this chapter is as follows. Section 3.2 contains a review of past literature on RP-based departure time choice models. In Section 3.3, we describe the GWR methodology used for the approximation of door-to-door travel times. Section 3.4 provides an overview of the data used in the GWR model, and the results of the GWR models are then shown in Section 3.5. Section 3.6 discusses the set-up of the departure time choice models, the respective data, as well as alternative model specifications. Section 3.7 presents the results of the choice models, and Section 8 concludes.

### 3.2 Related literature

A limited number of studies exist that estimate departure time choice models using RP data, either as only data source, or in combination with SP data. Most of the older RP-based studies on departure time choice behavior assume that travel times do not vary between days, hence ignoring travel time variability. For instance, Abkowitz (1981), Cosslett (1977) and Small (1982, 1987) use a dataset from the San Francisco Bay Area for the Urban Travel Demand Forecasting Project (UTDFP). By means of interpolation between peak and off-peak network values, driver-specific door-to-door travel times are calculated for 12 alternative travel times, which do not vary across days. A similar strategy is employed by McCafferty and Hall (1982) and Bhat (1998a,b).

More recent studies on RP-based departure time choice models account for travel time variability. While this renders the models more realistic, it also raises the requirements for travel time data. Either day-specific travel times data need to be gathered or a good way of approximating them needs to be found. As a consequence, only few of these studies take into account door-to-door travel times. One example is Börjesson (2008) who combines day-independent traffic simulation data and enriches them with camera data to generate travel times that vary by time of the day. However, she assumes that travel times on secondary roads do not differ between days, arguing that these are hardly ever congested. She does not comment on the possibility of varying travel times due to general factors such as weather conditions.

---

24 In general, caution should prevail with the use of reported travel times as these might be systematically biased. For example, we found that travel times reported by participants of the Spitsmijden experiment were on average overestimated with a factor 1.5 (see Chapter 4 of this thesis).

25 E.g. Börjesson (2008); Brownstone and Small (2005); Ghosh (2001); Small et al. (2005)
Lam and Small (2001) observe travel times directly along a 10 mile corridor for their joint route and departure time choice model, but not for the remaining links of the door-to-door trips. They estimate two models: The first model only takes into account travel times along this corridor, whereas the second model assumes that drivers face the same speeds as observed along the corridor also on the rest of their travel (except for a 5 mile access link, which is assumed to be uncongested). They find that the value of time decreases by 50% if they move from the first specification to the second specification. It can be expected that the second specification inflates the assumed travel time differences between travel moments, making the result in itself less surprising. Besides the first model of Lam and Small (2001), also Knockaert et al. (2009) and Tseng et al. (2011) only account for travel times on the specific highway link along which the regarding experiment took place, without testing alternative travel time specifications.

To our knowledge this study is the first one that uses GPS data for defining the attribute values of the choice alternatives in a scheduling model. However, in various papers GPS data have been used to analyze other types of travel choices, such as route choice behavior (e.g. Bierlaire et al., 2010; Carrion and Levinson, 2012; Ramos et al., 2012). Moreover, Fifer et al. (2011) use GPS data for defining realistic attribute values in their SP-based study on crash-risk reduction.

Independent of the context of choice modeling, the use of GPS data for travel time measurements has been discussed in various recent studies. Many of them investigate the problems arising from a limited availability of GPS data, which can manifest itself in two different ways: First, the sampling frequency of GPS observations for a given trip may be low, such that the spacing between two subsequent GPS observations from the same device becomes rather wide. Possible causes are external conditions such as tall buildings or tunnels (Chen et al., 2005) as well as the technical specifications of the GPS device (Jenelius et al., 2012; Westgate et al., 2011). Naturally, the consequence of low-frequency sampling is that it becomes difficult to infer which route a driver has followed – an information that is often considered desirable or even necessary for the computation of GPS-based travel times (depending on the methodology used). Various methods to deal with this so-called map matching problem have therefore been developed (e.g. Bierlaire et al., 2010; Quddus et al., 2007).

The second type of GPS data sparsity is present when the number of vehicles equipped with GPS devices is relatively low compared to the overall network volume. The respective literature generally suggests that GPS data from 4-6% of total vehicles in a given network are necessary for travel time measurements to be fairly precise (El Esawey and Sayed, 2012). Wang et al. (2012) emphasize that this percentage depends on the acceptable accuracy of the measurements as well as the sampling frequency of the GPS devices.

Different solutions have been presented in the literature to deal with the sparsity of GPS data due to limited coverage with GPS devices. One is to combine the GPS data with other travel time data sources, such as loop detector data, using Kalman filtering (Chu et al., 2005) or Bayesian methods (Choi and Chung, 2002). Another one is to use available travel time information from nearby network links,
3.2. Related literature

Taking into account that link travel times tend to be correlated. A distinction can then be made between models that emphasize temporal correlation patterns and those that focus on spatial correlation patterns. The ones belonging to the first category usually rely strongly on time-series data. Examples are models with Markov properties (e.g. Castro et al., 2012; Herring et al., 2010) or space–time autoregressive integrated moving average (STARIMA) models (e.g. Kamarianakis and Prastacos, 2005). Models that emphasize the spatial correlation pattern often use geostatistical kriging models (Aultman-Hall and Du, 2006; El Esawey and Sayed, 2012; Miura, 2010; Zou et al., 2012) or local regression models (e.g Hackney et al., 2007; Idé and Kato, 2009).

The problems related to GPS data sparsity are expected to decrease over the next decade, as more and more people adopt GPS-enabled smartphones and navigation devices, and technical advancements allow for higher sampling frequencies. Also various issues related to privacy, data ownership, standardization and data processing methodology are likely to be solved in the coming years (e.g. Krause et al., 2008; Leduc, 2008). However, the methods presented above are expected to remain relevant, since they provide generic methods of how GPS data can be used to measure as well as predict travel times.

The research contained in this chapter fits well into this recently emerging stream of literature on GPS data. It develops a method how to deal with second type of sparsity, which occurs if the number GPS-equipped vehicles is fairly low. The first kind of GPS data sparsity is not evident in this study, as GPS observations are sampled at fairly high frequency (almost every second). To deal with the very limited number of GPS-equipped vehicles, we employ both strategies presented above, the combination with other data sources as well as the estimation of models that account for travel correlations across links. So, first we combine GPS data with data from loop detectors. However, in contrast to the papers of Chu et al. (2005) and Choi and Chung (2002), we use GPS data for one part of the network and loop detector data on another part of the network, rather than combining both data sources on a given link. Second, we take into account travel time correlations across links. As the GPS observations are very sparse in our dataset, any models requiring time series data are infeasible. Among those models that focus on the spatial correlation across links, we chose to use geographically weighted regression (GWR), which is a spatial version of local regression, and goes back to the seminal papers of Stone (1977) and Cleveland (1979).

GWR is very flexible in its use and specification, and often easier to calibrate compared to models that employ kriging (Harris et al., 2010). In defining the correlation neighborhoods, we choose to use Euclidean distances rather than network distances. Some previous studies (e.g. Hackney et al., 2007; Zou et al., 2012) have defined neighborhoods based on (undirected) network distances, and found that these perform better in predicting speeds compared to the Euclidean distances. However, using the more simple Euclidean neighborhood definition

Note that these models are not per se specific to GPS data, but can also be used for other traffic data sources. For instance Kamarianakis and Prastacos (2005) apply their model to loop-detector data.
seems to be justifiable in our case, since we are not interested in trips throughout the entire network but only in those leading from the home locations to camera C1, and from camera C2 to the work locations. Due to the limited number of routes leading to these destinations (e.g. a limited number of highway entries and exits), two persons with home (work) locations close to each other in terms of Euclidean distance are likely to choose a similar route towards C1 (work) anyways, reducing the need to model network distance explicitly. For a similar reason, our method does not involve any map-matching on the home–C1 and C2–work links, meaning that we only consider speeds and distances rather than the actually chosen route along the GPS-observed links in our analysis. Note that therefore in this study a ‘link’ refers to an abstract connection between a home location and C1, or between C2 and a work location, without pinpointing this link to a specific route. The method is therefore also well applicable in situation where GPS data are sparse due to low sampling frequency, and map-matching would be difficult.

### 3.3 Geographically Weighted Regression

**Methodology**

The basic idea is to use geographically weighted regression (GWR) to explain speeds on those parts of the network for which infrequent GPS observations are available (the home–C1 and C2–work links) by the speeds along the C1–C2 link where continuous speed measurements are available from loop detectors. The estimated parameters that describe the relationship between the between the GPS-measured speeds and the continuously measured speeds are then used to approximate time-of-day- and day-specific speeds between home locations and C1, and C2 and the work locations by means of interpolation (if the regarding home or work location was not in the original sample).

GWR is a form of local regression that aims at the analysis of spatial data. Unlike the ordinary least squares (OLS) model, local regression models do not yield a global set of coefficients but local coefficients that result from fitting models to localized subsets of the data. Thus, in the GWR model coefficients can differ over space (in our application across home and work locations). The basic idea is that location specific coefficients tend to be more similar for (home or work) locations situated close to each other, and to diverge more strongly for locations further apart from each other. GWR has the advantageous property that all spatial information with respect to the start and ending points of the trips is fully used. No aggregation into areas (e.g. ZIP code areas) is required, and full advantage is taken of the spatial variation of the data.

The GWR model implies that for each observation a weighted least squares regression model is estimated, using a spatial weight matrix (e.g. Brunsdon et al., 1998; Charlton and Fotheringham, 2009; Fan and Gijbels, 1996). The weight matrix is determined by the relative geographic locations, implying that higher weights are attached to observations that are closer to the reference observation. The coefficients estimated from this model can then be used to predict speeds on the
GPS observed links for different days and times of the day. By means of spatial interpolation, we can also utilize them to approximate speeds for (home and work) locations for which no GPS observations are available.

The estimator of the locally weighted least squares model at the reference location \( u \) is given by:

\[
\hat{\lambda}(u) = [X^T W(u) X]^{-1} X^T W(u) y, \tag{3.1}
\]

where \( \hat{\lambda}(u) \) is the local parameter estimate, \( y \) is the \( N \times 1 \) column vector including the values of the dependent variable (the speeds on the GPS observed links) for all observations \( i = 1, \ldots, N \), and \( X \) is the \( N \times p \) matrix of covariates (for \( p \) covariates: most importantly, the speeds on the continuously observed link). \( W(u) \) denotes the spatial weight matrix. Its diagonal elements correspond to the weights between the reference location \( u \) and all other observations.

\[
W(u) = \begin{pmatrix}
    w_1(u) & 0 & \cdots & 0 \\
    0 & w_2(u) & \cdots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & w_n(u)
\end{pmatrix} \tag{3.2}
\]

The weights at location \( u \) with respect to observation \( i \), \( w_i(u) \), are determined by the kernel function \( f \), which takes the Euclidean distance \( d_i(u) \) between the locations associated with \( u \) and \( i \) and the bandwidth parameter \( h \), as inputs. The kernel function yields a weight that is decreasing in the distance between \( u \) and \( i \):

\[
w_i(u) = f(d_i(u), h) \tag{3.3}
\]

The functional form of \( f(d_i(u), h) \) is assumed to be Gaussian\(^{27}\).

\[
w_i(u) = \exp \left[ -\frac{1}{2} \left( \frac{d_i(u)}{h} \right)^2 \right]. \tag{3.4}
\]

The bandwidth \( h \) is a fixed, unknown parameter. It determines the distance decay of the weights and is expressed in the same units as the distances \( d_i(u) \). As the bandwidth increases, the weights become uniform and the local GWR model approaches the global OLS model. The optimal bandwidth trades off model fit (resulting from large bandwidths) and variance (resulting from small bandwidths). We use cross-validation (Bowman et al., 1998; Wheeler and Páez, 2010) to determine the optimal bandwidth. Cross validation defines that bandwidth as optimal that minimizes the root mean squared prediction errors (RMSPE), using a subset of the data for prediction. In the standard application of this method, the reference observation \( i \) is left out from the estimation in order to prevent a perfect fit of the

---

\(^{27}\)The scale of the weights does not matter as \( [X^T (\alpha W(u)) X]^{-1} X^T (\alpha W(u)) y = 1/\alpha [X^T W(u) X]^{-1} X^T a W(u) y = [X^T W(u) X]^{-1} X^T W(u) y. \)
model. However, in our application we leave out all observations that are attributed to the same driver $z$ as observation $i$:

$$\hat{h} = \arg\min_h \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_{-z(i)}(h))^2}$$  \hspace{1cm} (3.5)

This ‘leave-one-driver-out’ cross-validation criterion is adopted to account for the panel structure of the dataset. Observations attributable to the same driver tend to be clustered spatially as well as temporally (i.e. a driver often has similar departure times across working days)\(^{28}\). If only the reference observation $i$ is left out, the optimal bandwidth is likely to be too small, yielding very precise predictions for all trips undertaken by the corresponding driver $z(i)$, however, not for other drivers with similar start or end locations, who tend to depart at different times of the day. Since we intend to use the coefficients for out-of-sample predictions, it is crucial that the model is able to predict travel times well for all times of the day. We do not account for the panel setup directly in the estimation, for instance, by using fixed effects, since start and end points of the door-to-door trips for a given driver can differ between days.\(^{29}\)Introducing person-specific effects is therefore not useful as they might capture location-specific rather than driver-specific effects. Furthermore, we expect speed differences between individual drivers to be negligible, as speed-limits are enforced rather strictly and many links are congested during most morning peaks, leaving little room for speed differences between drivers.

**Functional form of the local specification**

In order to render the estimation results easily applicable for out-of-sample prediction of speeds for home and work locations that are not in the estimation sample, two models are estimated: One that measures the correlation between speeds on the home–C1 links and the C1–C2 link, and a second that measures the correlation between speeds on the C1–C2 link and the C2–work links.

We first need to establish the relationship between the time when the GPS-observed trip had started (at home (for home–C1 trips) and C2 (for C2–work trips), respectively), $\tilde{t}$, and the time when camera C1 is passed (which defines the C1–C2 speed used as an explanatory variable), $t$. The corresponding speeds on the home–C1 (C2—work) link for a specific home (work) location $u$ are then denoted by $v^g_{\tilde{t}}(u)$, and the speeds on the C1–C2 link by $v^{cont}_{t}$. Moreover, we denote the passage time at C1 that is defined as applicable for a given $\tilde{t}$ by $\bar{t}$.

In the home–C1 case, for an observed departure time from home, $\tilde{t}$, we use the C1–C2 speed also at time $\tilde{t}$ as explanatory variable ($\tilde{t} := \tilde{t}$). If we defined the C1–C2 speed upon arrival at C1, recursiveness would result, since then the passage time at C1 would be determined by the home–C1 speeds but at the same time the

\(^{28}\)See for instance Chapter 5 of this thesis.

\(^{29}\)These differences can result for multiple reasons: Variations in the parking location of the car, switching on (or off) the GPS device during the trip, or trip and GPS signal interruptions. See also Footnote 23.
home–C1 speeds would determine the passage time at C1. Using the definition of \( \tilde{t} := \tilde{t} \), speeds on the home–C1 link determine the passage time at C1 but not the other way round. In the C2–work case, recursiveness does not occur since the C1–C2 link is passed before the GPS-observed link. We therefore define \( \bar{t} \) as the passage time at C1 that results in the observed passage time at C2, \( \tilde{t} \), given C1–C2 speeds \( v_{cont}^t \):

\[
\bar{t} := \tilde{t} \quad \text{for home–C1 links}
\]

\[
\bar{t} := \arg \min_t |\tilde{t} - \left( t + \frac{\text{distance}_{C1-C2}}{v_{cont}^t} \right)| \quad \text{for C2–work links} \quad (3.6)
\]

The (location-specific) model that relates the GPS observed speeds to the C1–C2 speeds is assumed to have a local linear form:

\[
v_{gps}^t(u) = \lambda_0(u) + \lambda_1(u)v_{cont}^t \quad (3.7)
\]

Eq. 3.7 establishes a simple direct relation between the speeds on the two links. It proved to perform better in out-of-sample prediction than more complicated model structures. So, we tested models that include departure time as explanatory variable both in linear\(^{30}\) as well as nonlinear\(^{31}\) specifications. While the overall fit is comparable to the fit of the model based on Eq. 3.7, the more complicated functions tend to yield large outliers at the begin and the end of the peak. Furthermore, we also tested a model where the weight matrix was not only location but also time specific. While this method leads to small improvements in the predictive power of the models\(^{32}\), it has the disadvantage that a different set of coefficients is valid for different times of the day, making the model less general. Eq. 3.7, on the other hand, is applicable to all times of the day. Moreover, it does not require an assumption on the shape of the peak. Note that weekday- and weather-specific influences on door-to-door travel times, which tend to affect the entire network, are captured through by the observed speeds on the continuously observed link and the correlations of the home–C1 and C2–work speeds therewith.

3.4 Data

Experiment

The data used in this study were gathered from a peak avoidance experiment taking place in the Netherlands (in Dutch it is referred to as Spitsmijden experiment). Participants could receive a daily reward of 4 Euro for avoiding a specific highway link of 9.21 km length (C1–C2) during morning peak hours (6:30–9:30).\(^{33}\) The

\(^{30}\)E.g. \( v_{gps}^t(u) = \lambda_0(u) + \lambda_1(u)v_{cont}^t + \beta_3(u)\tilde{t} + \beta_4(u)\tilde{t}^2 \)

\(^{31}\)E.g. \( v_{gps}^t(u) = \lambda_1(u) - \beta_2(u)(1 + \beta_3(u)v_{cont}^t)e^{-(\beta_4(u)\tilde{t} - \beta_5)^2} + \beta_6v_{cont}^t \)

\(^{32}\)The root mean squared error (RMSE) decreases on average by 0.45 km/h for the home–C1 model, and by 1.18 km/h for the C2–work model.

\(^{33}\)Drivers participating in the experiment could not earn a reward on weekends or school vacation days, but also not if they had already exceeded the maximum number of days per week for which a reward could be obtained. This maximum is driver specific, and equal to the average number of weekly trips undertaken during the pre-experimental period (reference behavior).
experiment was set up with the goal to mitigate the negative effects of roadworks undertaken along this link during the time of the experiment.

Car drivers who were observed to pass the C1–C2 link multiple times a week were invited to participate in the experiment. In the first phase (11/2008–4/2009), 230 drivers obtained customized smartphones, also named Rabomobiel. These phones were equipped with a GPS receiver and transmitted information about the location of the phone to a central database. The GPS data gathered during this first phase are then used to approximate door-to-door travel times for the second phase of the experiment (09/2009-12/2009), where no GPS data were collected any longer and therefore travel time measurements were thus only available for the C1–C2 link. The number of participants in the second phase was close to 2000, of which approximately one third are considered in the analyses presented here. These are the ones for whom the preferred arrival times at work are known from a questionnaire. An in-depth description of the Spitsmijden experiment can be found in Appendix A of this dissertation.

To compute the attribute values of the departure time choice alternatives faced by the participants of the Spitsmijden experiment, the GPS-observed speeds for the home–C1 and C2–work links and the speeds derived from loop detectors along the C1–C2 link are used in the GWR estimations. The following two sections elaborate on the features of these travel time data sources.

**GPS-based speed measurements**

In their raw form, GPS measurements are data points with a location, speed and time stamp attached to them. From sequential data points from a given device, speeds and distances can be calculated.\(^{34}\) We take into account GPS-observed trips that pass either through camera location C1, C2 or both. We construct separate datasets for trips observed between home and C1, and C2 and work. A trip is defined to start when the speed at which the GPS receiver moves exceeds 5 km/h, and to end when the speed drops to less than 5 km/h for at least 10 minutes. To determine the average speed, the distance covered between the start and the end time needs to be known. The distance is defined as the sum of the distances between the GPS observations belonging to the same trip.\(^ {35}\) We only consider trips along the home–C1 and C2–work links that were observed during the morning period (between 5:00 and 11:00 a.m.) and on working days, corresponding to our analysis of departure time choices in the morning and on weekdays. We exclude

---

\(^{34}\)Also the chosen routes can be inferred from these data. However, as previously stated, we do not make use of any map-matching procedures in this study.

\(^{35}\)Since location measurements using GPS can be somewhat imprecise, it is possible that the GPS device transmits different locations although the vehicle is standing still. If these distances are added, the overall trip distance might be overestimated, resulting in a structural downward bias of travel times (e.g. Zito et al., 1995). For this reason, if speed drops below 5 km/h, we only take into account the distance between the location at which speed dropped below 5 km/h and the location at which the vehicle resumed a speed of above 5 km/h. It is reassuring that this approach yields very similar results compared to defining distance based on a ‘fastest network distance’ (retrieved from openrouteservice.org).
trips that result in speeds lower than 30 km/h or above 120 km/h, and those for which no close neighboring observations (with similar start or end locations)\textsuperscript{36} are available. Also, the largest distance between 2 GPS measurements (for a given trip) must not exceed 2 km.

Table 9 gives an overview of the descriptive statistics of the data for the home–C1 and C2–work links. Trips along the home–C1 links tend to be longer as well as faster in terms of speed compared to trips along the C2–work links. This finding can be attributed to the fact that, on average, a larger share of the home–C1 links consists of highways rather than local roads. Due to the higher average speeds on the home–C1 link, the mean and maximum distance between two subsequent GPS measurements are on average higher on the home–C1 link. Also, mean speeds on the C1–C2 link (after passing the home–C1 or before passing the C2–work link) are shown. They are similar to the speeds on the GPS-observed links but significantly lower than the free-flow speed of 100 km/h (i.e. the speed-limit). In contrast to the standard deviation of speeds for the GPS-observed links, which is also a consequence of heterogeneity between driver-specific links and the corresponding free-flow speeds, the standard deviation of speeds for the C1–C2 link can almost exclusively be attributed to congestion (as the C1–C2 link is defined identically for all drivers).

Moreover, Table 9 shows that departure (arrival) times as well as home (work) locations differ\textsuperscript{37} much more strongly across drivers than for a given driver (across trips), providing a good argument for the use of the ‘leave-one-driver-out’ cross validation criterion for bandwidth selection, as discussed in Section 3.3. Moreover, Table 9 demonstrates the importance of using door-to-door travel times. While the length of the C1–C2 link is 9.21 km, the average length of the home–C1 and C2–work links are equal to 30.91 km and 16.34 km, respectively. Hence, the omission of these parts of the trip that cannot be observed by means of continuous speed measurements would result in the omission of a very substantial part of the door-to-door trips.

**Loop-detector based speed measurements**

The GPS observed speeds discussed in the previous section are matched to C1–C2 speeds according to the temporal relations between the starting time of the home–C1 (C2–work) trip and the passage times on the C1–C2 link established in Eq. 3.6. The speeds on the C1–C2 link are available for any time of the day from a dense network of loop detectors. These speeds are aggregated in time (15-minute intervals) and space (from the single detectors towards the entire link), employing the trajectory method. A general description of this method can for instance be found in Van Lint (2010), whereas the exact approach used in this research is described in Modelit (2009). It can be shown that the speeds on the C1–C2 link

\textsuperscript{36}Less than 10 observations within a perimeter of 5 km of the reference location.

\textsuperscript{37}Differences between ‘home locations’ for a given driver might for instance occur because of different parking locations or because the GPS device has been switched on at different moments on different days.
Table 7: Descriptives GPS

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Home–C1</th>
<th>C2–Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of trips</td>
<td>2896</td>
<td>5896</td>
</tr>
<tr>
<td>Nr. of drivers</td>
<td>91</td>
<td>163</td>
</tr>
<tr>
<td>Nr. of days</td>
<td>144</td>
<td>144</td>
</tr>
<tr>
<td>Mean travel time (per trip)</td>
<td>23.13 min</td>
<td>15.14 min</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>14.19 min</td>
<td>5.83 min</td>
</tr>
<tr>
<td>Mean distance (per trip)</td>
<td>30.91 km</td>
<td>16.34 km</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>19.36 km</td>
<td>3.80 km</td>
</tr>
<tr>
<td>Mean speed (per trip)</td>
<td>78.69 km/h</td>
<td>69.29 km/h</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>21.88 km/h</td>
<td>16.65 km/h</td>
</tr>
<tr>
<td>Mean speed on C1–C2 (per trip)</td>
<td>80.02 km/h</td>
<td>74.46 km/h</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>20.30 km/h</td>
<td>19.21 km/h</td>
</tr>
<tr>
<td>Mean number of GPS measurements (per trip)</td>
<td>1288</td>
<td>784</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>775.26</td>
<td>252.64</td>
</tr>
<tr>
<td>Mean time btw. 2 GPS measurements</td>
<td>1.09 s</td>
<td>1.15 s</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>0.14 s</td>
<td>0.15 s</td>
</tr>
<tr>
<td>Mean distance btw. 2 GPS measurements</td>
<td>23.36 m</td>
<td>21.61 m</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>5.94 m</td>
<td>4.00</td>
</tr>
<tr>
<td>Maximum distance btw. 2 GPS measurements (on a given trip)</td>
<td>85.14 m</td>
<td>115.43 m</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>119.69</td>
<td>198.37</td>
</tr>
<tr>
<td>Overall st. dev. dep. (arr.) time at home/work</td>
<td>90.60 min</td>
<td>87.60 min</td>
</tr>
<tr>
<td>Median driver-specific st. dev. dep. (arr.) time at home/work</td>
<td>32.20 min</td>
<td>34.03 min</td>
</tr>
<tr>
<td>Overall median distance btw. home (work) locations</td>
<td>16.37 km</td>
<td>4.24 km</td>
</tr>
<tr>
<td>Median distance btw. driver-specific home (work) locations</td>
<td>1.43 km</td>
<td>0.35 km</td>
</tr>
</tbody>
</table>

do not differ significantly between the two project phases (11/2008–4/2009 and 09/2009-12/2009), providing confidence that we can safely use the relation derived between the GPS-based speeds and the camera observations during the first phase to make travel time predictions during the second phase.

3.5 Results: Geographically Weighted Regression

In this section we present the results obtained from the GWR model. We first determine the optimal bandwidths using the ‘leave-one driver-out’ cross-validation approach. Figure 10 shows that a minimum exists with respect to the RMSPE for both the home–C1 and the C2–work links. The optimal bandwidths are determined accordingly: 2.64 km for the home–C1 model and 1.05 km for C2–work model. Figure 11 shows the corresponding kernel weights attached to the distance between reference location \( u \) and another observation \( i \): The weights are subject to a steeper distance decay for the case of the C2–work links than for the home–C1 links. This is likely be the joint result of the relatively densely located work locations and as a consequence a lower variance in local traffic conditions, as well as the higher number of observations for the C2–work case. A comparison of the optimal GWR model to an OLS model with generic parameter values reveals that indeed the root mean squared error (RMSE) decreases and the R-squared increases substantially.
3.5. Results: Geographically Weighted Regression

if one moves from OLS to GWR (Table 8). While these results hold for both the home–C1 and the C2–work links, they are stronger for the home–C1 links. This confirms that spatial heterogeneity plays a larger role for the home–C1 link.

Figure 12.1 shows a map of the area within which the home and work locations are situated. Figures 12.2 and 12.3 show the spatial pattern that results if speeds are predicted using the location-specific coefficients derived from the GWR model, assuming a C1–C2 speed of 50 or 100 km/h, respectively. These figures can be interpreted as a peak (50 km/h) and an off-peak (100 km/h) scenario. As expected,
higher speeds are predicted for observations starting/ending along the highway, while lower speeds are predicted for observations starting/ending along local roads.

### 3.6 Departure time choice models

As a next step, we estimate departure time choice models. We use the location-specific coefficients derived from the GWR model to approximate door-to-door travel times and based on these the values of the other attributes of the utility function. Moreover, we show to which extent the estimated coefficients are different if less precise travel time definitions are used.
3.6. Departure time choice models

12.1: Map

12.2: Predictions: $C_1-C_2$ speed = 50 km/h

12.3: Predictions: $C_1-C_2$ speed = 100 km/h

Figure 12: Speed predictions
Utility function

The continuous departure time choice problem is re-formulated as a discrete problem with a finite number of departure time intervals. We use a standard multinomial logit (MNL) model for the estimation. A driver $z$ faces a sequence of $k = 1, \ldots, K$ choices among the alternatives $j = 1, \ldots, J$, where $J = 18$, as departure time intervals of 15 minutes between 5:30 and 10:00 a.m. are considered. $K$ is equal to the duration of the experiment of 75 working days. As most drivers were not observed to pass the C1–C2 link on all 75 days, we only use the driver-specific subset of days, during which a driver has been observed to travel in the estimation. The driver chooses the alternative that maximizes the following random utility function:

$$U_{zkj} = V_{zkj} + \epsilon_{zkj},$$

(3.8)

The random utility function consists of a deterministic component $V_{zkj}$ and a random component $\epsilon_{zkj}$ that follows a Gumbel distribution, with errors distributed identically and independently (iid) across observations. To account for a bias in the standard errors as a result of the panel nature of the data, the panel sandwich estimator is used (e.g. Daly and Hess, 2011). The choice probability of alternative $\tilde{j}$ is then given by:

$$P_{zk\tilde{j}} = \frac{\exp(V_{zkj})}{\sum_{j=1}^{J} \exp(V_{zkj})}.$$  

(3.9)

The formulation of the deterministic utility component builds on the scheduling model of Vickrey (1969) and Small (1982). It accounts for the trade-off between travel times and schedule delays. Since we take into consideration that travel times may vary across days, the attributes of the utility function are the expected measures of reward, $ER_{zkj}$, travel time, $ET_{zkj}$, schedule delay early, $ESDE_{zkj}$, and schedule delay late, $ESDL_{zkj}$. The corresponding coefficients are denoted by $\beta_R$, $\beta_T$, $\beta_E$ and $\beta_L$.

$$V_{zkj} = \beta_R \cdot ER_{zkj} + \beta_T \cdot ET_{zkj} + \beta_E \cdot ESDE_{zkj} + \beta_L \cdot ESDL_{zkj},$$

(3.10)

The values of time (VOT) and schedule delay early (VSDE) and late (VSDL) are then defined as the ratios between the coefficients of time and schedule delays and the reward coefficients. They can be interpreted as the values attached to changes in expected travel times and schedule delays, respectively. The ratios are multiplied by $(-1)$ in order to account for the fact that the monetary component is defined here as a reward.

$$VOT = -\beta_T / \beta_R \quad VSDE = -\beta_E / \beta_R \quad VSDL = -\beta_L / \beta_R$$  

(3.11)

When determining the expected values of the alternative specific attributes $A = \{R, T, SDE, SDL\}$, we take into account that travelers usually have more knowledge on travel times than just the time-of-day specific averages. We model this by assuming that the expected values of the attributes are based on a compound measure of average travel time across the entire experiment, thus across days.
3.6. Departure time choice models

$k = 1, \ldots, K$ and the realized travel time on the day of travel, $\hat{k}$, which we refer to as ‘current’ travel times. Even though current travel times are typically unknown in advance (hence, at the time the departure time decision is taken), they represent an upper threshold for the maximum extent of information available to drivers. The average travel times represent the long-run pattern of travel times over the time of the day, which participants of the experiment are likely to be aware of. We denote the (estimated) relative weight attached to the current travel times by $\theta$. The expectation for the attribute $A$ of the utility function is then given by:

$$EA_{zkj} = \theta \cdot A_{zkj} + (1 - \theta) \cdot \frac{1}{K} \sum_{k=1}^{K} A_{zkj}$$

(3.12)

If a driver $z$ is eligible for a reward on day $k$, the reward linked to departure time choice $j$, $R_{zkj}$, is equal to 4 Euro if it results in a passage time at C2 before 6:30 or a passage time at C1 after 9:30, and it is equal to 0 in all other cases. Schedule delays are a function of the difference between the arrival time, which depends on departure time $t_{zkj}$ and travel time $T_{zkj}$, and the preferred arrival time $PAT_z$, which is defined as the self-reported preferred arrival time in case no congestion would ever occur.

$$SDE_{zkj} = \max(0, PAT_z - t_{zkj} - T_{zkj})$$

(3.13)

$$SDL_{zkj} = \max(0, t_{zkj} + T_{zkj} - PAT_z)$$

It is likely that travel times are to some extent endogenous, since they are determined by the choice of the residential location (whereas the choice of the work location is arguably exogenous, in particular because we only focus on work locations in and close by The Hague). However, this does not automatically induce an endogeneity bias in the parameter estimates, since these are determined by attribute differences rather than absolute attribute values in discrete choice models. So, for an endogeneity bias to become evident, drivers would have to choose their residential location taking into account the actual recurrent congestion pattern over the time of the day. However, this seems to be fairly unlikely as residential location choices are long-term, infrequent decision processes, whereas the recurrent congestion pattern can change quite easily over time, for instance as a consequence of traffic management measures, or in this specific case also due to a road capacity extension (the respective construction works were taking place during the peak avoidance experiment). Endogeneity bias hence does not seem to be a major concern here.

**Use of the GWR coefficients**

We use the GWR coefficients from Table 8 to approximate door-to-door travel times for both chosen and unchosen departure time intervals, as well as for the approximation of the departure time from home.

---

38A similar definition of expected travel times was used by Tseng et al. (2011), however, they do not apply the compound measure to all attributes of the utility function but to travel times only.
For all drivers, home and work locations are known, as well as the shortest network distances between their home locations and C1, and between C2 and their work locations. It is not necessarily the case that these coincide with the start and end locations of the links that were considered in the original GWR models. In order to derive the location-specific coefficients for a specific location \( \hat{u} (\hat{\lambda}_0(\hat{u}), \hat{\lambda}_1(\hat{u})) \), we thus use spatial interpolation. More specifically, we apply the Gaussian kernel function with the optimal bandwidths (see Table 8) to the distance between \( \hat{u} \) and all locations \( u(i) \), which are associated with the \( i = 1, \ldots, N \) observations considered in the original GWR model. The GWR coefficients specific to location \( \hat{u} \) are then equal to the original GWR coefficients weighted by the kernel weights with respect to \( \hat{u} \).

\[
\hat{\lambda}_p(\hat{u}) = \frac{1}{\sum_{i=1}^N w_i(\hat{u})} \sum_{i=1}^N w_i(\hat{u}) \hat{\lambda}_p(u(i)) \quad \text{for } p = 0, 1
\]  

We can use these coefficients and the shortest network distances to approximate driver-, day-, and time-of-day-specific door-to-door travel times, \( T_{zkj} \), taking into account the relationship between \( \hat{t} \) and \( \tilde{t} \) as described in Eq. 3.6. Based on the resulting travel times, reward and schedule delay attributes are calculated. Finally, the departure time from home is defined as the departure time choice alternative \( j \) that results in a passage time at C1 closest to the observed one. In order to warrant the quality of the data, we do not consider those drivers in the departure time choice models who have a home (work) location that is more than 5 km away from the closest home (work) location considered in the original GWR estimation (17.73% of drivers). Moreover, we exclude drivers with a combination of \( \hat{\lambda}_0(\hat{u}), \hat{\lambda}_1(\hat{u}) \) that can lead to speed predictions larger than 110 km/h, given that the highest measured C1–C2 speed equals 107 km/h (1.77% of drivers).

### Data

Table 9 provides some descriptive statistics of the participants considered in the departure time choice models. It shows that, under uncongested conditions, a large majority of drivers prefer to arrive at work between 7:00 and 9:00 a.m. The average length of the home–C1 link and C2–work link is smaller than the corresponding distances considered in the original GWR estimation (see Table 9). This might be due to the fact that we do not include those drivers who have a home or work location that is further than 5 km away from the closest home (work) location considered in the original GWR estimation (see Table 9). This exclusion also average distances to the ‘closest GWR neighbor’ are small: 0.74 km for home locations and 0.23 km for work locations. The average values of the \( \hat{\lambda}_p \) and their standard deviations correspond closely to those found in the original GWR estimation (see Table 8).

---

\[\text{Retrieved from openrouteservice.org}\]

\[\text{We tested alternative weighting functions, such as using the coefficients of the closest neighbor or the closest 5 neighbors, or applying a uniform weighting function. However, the results of the departure time choice models are barely affected by the specification of this weighting function.}\]
Table 9: Descriptives of drivers considered in the choice model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value or Sample Average</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commute-related variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred arrival time at work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;7:00 a.m. (in %)</td>
<td>11.21</td>
<td>–</td>
</tr>
<tr>
<td>&gt;9:00 a.m. (in %)</td>
<td>12.31</td>
<td>–</td>
</tr>
<tr>
<td>Distance Home–C1 (in km)</td>
<td>22.38</td>
<td>15.69</td>
</tr>
<tr>
<td>Distance C2–Work (in km)</td>
<td>11.25</td>
<td>4.85</td>
</tr>
<tr>
<td>Distance to ‘closest GWR neighbor’: Home–C1</td>
<td>0.74</td>
<td>0.95</td>
</tr>
<tr>
<td>Distance to ‘closest GWR neighbor’: C2–Work</td>
<td>0.24</td>
<td>0.41</td>
</tr>
<tr>
<td>(\hat{\lambda}_{Home-C1})</td>
<td>49.75</td>
<td>16.10</td>
</tr>
<tr>
<td>(\hat{\lambda}_{C2-Work})</td>
<td>46.10</td>
<td>10.57</td>
</tr>
<tr>
<td>(\hat{\lambda}_{Home-C1})</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>(\hat{\lambda}_{C2-Work})</td>
<td>0.25</td>
<td>0.08</td>
</tr>
<tr>
<td>Choices</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr. of choices per individual</td>
<td>20.95</td>
<td>12.98</td>
</tr>
<tr>
<td>Nr. of individuals</td>
<td>455</td>
<td>–</td>
</tr>
<tr>
<td>Total nr. RP choices</td>
<td>9530</td>
<td>–</td>
</tr>
<tr>
<td>Duration of the RP experiment (working days)</td>
<td>75</td>
<td>–</td>
</tr>
</tbody>
</table>

Model specifications

We define three models that differ in the computation of the speeds on the GPS-observed links, and that represent three ways in which a researcher might tackle the difficulties of observing only part of a trip. They have in common that speeds along the C1–C2 link are based on observed speeds. However, for a given driver, day and departure time, the C1–C2 speeds may still differ across the models, since the passage time of the C1–C2 link and therefore also the speed along this link are determined by the speed that is assumed on the home–C1 link.

Model 1 represents the model with the most sophisticated definition of door-to-door travel times, given the available data sources. Speeds on the home–C1 and C2–work links are then computed from the GWR coefficients, \(\hat{\lambda}_0\) and \(\hat{\lambda}_1\) (as derived in Eq. 3.14), and vary across space, days and time of the day. Model 2 represents the case when speeds differ between days and time of the day, however, not across space. Home–C1 and C2–work speeds are thus independent from the home and work location. This setup is achieved by using the OLS coefficients \(\lambda_0\) and \(\lambda_1\) (see Table 8) rather than the GWR coefficients to determine home–C1 and C2–work travel times. Finally, Model 3 represents the case where travel times do not vary across space, days and time of the day. Speeds on all home–C1 and C2–work links are fixed to 80 km/h\(^{41}\) (i.e. \(\lambda_0 = 80, \lambda_1 = 0\)). While we do not explicitly consider a model where only C1–C2 travel times are accounted for (such a model would require a different specification of the departure time alternatives (at C1) as well as of the preferred arrival time (at C2)), Model 3 captures the main property of a

\(^{41}\)This corresponds approximately to the mean C1–C2 speed.
model that is based on the C1–C2 link only, namely the effect of ignoring travel time correlations between the GPS-observed and the C1–C2 link.

Besides these three models, that differ in how home–C1 and C2–work travel times are defined, we also distinguish between two scenarios that differ in how departure times from home are defined, representing different amounts of information that a researcher might have. In the first scenario, departure times from home are calculated using the coefficients corresponding to the respective model. For instance, in Model 2, both (home–C1, C2–work) travel times as well as departure times are then computed using the OLS coefficients. As a consequence, for a given driver and day, the departure time from home may differ across the models. We refer to this scenario as ‘unknown departure time’, as it is a proxy for the situation where departure times from home are not observed, and therefore need to be computed by the researcher. In the second scenario, the GWR coefficients are used to compute departure times in all models. We interpret these departure times as ‘known’ (observed) departure times, representing for instance the case when drivers are asked to fill in their departure times in a log-book. Since departure time choices are defined equally for all models in the second scenario, this also allows us to compare the models directly in terms of their fit. Clearly, for Model 1 the ‘known’ and the ‘unknown’ departure time scenario coincide.

### 3.7 Results: Departure time choice models

Table 10 shows the results of the GWR-based Model 1, as well as the results of Models 2 and 3 for the two scenarios defined in the previous section. As the scale of the utility may differ between the models, we focus on comparing the estimated VOT, VSDE and VSDL, which are independent of the scale (e.g. Train, 2003).

We obtain fairly high VOT estimates in all models (between 34 and 65 Euro/h). These can partly be explained by the rather high incomes of the participants (see Appendix A), which tend to be correlated positively to the VOT (e.g. Small et al., 2005). Also the fact that variability is not included as a separate term may lead to a upward bias of the VOT. Finally, it is reassuring that similarly high values have been obtained from an SP experiment conducted among the same set of drivers (see Chapter 4 of this thesis).

Compared to the VOT of 42 Euro/h obtained in Model 1, VOTs obtained in Model 2 are relatively low (34 Euro/h if departure times are defined as ‘unknown’, and 38 Euro if departure times are defined as ‘known’). One would expect that the finding that OLS coefficients predict travel times less precisely than GWR coefficients (see Table 8) results in a measurement error of the travel time, which biases downwards the time coefficient (Bhutta and Larsen, 2011; Train, 2003). Instead we find that the lower VOTs are mainly the result of higher reward coefficients rather than lower time coefficients. A possible explanation is related to the setup of the peak avoidance experiment under consideration: Since rewards tend to be (negatively) correlated to travel times (due to off-peak rewarding), the reward coefficient might pick up also utility gains from lower travel times. This effect may be stronger when travel times are measured imprecisely. Since only two
3.7. Results: Departure time choice models

Table 10: Estimation results: departure time choice models

<table>
<thead>
<tr>
<th>Departure time</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_R )</td>
<td>0.13</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>( t\text{-stat.} )</td>
<td>(3.57)</td>
<td>(4.06)</td>
<td>(4.49)</td>
<td>(3.90)</td>
<td>(3.92)</td>
<td></td>
</tr>
<tr>
<td>( \beta_T )</td>
<td>-5.56</td>
<td>-5.30</td>
<td>-8.62</td>
<td>-5.57</td>
<td>-9.21</td>
<td></td>
</tr>
<tr>
<td>( t\text{-stat.} )</td>
<td>(-4.81)</td>
<td>(-3.91)</td>
<td>(-3.68)</td>
<td>(-4.08)</td>
<td>(-3.98)</td>
<td></td>
</tr>
<tr>
<td>( \beta_E )</td>
<td>-1.73</td>
<td>-1.71</td>
<td>-1.67</td>
<td>-1.73</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>( t\text{-stat.} )</td>
<td>(-10.29)</td>
<td>(-9.95)</td>
<td>(-10.04)</td>
<td>(-10.05)</td>
<td>(-10.10)</td>
<td></td>
</tr>
<tr>
<td>( \beta_L )</td>
<td>-1.44</td>
<td>-1.48</td>
<td>-1.67</td>
<td>-1.46</td>
<td>-1.50</td>
<td></td>
</tr>
<tr>
<td>( t\text{-stat.} )</td>
<td>(-15.72)</td>
<td>(-15.38)</td>
<td>(-15.22)</td>
<td>(-15.41)</td>
<td>(-15.27)</td>
<td></td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.15</td>
<td>0.17</td>
<td>0.14</td>
<td>0.15</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>( t\text{-stat.} )</td>
<td>(4.02)</td>
<td>(3.70)</td>
<td>(3.52)</td>
<td>(3.60)</td>
<td>(3.23)</td>
<td></td>
</tr>
<tr>
<td>VOT (Euro/h)</td>
<td>42.27</td>
<td>33.56</td>
<td>49.92</td>
<td>38.05</td>
<td>64.74</td>
<td></td>
</tr>
<tr>
<td>VSDE (Euro/h)</td>
<td>13.15</td>
<td>10.85</td>
<td>9.68</td>
<td>11.81</td>
<td>11.68</td>
<td></td>
</tr>
<tr>
<td>VSDL (Euro/h)</td>
<td>10.95</td>
<td>9.34</td>
<td>8.74</td>
<td>9.96</td>
<td>10.56</td>
<td></td>
</tr>
<tr>
<td>Nr. Obs.</td>
<td>9530</td>
<td>9530</td>
<td>9530</td>
<td>9530</td>
<td>9530</td>
<td></td>
</tr>
<tr>
<td>LogLikelihood</td>
<td>-23450</td>
<td>-23454</td>
<td>-23485</td>
<td>-23485</td>
<td>-23535</td>
<td></td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.149</td>
<td>0.148</td>
<td>0.147</td>
<td>0.147</td>
<td>0.146</td>
<td></td>
</tr>
</tbody>
</table>

reward levels exist (0 and 4 Euro), rewards are less sensitive towards travel time mis-specifications. This hypothesis is supported by the finding that when Model 1 and Model 2 are compared for the case of ‘known’ departure times (which are identical across the models), the t-statistic of the time coefficient is higher for Model 1 and the t-statistic for the reward coefficient is higher in Model 2.

Comparing the VOT in Models 1 and 3, we find substantial overestimation of the VOT in Model 3 (50 Euro/h if departure times are defined as ‘unknown’, and 65 Euro for Model 3 if departure times are defined as ‘known’). In the scenario of ‘known’ departure times, the VOT of Model 1 is overestimated by 53%. This finding can be attributed to the fact that Model 3 ignores travel time correlations between the GPS-observed links and the C1–C2 link, which have been proven to exist (see Table 8). Not accounting for the correlations leads to an underestimation of the difference between peak and off-peak travel times, and therefore to an overestimation of the absolute time coefficient and the VOT.

Unlike the VOT, the VSDE and the VSDL are similar across all models, ranging between 10 and 13 Euro/h and between 9 and 11 Euro/h, respectively. This is also an indication that the assumption of a fixed speed of 80 km/h is a good representation of the average speed on the home–C1 and C2–work links, as it does not lead to a structural overestimation of schedule delay early (in the case of a too high speed) or schedule delay late (in case of a too low speed). Also the relative weight attached to the current travel times relative to average travel times, \( \theta \), is quite stable across the models, ranging between 0.12 and 0.17. Departure time decisions are thus more affected by past travel times (and the resulting attributes) than by travel times on the day of travel. This is plausible, since travel times on the day of travel are unknown.
in advance, and, moreover, drivers may find it difficult to change their routines from day to day. However, the fact that $\theta$ is significantly different from 0 in all models shows that travel time expectations differ across days. This finding proves the relevance of using models that are able to approximate day-specific travel times on those links for which no direct travel time measurements are available rather than using reported travel times or travel times derived from network models, both of which are usually not day-specific.

As mentioned above, it is not useful to compare the fit of the models in the scenario where departure times are ‘unknown’, and therefore based on the coefficients of the corresponding model, as the dependent variables differ. If we compare the model fit for the scenario where departure times are ‘known’, and therefore all models have the same dependent variable, we find that, as expected, Model 1 yields the highest log-likelihood, followed by Model 2 and Model 3.

### 3.8 Conclusions

We use geographically weighted regression (GWR) to estimate the location-specific relationship between speeds on links for which only infrequent GPS data are available and speeds on a main link for which continuous speed measurements are available. We find that the predictive power of these models improves considerably if one uses GWR rather than OLS. We use the coefficients resulting from the GWR model to approximate driver-, day- and time-of-day-specific door-to-door travel times. These are then used in the estimation of departure time choice models with attributes based on door-to-door travel times. We are able to show that not accounting for spatial variation of speeds leads to a downward bias of the value of time (VOT), whereas not accounting for the temporal variation in speeds results in an upward bias of the VOT. These findings seem to be generally valid in settings where speeds depend on the start and end locations of the trips and are positively correlated across trip parts.

Our study also demonstrates how an RP model can be set up without using any reported data – besides the preferred arrival time – neither for travel nor for departure times, thus avoiding biases that might result from the use for reported data. We also show that the estimates obtained from RP models strongly depend on how (expected) travel times are defined, thus, explaining some of the variation in the values of time and schedule delay found in the literature. This finding contrasts existing literature on the comparison between SP and RP data, which usually emphasizes the effect of the SP design and the laboratory setting on the respective estimates, whereas the coefficients obtained from the RP models are considered as ‘true’ estimates. Here, we show that also RP estimates are heavily dependent on how the attributes of the choice model are specified.

We focused on the application of the GWR method to calculate door-to-door travel times as an input for the departure time choice experiments, however, the method is generally applicable for travel time predictions in situations where continuous speed measurements are only available on some links and GPS data are available on other links. We expect such travel time predictions to gain relevance in the near future as more and more GPS data become available.
Overstating vs. Overreacting: Commuters’ perceptions of travel times
4.1 Introduction

Studies that focus on the relation between reported travel times and actual travel times usually come to the conclusion that individuals tend to overstate travel times (e.g. Burnett, 1978; Henley et al., 1981; MVA Consultancy, 1987; O’Farrell and Markham, 1974; Van Exel and Rietveld, 2009). However, only little research has been conducted to investigate whether the overstating of travel times is caused by misperceptions that also translate into actual travel behavior, hence affecting the choice of departure times, travel modes and routes, or whether it is due to reporting errors only. Depending on whether travelers only overreport travel times, or actually misperceive and thus overestimate them, different policy implications can be derived.

If a mismatch between stated and actual travel time is due to reporting errors only, rather than actual misperceptions, the main policy conclusion would be that reported travel times are untrustworthy, and should therefore not be used as a representation of actual travel times. In the reverse case, if the misperceptions are ‘real’, they will affect actual as well as hypothetical travel decisions in situations when travel times differ between choice alternatives. As a consequence, also the estimate of the travelers’ willingness to pay for reducing travel time, which is usually derived from their choice behavior in revealed preference (RP) or stated preference (SP) experiments, is affected. RP data contain choices undertaken in real-life circumstances, whereas SP data are collected in laboratory settings, where participants are asked to decide between hypothetical travel alternatives. Recent meta-analyses of relevant empirical research on the value of (travel) time (VOT) can for instance be found in Zamparini and Reggiani (2007), Small and Verhoef (2007), Shires and De Jong (2009) and Abrantes and Wardman (2011).

Misperceptions of travel times affect the VOT derived from RP and SP experiments in different ways. So, if a traveler perceives travel times as longer than they actually are, he is expected to ‘overreact’ to travel times in real-life situations. If his VOT is then estimated based on his RP behavior and using a model that assumes objective travel time measurements, the resulting hourly VOT represents his willingness to pay for a reduction of one hour of objective travel time rather than an hour of perceived travel time. Because he perceives travel times as longer than they actually are, the estimated VOT will be relatively high. Specifically, it will exceed the VOT that would result if the attributes of the choice alternatives were defined in perceived terms by a factor equal to the ratio of perceived and actual travel times. However, usually the VOT derived from a model that defines the attributes of the choice alternatives in objective terms is the relevant measure to use in policy appraisals, as changes in travel times due to policy interventions are almost always predicted in objective terms too (for instance, from traffic assignment models).

The consequences of travel time misperceptions are more severe if the VOT is derived from SP experiments. The underlying rationale suggested by Brownstone and Small (2005) is that if drivers overestimate travel time in reality, it is likely that

42 Besides travel time perceptions being usually unknown, it is not obvious either how travel time perceptions change with the implementation of new policies (e.g. Small and Verhoef, 2007, p.21).
in an SP setting travelers react to stated travel times as if they were overestimated as well, resulting in a relatively low VOT. For instance, an individual who perceives a travel time of 10 minutes as 20 minutes probably reacts to a travel time of 20 minutes in an SP setting in the same way as she would to a travel time of 10 minutes in reality. The value of time derived in an SP setting would therefore be half of the value of time that would be derived in an RP setting. As a consequence, if the resulting SP-based VOT is used in policy evaluation and thus applied to objectively measured changes in travel times, substantial biases can arise, also because the benefits (or costs) due to changes in travel times constitute a major category in the appraisal of most transport policies. Hensher (2001), for instance, estimates them at 60% of total user benefits.

In fact, it is a frequent outcome that RP estimates of the VOT tend to be higher than SP estimates, by roughly a factor 2 (e.g. Brownstone and Small, 2005; Ghosh, 2001; Hensher, 2001; Isacsson, 2007; Small et al., 2005). Besides the hypothesis that drivers misperceive travel times in reality, also other explanations have been put forward. Most of them can be summed up under the label ‘hypothetical bias’, which essentially arises if people behave differently in an SP setting than in an RP setting (see for instance Louviere et al. (2000) and Carlsson (2010) for overviews). So, it has been speculated that people are more sensitive towards monetary attributes in a hypothetical choice situations than in real life, where they might face scheduling restrictions, which they did not anticipated when making the choices in the SP experiment. Recent research shows that hypothetical biases can be reduced by providing choice options to the respondents that are as close as possible to the choices they face in real life (e.g. Hensher, 2010).

In this study, we first compare reported and actual travel times to test how closely they correspond to each other, and whether there is a structural bias towards over- or underreporting of travel times. Various specifications of ‘travel time ratios’, which represent the ratio of reported and actual travel times, are computed. We then test whether the extent of over- or understating can be explained by driver- and link-specific characteristics. Thereafter, the variation between individual drivers regarding the extent of over- and understating is used to identify whether a higher travel time ratio impacts actual and hypothetical travel choice behavior and thus influences the travel time valuations derived from RP and SP choice models. To better be able to test whether the relation between the RP- and SP-based VOT as suggested Brownstone and Small (2005) holds, we estimate a joint SP-RP model.

This chapter contributes to the literature on time perception by explicitly distinguishing between reported durations and perceived durations, whereas these are usually assumed to be identical in existing research. We will argue that at least in the context of trip durations, these two measures should not be assumed equal. The main line of reasoning, based on the results we obtain, is that if choice behavior can be better explained by actual, observed travel times rather than reported travel times, differences between reported and actual travel times should be interpreted as reporting errors rather than travel time misperceptions. To our knowledge Ghosh (2001) has been the only one so far who has tested whether travel choice behavior is affected by the (person-specific) extent of over- and underreporting of travel
times. He finds that the results of the RP model, specifically the value of time, are barely affected by reported travel times, and in do not explain the difference between SP and RP estimates. We are not aware of any comparable research using SP data. Finally, this chapter also adds to the recent literature on models that combine SP and RP data. Only few studies on travel choice behavior have so far been undertaken with both data sources as input (e.g. Börjesson, 2008; Brownstone and Small, 2005; Ghosh, 2001; Small et al., 2005).

This chapter models travel choice behavior in the form of scheduling decisions. In the face of substantial peak hour congestion present in most urban areas, scheduling models have gained an increasing amount of interest in the last few decades. They are able to shed light on the effects of policies such as capacity extension and road pricing on departure time choices, and hence, at the aggregate level, on peak congestion as a whole. The workhorse model for the analysis of scheduling choices has been developed by Vickrey (1969). He was the first to define departure time decisions as a result of trade-offs between travel times and schedule delays, which characterize the extent of earliness and lateness with respect to the preferred moment of arrival. His work was later on extended along theoretical (e.g. Arnott et al., 1993, 1994), as well as empirical lines of research (e.g. Noland and Small, 1995; Small, 1982). Small (1982) was the first to estimate monetary valuations of travel time and schedule delays, while Noland and Small (1995) were the first to incorporate stochasticity of travel times in the scheduling model. Empirically, departure time decisions are usually represented as choices between discrete departure time alternatives. The respective models can then be estimated using discrete choice analysis (McFadden, 1974; Train, 2003). In this chapter, besides the standard multinomial logit model, we employ panel latent class models, which are able to account for unobserved heterogeneity between individuals as well as the repeated nature of the underlying data.

All RP data used in this study have been collected in the context of a real-life peak avoidance experiment that took place in the Netherlands in 2009. It included approximately 2000 participants who were able to gain a monetary reward for not using a specific highway link during morning peak hours. Their travel behavior along this link was monitored using cameras capable of number plate detection. In addition to these RP data, reported travel times were gathered from a questionnaire, and an SP experiment was conducted among the same set of drivers. While the observed behavior was only measured directly along this specific highway link, day- and time-of-day-specific door-to-door travel times can be approximated applying the method described in Chapter 3 of this dissertation, which takes GPS data as an input. Taken together, this combination of data sources provides a unique opportunity to compare reported and actual travel times, and to test whether differences between reported and actual travel times also manifest themselves in actual and hypothetical travel choice behavior.

This chapter is structured as follows. Section 4.2 contains a review of related literature, while Section 4.3 discusses the specification of the travel time ratio and evidence on how it differs across commuters and links. Section 4.4 introduces the theoretical framework that allows for testing whether over- and underreported
travel times affect choice behavior in revealed and stated preference settings. Section 4.5 gives an overview of the SP and RP data used in the analysis, while Section 4.6 contains the econometric framework and the estimations of the choice models. Finally, Section 4.7 concludes.

4.2 Related literature

This study is embedded in a wide body of literature that deals with differences between objective and subjective time measurements. Clearly, differences between objective and subjective measurements do not only occur in the time domain, but in almost all aspects of life. However, we confine this literature review to studies on time perception, with an emphasis on the perception of travel time (and closely related measures such as speeds and distances).

If a difference exists between actual and perceived durations, one would expect that the misperceptions affect reported durations as well behavior in (real and hypothetical) choice situations where time plays a role. Since perceived durations are difficult to observe directly (at least in situations when reporting errors might be present), this chapter will draw conclusions on the existence and extent of misperceptions from whether they can be observed consistently in both reported durations and (hypothetical as well as actual) choice behavior. If this is not found to be true, other factors that affect either only behavior or only reported data must be relevant. So, reported durations might be subject to reporting errors, for instance if the reporting task is misunderstood. For obvious reasons these will not translate to choice behavior. Another possibility is the presence of behavioral biases, which are mainly relevant in choice behavior but not in the reporting of durations. These can for instance be driven by processes of bounded rationality, including heuristic choices, which are characterized by an incomplete and/or incorrect processing of information (e.g. Tversky and Kahneman, 1974), or hypothetical biases relevant in the SP domain (e.g. Liljas and Blumenschein, 2000).

The focus of this chapter lies on the relation between objective travel times, reported travel times and choice behavior. Conclusions on the presence of reporting errors and/or behavioral biases will only be drawn implicitly. Figure 13 provides an overview of this underlying framework of concepts employed in this study. The gray-shaded fields indicate the data sources used in the analysis, while the transparent fields indicate the latent concepts of biases and errors, which can usually not be observed directly.

The starting point of this literature review is the common finding that objective and subjective durations of activities are not always identical, and in fact diverge quite strongly under specific circumstances (for an extensive overview see for instance Grondin, 2010). Despite a large number of studies undertaken on this topic (especially in the field of psychology), evidence on when durations tend to be over- or underestimated differs substantially across studies. One main point of consensus, however, is that the perception of durations becomes increasingly inaccurate if the cognitive load during the time interval under question is high, supposedly because fewer mental resources are available for temporal processing.
This has been found to hold also in the context of car travel (e.g. Baldauf et al., 2009). Another pattern that is relevant for travel behavior is that familiar tasks (such as commuting) tend to be perceived as shorter than they actually are. The opposite holds true for activities that are not well predictable (such as traveling under the risk of non-recurrent delays) (Li, 2003; Van de Ven et al., 2011). Existing research on time perception also points out the role of emotions. Most studies suggest that time passes slower under very stressful conditions (e.g. Droit-Volet and Meck, 2007). The evidence is more mixed when it comes to situations with very low arousal. While some studies claim that in such situations time will seemingly pass faster (e.g. Block and Zakay, 1996), others report the opposite finding (e.g. Flaherty, 1999; Glicksohn, 2001).

Various methods to measure time perceptions have been established in the literature. The most prominent one, which is also used in this chapter, is verbal estimation. It implies that individuals are asked to state the perceived duration of a

---

43In their meta-analysis, Block et al. (2010) report that the direction of the misperceptions depends on whether respondents have been informed in advance that they need to judge the duration of a certain time span (prospective timing), or not (retrospective timing). They find that if cognitive load increases, time is perceived to pass faster in the case of prospective timing, and slower if retrospective timing applies.
4.2. Related literature

certain time span.\textsuperscript{44} The reported travel times are then usually interpreted as being equal to perceived travel times. While this seems to be a reasonable assumption under controlled experimental conditions and short time spans (some studies on time perception cover time intervals of only few seconds), other situations where the analyst has less control over the experiment and where durations are generally longer are probably more prone to reporting errors. Reporting errors can then drive a wedge between reported and perceived durations. Possible causes are strategic behavior, social desirability bias, inaccurate recall, or a misunderstanding of the reporting task. In general, it is difficult to disentangle reporting errors from misperceptions, as for instance the study by Tai-Seale and McGuire (2012) on the duration of primary-care office visits illustrates. Reporting errors in the context of travel behavior have for example been encountered by Kitamura and Bovy (1987).

Also most studies on time perception in the field of travel behavior implicitly assume that reported travel times represent travel time perceptions. They generally find that reported travel times are substantially overstated (e.g. Burnett, 1978; Henley et al., 1981; MVA Consultancy, 1987; O’Farrell and Markham, 1974; Van Exel and Rietveld, 2009).\textsuperscript{45} For instance, Henley et al. (1981) suggest that on average individuals overstate actual travel times by 20%, with the relative bias becoming stronger for longer travel times. While the explanations provided for these findings are diverse, and often fairly specific to the context of the studies, several papers attempt to explain heterogeneity in the biases by person-, mode- and route-specific covariates (e.g. Burnett, 1978; Ghosh, 2001; Henley et al., 1981; Van Exel and Rietveld, 2009). Some of these studies also refer to the possibility of reporting errors. One example is Van Exel and Rietveld (2009) who find that regular car users overstate public transport travel time by 46%. They mention that this result might be a consequence of car users trying to consciously or unconsciously justify their mode choice in favor of the car rather than public transport.

While the literature that is concerned with the perception of time in general mainly focuses on the comparison of reported and actual durations, the literature on travel behavior has attempted to draw conclusions on travel time misperceptions also based on the behavior of travelers in hypothetical and actual choice situations. So, Brownstone and Small (2005) speculate that the common result of the VOT being higher under congested and slow traffic conditions (e.g. Hensher, 2001; Recarte and Nunes, 1996; Wardman, 2001; Wardman and Nicolás Ibáñez, 2012; Zhang et al., 2005) might be an indication that the time spent under such traffic conditions seems to pass slower, supposedly due to the annoyance with heavy traffic.

\textsuperscript{44}The most relevant alternative methods are reproduction (respondents need to reproduce a duration) and comparison of durations (Grondin, 2010).
\textsuperscript{45}Also Rietveld et al. (1999) report that travel times are overreported by 24%, however, they mainly attribute this finding to a wrong indicator of actual travel times, rather than misperceptions.
4.3 Specification of the travel time ratio

Data

The computation of the travel time ratio involves the comparison of reported and actual travel times. All data required to accomplish this comparison have been collected in the context of a peak avoidance experiment (Spitsmijden in Dutch). It took place in the Netherlands along a 9.21 km long stretch of the A12 highway leading to The Hague. As this link is confined by two cameras that are used for number plate detection, we refer to its start and end location as C1 and C2, respectively. The main goal of the experiment was to reduce traffic along this link in order to mitigate the impact of planned road works. While the entire experiment lasted for more than a year, the analysis conducted in this study focuses on the time frame between September 2009 and January 2010 for reasons of data availability. Approximately 2000 commuters participated in the experiment during that time. Before they entered the experiment, their average number of weekly trips through the C1–C2 link were measured and defined as reference behavior. Between September and December, they were able to obtain a reward of 4 Euro for each avoided trip along the C1–C2 link during morning peak hours (6:30-9:30 a.m.), relative to their reference behavior. Appendix A provides a more detailed explanation of the setup of this experiment.

The main data source for reported travel time data is an online questionnaire conducted among the participants of the experiment, resulting in a response rate of approximately 30%. Respondents were asked several questions addressing their perceptions of travel times. Another survey, which was conducted at the time when drivers entered the experiment, is used as supplementary data source. Multiple data sources are used to define actual travel times. Travel times through the C1–C2 link are observed individually for each driver using number plate detection at the begin and the end of the link. Home–C1 and C2–work travel times, however, need to be approximated. The corresponding method is described in Chapter 3. It involves geographically weighted regression (GWR), which is used to measure the extent of correlation between speeds on the C1–C2 and home–C1 as well as C2–work links respectively. As a result, day- and time-of-day-specific travel times can be approximated for almost all combinations of home and work locations in the dataset. Chapter 3 demonstrates that this method yields reliable predictions of travel times.

Definitions

We first provide the exact definitions of the reported and actual travel times, as well as of the travel time ratio, which is defined as the ratio of reported and actual travel times.\textsuperscript{47}

\textsuperscript{46}We do not consider those persons who have a home (work) location that is more than 5 km away from the closest home (work) location considered in the original GWR estimation.

\textsuperscript{47}Note that in the literature on time perception this ratio is also referred to as duration judgement ratio (e.g. Block et al., 2010).
4.3. Specification of the travel time ratio

Reported travel times are derived from the main questionnaire where participants were asked to state their average travel time on the home–C1, C1–C2, C2–work links, taking into account their most recent 20 morning commute trips. To prompt respondents to give realistic answers, maps of the location of the C1–C2 link were shown next to the questions. Average home–work travel times are then computed as the sum of the average travel times indicated for each of the three sub-links. We label this definition as ‘standard’, in order to distinguish it from an alternative measure of reported (home–work) travel times. The alternative measure is derived from the supplementary survey, where drivers were asked for their average home–work travel time, thus, not distinguishing between any sub-links. To emphasize this property, this definition is labelled ‘total’. We use it to test whether the results of the ‘standard’ measure might be biased because they are based on the sum of travel times along the sub-links rather than total home–work travel times. Reported travel times are denoted by \( T^R_{lz} \), where \( l \) indicates the link \( l = \{ \text{Home–C1, C1–C2, C2–Work, Home–Work} \} \) and \( z \) indicates the participant.

Actual travel times are derived from driver-specific observations along the C1–C2 link and the corresponding approximated travel times on the home–C1 and C2–work links. They are denoted by \( T^A_{lz} \), where \( l \) again denotes the link and \( z \) the driver. Moreover, \( d = 1, \ldots, D \) is the index attached to the observed passages of the C1–C2 link, where \( d = 1 \) indicates the most recent trip before filling out the survey, \( d = 2 \) the second most recent trip, continuing up to \( D \), which thus denotes the overall number of observed passages between September 1, 2009 and the date a specific driver has completed the questionnaire. For an observed passage to be considered in this list of past trips, certain conditions must be fulfilled. For the ‘standard’ definition, we assume that passages of the C1–C2 link between 6:00 and 11:00 are considered commuting trips. In an alternative definition, labelled ‘peak’, only those trips that result in a passage of the C1–C2 link between 6:30 and 9:30 are considered commuting trips. Although the latter definition reduces the number of observed trips that are available for the analysis, we use it in order to compare whether the standard definition induces biased results of the travel time ratio by taking into account an excessively broad peak, which possibly leads to lower average actual travel times (due to lower off-peak travel times), and as a consequence to an upward bias in the travel time ratio.

Not for all drivers data on 20 recent commuting trips are available (as suggested to be taken into account when filling in the questionnaire). In this case, we take into account as many observations as are available. In the reverse case, when more than 20 commuting trips with passages of the C1–C2 link have been observed, the 20 most recent ones enter the analysis. The average actual link- and driver-specific travel time is therefore defined as follows:

\[
\bar{T}^A_{lz} = \frac{1}{\min[20, D]} \sum_{d=1}^{\min[20, D]} T^A_{lz}
\]

(4.1)

The link- and driver-specific travel time ratio \( \tau_{lz} \) is then defined as the ratio between the reported average travel time and the average actual travel time. The travel time ratio will thus be larger than 1 for all cases of overstating, and between 0
and 1 for case of understating travel times. An identical formulation has for instance been used by Parthasarathi et al. (2011) for comparing reported and observed travel times.

\[
\tau_{iz} = \frac{T_{iz}^R}{\bar{T}_{iz}^A}
\]  

Studies on travel time perceptions often adopt a power law function in order to control for nonlinearities in the relation between reported and actual durations, with the general conclusion that the exponential term is close to 1, and nonlinearities are therefore small (e.g. Allan, 1979; Eisler, 1976).\textsuperscript{48} Leiser and Stern (1988) showed that also in the context of travel times, the linear formulation performs almost as well the power law function. For this reason as well as the fact that one of our main interests in this research is to use the travel time ratio as an input to departure time choice models, we do not adopt these more complex specifications here.

**Descriptives**

Table 11 provides the descriptive statistics for the reported and observed average travel times as well as for the travel time ratio. The two bottom lines of the table give the results for the alternative measures of the reported (‘total’) and observed (‘peak only’) home–work travel times. In total, valid measures for both reported and actual travel times for the ‘standard’ definitions can be derived for 540 respondents.

<table>
<thead>
<tr>
<th>Link</th>
<th>Label</th>
<th>Nr. Obs.</th>
<th>Reported Mean</th>
<th>St. D.</th>
<th>Observed Mean</th>
<th>St. D.</th>
<th>( \tau ) Mean</th>
<th>Median</th>
<th>St. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H–C1</td>
<td>–</td>
<td>540</td>
<td>31.12</td>
<td>19.47</td>
<td>22.18</td>
<td>16.70</td>
<td>1.67</td>
<td>1.48</td>
<td>0.68</td>
</tr>
<tr>
<td>C1–C2</td>
<td>–</td>
<td>540</td>
<td>15.67</td>
<td>5.91</td>
<td>9.84</td>
<td>2.71</td>
<td>1.66</td>
<td>1.56</td>
<td>0.67</td>
</tr>
<tr>
<td>C2–W</td>
<td>–</td>
<td>540</td>
<td>15.95</td>
<td>7.68</td>
<td>11.05</td>
<td>6.91</td>
<td>1.74</td>
<td>1.51</td>
<td>1.00</td>
</tr>
<tr>
<td>H–W</td>
<td>stand.</td>
<td>540</td>
<td>62.74</td>
<td>22.74</td>
<td>43.07</td>
<td>18.02</td>
<td>1.53</td>
<td>1.49</td>
<td>0.38</td>
</tr>
<tr>
<td>H–W</td>
<td>total</td>
<td>527</td>
<td>64.59</td>
<td>21.45</td>
<td>43.07</td>
<td>18.02</td>
<td>1.62</td>
<td>1.56</td>
<td>0.44</td>
</tr>
<tr>
<td>H–W</td>
<td>peak</td>
<td>486</td>
<td>62.74</td>
<td>22.74</td>
<td>43.92</td>
<td>17.60</td>
<td>1.50</td>
<td>1.45</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Overstating is found to be a persistent phenomenon for all sub-links as well as for all definitions of the travel time ratio, with its average ranging from 1.50 to 1.74 (the median is slightly lower, ranging from 1.45 to 1.56). The first three rows of Table 11 demonstrate that the average travel time ratio exceeds 1 on all sub-links, hence, rejecting the hypothesis that (on average) drivers overreport travel times only on some of the sub-links (e.g. those affected most by congestion) and compensate this overreporting by understating travel times on the remaining links of the commute.\textsuperscript{49} A comparison of the ‘total’ definition of the travel time ratio

\textsuperscript{48}Eisler et al. (2007) do not only account for linearities but also for subjectivity in the point of time when an activity is experienced to start as well as discontinuities.

\textsuperscript{49}The finding that the average home-work travel time ratio is smaller than the corresponding travel time ratios on the sub-links is driven by some outliers on the sub-links.
4.3. Specification of the travel time ratio

to the ‘standard’ one shows that the extent of overstating is even higher if drivers are asked for the total average home–work travel time rather than the average travel times on the sub-links (1.62 vs. 1.53). We can therefore conclude that the overreporting found for the ‘standard’ definition is not caused by the fact that the survey questions referred to the travel times on the sub-links rather than door-to-door travel times. Finally, the finding that the travel time ratio for the ‘peak’ definition is almost as high as for the ‘standard’ definition (1.50 vs. 1.53) proves that the overreporting found for the ‘standard’ definition is not due to using an excessively broad definition of the morning peak.

Figure 14 shows a histogram of the person-specific home-work travel time ratio, based on the ‘standard’ definition. Only a very minor share of respondents (4.44%) have a travel time ratio less than 1 and, therefore, understate travel times. The figure also demonstrates that the distribution of the travel time ratio is fairly symmetric with a mean of ca. 1.5 and few outliers at the right side of the distribution.

Figure 14: Histogram travel time ratio (home–work)

Heterogeneity

This section intends to explain the variations in the travel time ratio across drivers and across links by regressing the ‘standard’ travel time ratios on driver and link-specific characteristics. The descriptive statistics of the variables considered in the regressions are provided in Table 12. In addition to these, we also tested the significance of various socio-economic variables (age, gender, education, income\(^\text{50}\), children, flexibility of working hours), alternative measures of the frequency of

\(^{50}\)Our finding that the travel time ratio is independent of income is in contrast to the result obtained by Burnett (1978) who concludes that individuals with low income overstate travel times most, supposedly due to lower education levels or less experience in traveling.
travel (as a proxy for experience), measures of travel time variability (specifically, percentile differences and standard deviations), characteristics of the most recent trip, and person-specific outliers in terms of travel times. However, all of them were found to be insignificant. For various other variables that might affect the travel time ratio, such as road type or free-flow speed, no data were readily available.

Table 12: Descriptive statistics of driver- and link-specific variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Specification</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of trips</td>
<td>within the last 20 working days</td>
<td>13.01</td>
<td>6.76</td>
</tr>
<tr>
<td>Share of trips outside peak</td>
<td>since September 1, 2009</td>
<td>0.20</td>
<td>0.29</td>
</tr>
<tr>
<td>Mean speed home–work</td>
<td>in km/h</td>
<td>70.50</td>
<td>12.05</td>
</tr>
<tr>
<td>Mean speed home–C1</td>
<td>in km/h</td>
<td>74.89</td>
<td>18.65</td>
</tr>
<tr>
<td>Mean speed C1–C2</td>
<td>in km/h</td>
<td>66.91</td>
<td>13.82</td>
</tr>
<tr>
<td>Mean speed C2–work</td>
<td>in km/h</td>
<td>61.94</td>
<td>12.05</td>
</tr>
<tr>
<td>Distance home–work</td>
<td>in km</td>
<td>50.57</td>
<td>25.38</td>
</tr>
<tr>
<td>Distance home–C1</td>
<td>in km</td>
<td>29.82</td>
<td>24.60</td>
</tr>
<tr>
<td>Distance C1–C2</td>
<td>in km</td>
<td>9.21</td>
<td>0.00</td>
</tr>
<tr>
<td>Distance C2–work</td>
<td>in km</td>
<td>11.53</td>
<td>6.09</td>
</tr>
</tbody>
</table>

In the first model, the travel time ratios for the home–work link are used as dependent variable, while in the second model the travel time ratios of all three sub-links are considered. Therefore, in the first model, the set of explanatory variables consists of person-specific variables as well as variables that characterize home–work trips. In addition to those variables, the second model also accounts for variables that are defined at the level of the sub-links. A standard ordinary least squares (OLS) estimation is conducted for Model 1. For Model 2, random effects (RE) regression is used in order to account for the panel structure of the dataset that is due to the fact that for each driver three partial links have been defined. Table 13 shows the according results.

Both model estimations result in an R-square of 0.22 to 0.23. This is an indication that a considerable share of the travel time ratio can indeed be explained by driver- and link-specific characteristics. The first model shows that the home–work travel time ratio is relatively high for drivers with short commuting distances and high average speeds on their commutes. In addition, we find some evidence that those divers tend to overstate travel times more who have little experience in commuting\(^5\), which is consistent with earlier studies finding that unfamiliar tasks are perceived longer than familiar ones (e.g. Boltz et al., 1998). Also individuals who travel relatively often during the peak, meaning that they pass the C1–C2 link between 6:30 and 9:30, tend to overstate travel times. This is in agreement with the hypothesis discussed in Section 4.2 stating that travel times under congested traffic conditions may be perceived as longer. All of these results are also confirmed at the link level.

\(^5\)This variable is a dummy that assumes the value 1 if a driver has undertaken less than 4 trips since the start of the reward period in September 2009.
4.4 Behavioral models: Theoretical framework

In this section the theoretical framework will be introduced that allows us to test whether the biases found in the reported travel time data can be found back in hypothetical and real-life departure time choice behavior. If this is the case, we can conclude that the reported travel times are an indicator of travel time perceptions. Otherwise, differences between reported and actual travel times should rather be attributed to reporting errors.

### Table 13: Regression results: Explanation of the travel time ratio

<table>
<thead>
<tr>
<th>Variable</th>
<th>τ Home–Work Coefficient</th>
<th>t-stats.</th>
<th>τ Home–C1,C1–C2,C2–Work Coefficient</th>
<th>t-tstats.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.25</td>
<td>13.55</td>
<td>1.52</td>
<td>8.23</td>
</tr>
<tr>
<td>Little experience</td>
<td>5.59×10^{-2}</td>
<td>1.78</td>
<td>3.54×10^{-2}</td>
<td>0.56</td>
</tr>
<tr>
<td>Share of trips outside peak</td>
<td>-0.27</td>
<td>-4.71</td>
<td>-0.32</td>
<td>-4.18</td>
</tr>
<tr>
<td>Mean speed home–work (MSHW)</td>
<td>1.07×10^{-2}</td>
<td>6.71</td>
<td>5.82×10^{-3}</td>
<td>2.64</td>
</tr>
<tr>
<td>Distance home–work (DHW)</td>
<td>-8.87×10^{-3}</td>
<td>-12.26</td>
<td>-6.83×10^{-3}</td>
<td>-6.93</td>
</tr>
<tr>
<td>Mean speed link/MSHW</td>
<td>–</td>
<td>–</td>
<td>0.69</td>
<td>5.29</td>
</tr>
<tr>
<td>Distance link/DHW</td>
<td>–</td>
<td>–</td>
<td>-2.25</td>
<td>-18.83</td>
</tr>
<tr>
<td>dummy home–C1</td>
<td>–</td>
<td>–</td>
<td>0.55</td>
<td>11.48</td>
</tr>
<tr>
<td>dummy C2–work</td>
<td>–</td>
<td>–</td>
<td>0.18</td>
<td>4.40</td>
</tr>
<tr>
<td>Estimation Method</td>
<td>OLS</td>
<td></td>
<td>RE</td>
<td></td>
</tr>
<tr>
<td>Nr. of Observations</td>
<td>540</td>
<td></td>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>R-square</td>
<td>0.23</td>
<td></td>
<td>0.22</td>
<td></td>
</tr>
</tbody>
</table>

The relatively high extent of overreporting on the home–C1 and C2–work links might be explainable by drivers mistakenly not only taking into account in-vehicle commuting time, but also the time required to walk to the car, and from the car to the office, although they were explicitly asked to only consider in-vehicle commuting time when answering the questions. This effect is likely to be more important on the home–C1 and the C2–work links than on the intermediate C1–C2 link. Since this out-of-vehicle commuting time is probably fairly independent from the commuting distance, it is also a possible explanation for the relatively high travel time ratio for shorter commutes. The finding that travel times along the sub-links with higher speeds (home–C1, C2–work) are overreported more strongly might result because drivers know the length of the sub-links rather than the corresponding travel times (for instance from traffic signs installed along the roads). When prompted to indicate travel times on the sub-links, they might attribute their perceived total travel times to the sub-links more or less proportionally to their lengths, not taking into account the differences in speeds across links. A similar pattern has been found by Leiser et al. (1991).
Basic model

Commuters choose their optimal day-specific departure time from home subject to expected recurrent and non-recurrent traffic congestion as well as their preferred arrival time at work. The utility maximization involves the trade-off between travel time, schedule delays and monetary trip costs (in this study, monetary rewards) (e.g. Small, 1982; Vickrey, 1969).

Schedule delays indicate the extent of earliness and lateness with respect to the preferred arrival time (PAT). The latter is often defined as the preferred moment of arrival at work under uncongested conditions (e.g. Börjesson, 2009), or as work starting time (e.g. Noland and Small, 1995; Small, 1982). Since earliness and lateness might be valued differently, we distinguish between schedule delays early (SDE) and schedule delays late (SDL). $T(t_h)$ denotes the travel time associated with the departure time from home $t_h$. Note that for simplicity, we drop here the indices related to individual $z$, alternative $j$ and choice situation $k$.

\[
SDE = \max[\text{PAT} - t_h - T(t_h), 0] \tag{4.3}
\]
\[
SDL = \max[t_h + T(t_h) - \text{PAT}, 0]
\]

Besides schedule delays SDE and SDL, we define travel time $T$ and reward $R$ as attributes of the utility function. We account for stochastic travel times by taking into account the expectations $E[.]$ of the attribute values rather than the deterministic values (e.g. Noland and Small, 1995). The coefficients associated with the reward, travel time and schedule delays early and late are then denoted by $\beta_R, \beta_T, \beta_E$ and $\beta_L$. The expectation of the systematic part of the utility function, $E[V]$ assumes the following additive form:

\[
E[V] = \beta_R * E[R] + \beta_T * E[T] + \beta_E * E[SDE] + \beta_L * E[SDL] \tag{4.4}
\]

The values of time (VOT) and schedule delay early (VSDE) and late (VSDL) can then be defined based on the ratios between the coefficients of time and schedule delays and the reward coefficient. These characterize the values attached to changes in expected travel times and schedule delays, respectively. The ratios need to be multiplied by $(-1)$ in order to account for the fact that the monetary component is defined here as a reward rather than a cost:

\[
VOT = -\frac{\beta_T}{\beta_R} \quad \text{VSDE} = -\frac{\beta_E}{\beta_R} \quad \text{VSDL} = -\frac{\beta_L}{\beta_R} \tag{4.5}
\]

Below the RP and SP models employed for testing the effect of reported travel times on choice behavior are discussed. They both use the same utility function as given in Eq. 4.4, but then extended by additional additive terms.

RP model

The systematic part of the utility function associated with RP-based departure time choices, denoted by $E[V_{RP}]$, consists of the additive terms already included in $E[V]$ (see Eq. 4.4), as well as terms that indicate the difference between the attribute
values that would result if drivers would over- or underestimate travel times to the same extent as they over- or underreport them and objective attribute values. Such difference terms are included for all four attributes: rewards, travel times and schedule delays early and late. It seems unlikely that individuals would misperceive travel times, but not take into consideration the effects of the misperception on the other attributes of the utility function. So, for instance, a driver who overestimates travel times is expected to structurally underestimate schedule delays early. The coefficients corresponding to those differences as well as the attribute values that would result if travel times were misperceived are labelled with the superscript $\tau$.

\[
E[V_{RP}^\tau] = \beta_R * E[R] + \beta_T^\tau * (E[R^\tau] - E[R]) + \\
\beta_T * E[T] + \beta_T^\tau * (E[T^\tau] - E[T]) + \\
\beta_E * E[SDE] + \beta_E^\tau * (E[SDE^\tau] - E[SDE]) + \\
\beta_L * E[SDL] + \beta_L^\tau * (E[SDL^\tau] - E[SDL])
\] (4.6)

The coefficients related to the differences are expected to be insignificant if drivers take into account the objective attribute levels in their decision making rather than those attribute levels associated with over- or underestimation. On the contrary, if $\beta_R^\tau, \beta_T^\tau, \beta_E^\tau, \beta_L^\tau$ have the same sign and are roughly equal in size to $\beta_R, \beta_T, \beta_E, \beta_L$, respectively, we can conclude that the behavior of the individuals in real life is consistent with the travel times they reported. In turn, we can interpret reported travel times as an indicator of travel time perceptions.

**SP model**

For departure time observations derived in an SP setting and the corresponding systematic part of the utility function, $E[V_{SP}^\tau]$, only the difference between reported and objectively measured travel times is included (in addition to the attributes present in Eq. 4.4). The reason is that in the SP experiment, rewards as well as actual arrival times at work are presented to the respondents. As a consequence, it is unlikely that respondents perceive the reward and the schedule delays differently from the ones presented to them, even if they misperceive travel times.

\[
E[V_{SP}^\tau] = \beta_R * E[R] + \beta_T * E[T] + \beta_T^\tau * (E[T^\tau] - E[T]) + \\
\beta_E * E[SDE] + \beta_L * E[SDL]
\] (4.7)

The interpretation of $\beta_T^\tau$ is similar to its interpretation in the RP utility function. However, here we expect a positive $\beta_T^\tau$ in the case of overestimation. In that case, the monetary value attached to the difference in overstated and objectively measured travel times ($VOT^\tau = -\beta_T^\tau/\beta_R$) would be negative, hence, lowering the overall value attached to travel time ($VOT + VOT^\tau$) as suggested by Brownstone and Small (2005).

4.5 Behavioral models: Data

**RP data**

Compared to SP data, RP data have the advantage that they are based real choices rather than hypothetical ones. This comes with the disadvantage that there may be
strong correlations between the model variables, or that certain attributes simply
do not exist yet because they are part of a new product or a new situation, as it is
common in marketing and environmental studies.

In this study, RP data are defined in a similar way as the actual travel times used
to compute the travel time ratio. The main difference is that for choice analysis
not only the attributes of the chosen but also of the unchosen alternatives need
to be known. Hence, for each departure time alternative \( j \) for driver \( z \) and choice
situation \( k \), expected rewards, travel times and schedule delays need to be derived.
Choice situations refer here to days during which drivers have been observed to
pass the C1–C2 link, along which the reward experiment took place. Our models
thus describe the departure time choice conditional on making a commute trip, and
the overall number of RP choice situations is therefore driver-specific. Departure
time alternatives are defined as discrete intervals, each of them being equal to
a 15-minute interval during the morning peak. There are 17 choice alternatives,
ranging from 5:30 a.m. to 9:45 a.m.. Chosen departure times need to be situated
within this time frame, otherwise the according observations are dropped.\(^{52}\) This
is because trips that take place outside this interval are usually undertaken for
purposes other than regular commuting, implying that also different, unobserved
preferred arrival times (PATs) apply.

The person-specific PATs that apply for regular commuting trips (and relative
to which schedule delays are defined) are derived from a questionnaire. They are
defined as the preferred moment of arrival at work if there was no congestion (ever).
Only drivers with a PAT between 6:30 and 9:30 are taken into account, since drivers
with very early or late PATs face barely any trade-off between schedule delays and
travel times.\(^{53}\) Rewards vary by time of the day. They are equal to 4 Euro if a driver
passes C2 before 6:30 a.m. or C1 after 9:30 a.m., respectively, and 0 otherwise.
Moreover, the maximum number of weekly rewards cannot exceed the number of
weekly peak trips that a driver has undertaken before the start of the experiment.
If the weekly maximum has been reached, no further rewards can be obtained
during the rest of the week. Moreover, no rewards were distributed during school
holidays and weekends. While rewards were only distributed between September
and December 2009, we also include post-measurements from January 2010 in
order to account for counterfactual behavior.

Given our knowledge of the PAT, the rewards and the driver-, day- and time-
of-day-specific door-to-door travel times (which are again computed using the
method presented in Chapter 3), the attribute levels specific to day \( k \), individual \( z \)
and choice alternative \( j \) can be derived. Expected attribute levels for attributes \( A = \{R, T, SDE, SDL\} \) are defined as a compound measure of actual, realized attribute
levels on the day of travel and the person-specific past experience of commuting
trips. While actual realizations are in principle unknown in advance (hence, at the
time the departure time decision is taken), they represent an upper benchmark
for the maximum extent of information possibly available to drivers. The person-
specific past experience is defined as the average time-of-day-specific attribute

\(^{52}\) Roughly, 5.6% of all RP observations are dropped due to this restriction.

\(^{53}\) Only, 2.9% of all participants have a PAT that is either before 6:30 or after 9:30.
level on the last \( d \) days during which a certain person has been observed to travel preceding the regarding day of travel \( k \) \((A_{zkj})\). Here we take into account the last 9 trips.\(^{54}\) We estimate the weight \( \theta \) that drivers attach to the actual realizations relative to the past experience; \( \theta \) will be equal to 0 if travel time expectations are based exclusively on the past experience, while it will be equal to 1 if they are perfectly informed about actual realizations, and thus do not take into account any past experiences. The expectation for attribute \( A \) of the utility function for driver \( z \), day \( k \) and departure time choice alternative \( j \) is thus given by:

\[
E[A_{zkj}] = \theta \cdot A_{zkj} + (1 - \theta) \cdot \frac{1}{9} \cdot \sum_{d=1}^{9} A_{zkjd} \tag{4.8}
\]

**SP data**

In SP experiments, respondents are asked to make hypothetical choices between alternatives with varying departure times, travel time levels and costs (or rewards). Because the researcher determines the attribute levels, problems of collinearity between attribute levels can be more easily avoided than in RP settings. However, SP-based estimates will be biased if respondents choose differently in a laboratory setting than in real life. These so-called hypothetical biases can be reduced by designing the SP experiments such that the realism of the choices and attribute levels is enhanced; for instance, by pivoting the design values around the status quo behavior of respondents (Hensher, 2010). This strategy is also adopted in the SP experiment considered in this chapter. In this experiment, the travel time attributes are designed such that they are situated between the minimum and maximum travel time reported by the driver, and the preferred arrival time shown to the respondents is identical to their reported PAT.\(^{55}\)

The SP choice experiment is composed of 10 choices between 2 different departure time alternatives. Each alternative comprises two possible travel time realizations, which occur with a certain probability and, in turn, also affect rewards and schedule delays. Figure 15 shows an example of the choice screen presented to the respondents. In the estimation of the SP choice models, the expected attribute values associated with each choice alternative are defined as weighted average across the two possible travel time realizations. After filling in the choice experiment, respondents were asked to indicate whether they answered some or all choice questions at random. Those who confirmed they did, are not taken into account in the analysis. The efficiency of the choice experimental design was thoroughly tested using extensive simulation in order to assure that a broad

\(^{54}\)This is a fairly reasonable assumption, since these past trips constitute the person-specific past commuting experience. Moreover, it is consistent with the definition of the travel time ratio, which is also based on past trips. Compared to a generic definition of past experiences, the person-specific definition comes with the disadvantage that observations drop out for which no sufficient number of previously experienced trips (here: 9) is available. Using a significantly lower number of past trips for the definition of the past experience, however, induces too much variation from outliers.

\(^{55}\)Since these turned out to overstated, the travel time attributes in the SP experiment were on average longer than travel times in reality. We will discuss possible implications in Section 4.6.
range of parameters can be reproduced (e.g. Koster and Tseng, 2010).\textsuperscript{56} A detailed description of the design of the SP experiment can be found in Knockaert et al. (2012).

**Dataset descriptives**

Table 14 shows some descriptive statistics for those participants who are taken into account in the SP and RP scheduling choice models. It shows that the majority of the participants to both experiments is between 35 and 50 years old, around a quarter of the participants is female and, probably most notable, many of them have a fairly high income. The table also provides insights on the distribution of the PAT under uncongested conditions. PATs show to be concentrated in the time interval between 7:30 and 8:30 for roughly half of all participants. The travel time ratio of the individuals taken into account in the choice analysis is almost identical to the travel time ratio derived for the entire sample (again using the ‘standard’ definition).\textsuperscript{57}

### 4.6 Behavioral models: Estimations

**Econometric framework**

We estimate standard multinomial logit (MNL) models and more advanced panel latent class models that take into account heterogeneity between individuals as well as the panel nature of the data. While we will estimate the MNL models separately for SP and RP data, a latent class model is estimated with pooled data sources. All models in this section are introduced using a general notation that also accounts for the pooled version of the models. Usually, the underlying notion of a joint analysis of SP and RP data is that SP data are able to correct for deficiencies in the

\textsuperscript{56}Moreover, the layout and the phrasing was tested using focus groups and an online test.

\textsuperscript{57}The sample is here somewhat smaller than in the computations of travel time ratio, because we exclude drivers who choose randomly in the SP experiment.
Table 14: Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value or Fraction of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-economic characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>&lt; 35 years</td>
<td>0.25</td>
</tr>
<tr>
<td>&gt; 50 years</td>
<td>0.21</td>
</tr>
<tr>
<td>Female</td>
<td>0.25</td>
</tr>
<tr>
<td>Monthly (net) income</td>
<td></td>
</tr>
<tr>
<td>&lt; 3500 Euro</td>
<td>0.31</td>
</tr>
<tr>
<td>&gt; 5000 Euro</td>
<td>0.14</td>
</tr>
<tr>
<td>unknown</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Commute-related variables</strong></td>
<td></td>
</tr>
<tr>
<td>Preferred arrival time (PAT)</td>
<td></td>
</tr>
<tr>
<td>&lt; 7:30 a.m.</td>
<td>0.25</td>
</tr>
<tr>
<td>&gt; 8:30 a.m.</td>
<td>0.27</td>
</tr>
<tr>
<td>Mean travel time ratio (τ)</td>
<td>1.54</td>
</tr>
<tr>
<td><strong>Choices</strong></td>
<td></td>
</tr>
<tr>
<td>Nr. of individuals</td>
<td>493</td>
</tr>
<tr>
<td>Share post-measurements in RP</td>
<td>0.33</td>
</tr>
<tr>
<td>Total nr. SP choices</td>
<td>4930</td>
</tr>
<tr>
<td>Total nr. RP choices</td>
<td>2997</td>
</tr>
</tbody>
</table>

RP data such as correlations between attributes or the lack of identification of some attributes or attribute ranges (Louviere and Hensher, 2001; Louviere et al., 2000). Estimations that pool RP and SP data are most advantageous when the RP and SP choice situations are similar and are also perceived similarly by the participants (e.g. Börjesson, 2008). In that case, common coefficients for both SP- and RP-based observations may be possible if individuals are found to react in consistent ways to trade-offs they face in both RP and SP choice situations.

We define the utility $U_{zkj}$ as the utility associated with choice $k = 1,\ldots,K_z$, individual $z = 1,\ldots,Z$ and choice alternative $j = 1,\ldots,J_{zk}$. The utility function consists of the systematic component as defined in Eq. 4.4 and a stochastic component $\epsilon_{zkj}$. The definition of the systematic component depends on the data source as long as not all coefficients are common across SP and RP observations. We denote the RP- and SP-based systematic components by $V_{zkj}^{RP}$ and $V_{zkj}^{SP}$, respectively. The panel is unbalanced since the number of RP choices differs across drivers (the number of SP choices is equal to 10 for all drivers). Also, the number of available alternatives, $J_{zk}$ differs across choices. It equals 2 for SP choices and 17 for RP choices. The indicator function $1^{RP}$ is used to distinguish RP and SP choices and, hence, the corresponding systematic parts of the utility function. It is equal to 0 for SP choices ($k \in \{1,\ldots,10\}$) and equal to 1 for RP choices ($k \in \{11,\ldots,K_z\}$). Moreover, any difference in the variance of $\epsilon_{zkj}$ between SP and RP observations is taken into account by defining a multiplicative scale factor $\lambda$ that is relevant for SP
observations (while the scale of RP observations is fixed to 1).

The random utility function to be maximized is then given by:

\[ U_{zkj} = 1_{RP} \cdot V_{zkj}^{RP} + (1 - 1_{RP}) \cdot \lambda \cdot V_{zkj}^{SP} + \epsilon_{zkj} \]  

(4.9)

For the MNL models, the random component \( \epsilon_{zkj} \) is assumed to follow a Gumbel distribution, with errors distributed identically and independently (iid) across observations. The probability of driver \( z \) choosing alternative \( j \) in choice \( k \), \( P_{zkj} \), is then defined as follows:

\[ P_{zkj} = 1_{RP} \cdot \frac{\exp(V_{zkj}^{RP})}{\sum_{j=1}^{J(zk)} \exp(V_{zkj}^{RP})} + (1 - 1_{RP}) \cdot \frac{\exp(V_{zkj}^{SP})}{\sum_{j=1}^{J(zk)} \exp(V_{zkj}^{SP})} \]  

(4.10)

The corresponding loglikelihood function is given below. For simplicity, we use the notation of \( \bar{P}_{zk} \) to indicate the chosen alternative in choice \( k \) by driver \( z \):

\[ \ln L = \sum_{z=1}^{Z} \sum_{k=1}^{K_z} \ln \left[ \sum_{q=1}^{Q} H_{zq} \left( \prod_{k=1}^{K_z} \bar{P}_{zk|q} \right) \right], \]  

(4.11)

Moreover, we use a panel latent class model, allowing for heterogeneity among drivers and taking into account the panel setup of the underlying data. Latent class models assume that drivers can be sorted into a set of \( q = 1, \ldots, Q \) classes. Preferences can then vary between the classes, while they are assumed homogenous within each class. This means that (some or all) coefficients estimates will be class specific. The term ‘latent’ derives from the fact that heterogeneity is unobserved by the analyst. Class membership of a specific driver as well as the size of each class are unknown in advance. The analyst has only control over the number of classes \( Q \). Hence, in addition to the (class-specific) coefficients, also the driver-specific probabilities of being member of a specific class need to be estimated. In line with the notation used by Greene and Hensher (2003), these are denoted by \( H_{zq} \). While various formulations are possible for the estimation of class membership, we adopt a simple multinomial logit form here (without explanatory variables), which ensures that relative class sizes sum up to 1, without actively restricting the parameter values. The loglikelihood function is then given by the following equation:

\[ \ln L = \sum_{z=1}^{Z} \ln \left[ \sum_{q=1}^{Q} H_{zq} \left( \prod_{k=1}^{K_z} \bar{P}_{zk|q} \right) \right], \]  

(4.12)

where \( \bar{P}_{zk|q} \) is equal to the probability associated with the chosen alternative by driver \( z \) in choice situation \( k \) conditional on driver \( z \) being member of class \( q \). And the multiplicative term \( \prod_{k=1}^{K_z} \bar{P}_{zk|q} \) therefore represents the likelihood of the

---

58 Differences in scale are only relevant if common coefficients for the SP and RP observations are estimated.

59 A general discussion on latent class analysis can for instance be found in McCutcheon (1987).
sequence of choices \( k = 1, \ldots, K_z \) made by driver \( z \), again conditional on class membership.

In contrast to mixed logit models, which are probably the most common method to capture unobserved heterogeneity in discrete choice analysis, the estimation of panel latent class models does not demand any assumptions regarding the shape of the distribution of a given parameter across individuals. Only the number of classes needs to be set by the analyst. Moreover, for a relatively small number of classes, it is usually not necessary to assume that specific coefficients cannot differ across classes. Such an assumption is frequently made in mixed logit models in order to ensure convergence of the respective models.

**Estimation results: Multinomial logit models**

Table 15 gives the results for the multinomial logit models. RP and SP data have not been pooled in these models. The first two models do not take into account the travel time ratio, while the third and the fourth model presented in the table do (using the specifications of the utility functions introduced in Eqs. 4.6 and 4.7 for RP and SP observations, respectively). Robust standard errors are used for the computations of the t-statistics. All valuations are given in Euro per hour.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Without ( \tau )</th>
<th>With ( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_T )</td>
<td>-5.40</td>
<td>-5.03</td>
</tr>
<tr>
<td>( \beta_E )</td>
<td>-1.71</td>
<td>-13.09</td>
</tr>
<tr>
<td>( \beta_L )</td>
<td>-1.54</td>
<td>-11.51</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.21</td>
<td>4.82</td>
</tr>
<tr>
<td>( \beta_R^{\tau} )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_T^{\tau} )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_E^{\tau} )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_L^{\tau} )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>VOT</td>
<td>40.30</td>
<td>43.36</td>
</tr>
<tr>
<td>VSTL</td>
<td>11.49</td>
<td>21.68</td>
</tr>
<tr>
<td>VSTL</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nr. Obs.</td>
<td>2997</td>
<td>4930</td>
</tr>
<tr>
<td>LogLik.</td>
<td>-7336</td>
<td>-2753</td>
</tr>
<tr>
<td>Ps. R² adj.</td>
<td>0.136</td>
<td>0.193</td>
</tr>
</tbody>
</table>

Unlike in most earlier research, we find that the VOT is quite close for the RP- and SP-based models. This might be a consequence of designing the SP choice experiment in a way that the trade-offs are person-specific and closely resemble those faced in reality. The finding that the VOT is relatively high in all models can be explained by the fact that many participants belong to the upper income
classes (see Table 14). The values of schedule delay early and late, however, depend quite strongly on whether RP or SP data are used. So, the VSDE and the VSDL are significantly higher for the SP-based models. A possible explanation is that SP choices represent short-run decisions that are characterized by little or no possibility for re-scheduling and thus relatively high costs of schedule delays (see Chapter 5). This is a direct consequence of the setup of the SP experiment, where drivers are told that the actual travel time realization on the day of travel is entirely random (given the set of possible realizations), and can thus not be anticipated. RP-based departure time choices, in contrast, are likely to be the result of long-run adaptation processes, for instance in anticipation of participation in the reward experiment. Due to this higher degree of flexibility, scheduling restrictions are less binding, and schedule delay costs tend to be lower as a consequence.

Also the effect of including non-linear terms for the travel time and scheduling attributes was tested. This test is important since the attribute values presented in the SP experiment were based on the reported minimum, average and maximum travel times, which turned out to be overestimated. The travel time attributes for the SP survey are thus on average longer than the RP attribute values. If non-linearities in the valuations exist, RP- and SP-based valuations are therefore affected differently. However, while we find that some of the non-linear terms were significant, they did not lead to substantial changes in the VOT, VSDE and VSDL. Furthermore, we find that the relative weight of the current travel time to the past experience, $\theta$ is equal to roughly 0.2, regardless of whether the travel time ratio is taken into account or not. Travel time expectations are thus mainly based on past experience, however, updated by day-specific information.

From the estimation results of the RP-based model that accounts for the travel time ratio (using the ‘standard’ definition for the travel time ratio from Table 11 and the model specification in Eq. 4.6), we do not find convincing evidence that drivers misperceive travel times to a similar extent as they over- (and in few cases under-) report them in the questionnaire. All coefficients attached to the differences between the attribute levels that would result if travel times were misperceived and the objectively measured attributes are insignificant, except for the reward coefficient. However, this finding does not indicate that drivers actually misperceive travel times, not in the least place because it has a negative sign (whereas a positive sign would be expected). Note that the overall VOT is again given by VOT plus VOT$^\tau$.

---

60 This is in contrast to the results obtained in the meta-analysis by Tseng et al. (2005). Other studies, however, do not find any pattern in the relation of RP- and SP-based scheduling values (e.g. Li et al., 2009).

61 As a robustness check, the models were also re-estimated using the alternative definitions of the travel time ratio. Qualitatively the results do not change if the alternative definitions are adopted.

62 One likely explanation is related to the relatively large share of drivers who pass C2 just before the start of the peak (i.e. 6:30). The difference between the reward attribute when travel times are overestimated and the objective reward attribute might then be negative (i.e. 0–4) for one or more quarters before 6:30 (depending on the extent of overestimation). In other words, for a given departure time, one would obtain a reward under objectively measured travel times. However, if travel times were overestimated, the C2 link would only be passed after 6:30, hence during the peak period. The coefficient $\beta_T^L$ might thus capture the propensity of many drivers to pass C2 just before 6:30.
In the SP-based model that accounts for the travel time ratio, $\beta^T$, it is found to be positive, as suggested by Brownstone and Small (2005). However, it is only significant at the 14% level. In the following section, we will estimate a joint SP-RP panel latent class model that allows for heterogeneity among drivers and misperceptions of travel times in the SP domain only. This more sophisticated estimation procedure might help identifying if and to which extent behavior in hypothetical and real-life scheduling are consistent with biases in reported travel times choices.

**Estimation results: Panel latent class model**

We estimate a panel latent class model with three classes ($Q = 3$). Various statistical criteria such as the BIC or the AIC have been suggested as adequate criteria for the selection of the number classes (e.g. Greene and Hensher, 2003). However, Scarpa and Thiene (2005) note that this selection must also account for significance of parameter estimates and be tempered by the analyst’s own judgment on the meaning of the parameter signs. Here, the statistical criteria suggest a higher number of classes. But probably due to the correlations in the attribute values, which are especially evident in the RP setting, many coefficients then assume the opposite sign of what would be expected, leading to negative valuations of time and schedule delays, which are clearly not useful from an economic point of view. Nevertheless, the suggested higher number of classes is an indication that there is a substantial amount of heterogeneity present in the sample of participants.

In the model presented in Table 16, all coefficients are class-specific except for $\theta$, which does not differ substantially between classes, and has therefore been restricted to be equal for all classes. Moreover, except for the reward coefficient, all coefficients are data-source-specific (clearly, $\theta$ is not, as it is only applicable for RP data). Using a likelihood ratio test, it can be shown that restricting the reward coefficient to be equal for SP and RP choices is legitimate. There is one class with a reward coefficient close to 0 (’Class 3’), thus representing drivers who do not take into account the reward neither in RP nor in SP choices. We do not compute the monetary valuations for this class as these would be close to infinity (and negative).

We find that for all classes, the VOT is significantly higher for RP-choices than for SP-choices. This result remains true if the travel time ratio is not taken into account (hence, if $\beta^{SP}_T$ is restricted to 0), as well as if the reward coefficient is not restricted to be equal for RP and SP choices (although to a lesser extent in the latter case). This finding is consistent with earlier studies that compare RP and SP estimates of the VOT.

While we find for all three classes a positive $\beta^{SP}_T$, and therefore negative values for $VOT^T$, all three estimates of $\beta^{SP}_T$ are not significant (just as in the MNL model). So, while the sign of these coefficients is in line with the argument brought forward by Brownstone and Small (2005) that drivers who overestimate travel times in reality will ‘underreact’ to travel times in a hypothetical setting, the size of this effect seems

---

63 The model specification is provided in Eq. 4.7.
to be negligible in our dataset. We can therefore conclude that there is no evidence that travelers overestimate travel times to a similar extent as they overreport them, and that, as a consequence, reported travel times do not correspond to travel time perceptions. This result holds also if a higher number of classes is introduced.

### 4.7 Conclusions

In this study, we first compare reported and observed travel times. We find very robust evidence that on average travelers overstate reported travel times by a factor of ca. 1.5. We are also able to identify some main determinants of the extent of overstating, such as link length and average speed, and to test some of the hypotheses brought forward in the literature on why travelers tend to overstate travel times.

Next, we investigate whether the person-specific travel time ratio affects actual and hypothetical departure time choice behavior, using (multinomial logit and panel latent class) discrete choice models. Revealed preference (RP) and stated preference (SP) data, which have been gathered from the participants of a large-scale peak avoidance experiment, are used for this purpose. From observed scheduling behavior in actual as well as in the hypothetical choice situations, we do not find evidence that drivers overestimate travel times, in particular not to a similar extent as they overreport them. We therefore conclude that the travel time
ratio is mainly a reporting error, and that reported travel times should not be used as an indicator of actual travel time perceptions.

The reporting error (being distinct from travel time misperceptions that translate to actual behavior) can arise for multiple reasons. First, even though the questionnaire explicitly focused on in-vehicle time, drivers might also take into account the time they take to get from their home to their car as well as the time they require to get from their car to their work place when reporting travel times in the questionnaire. Second, the results might also be driven by a social desirability bias. For instance, drivers might assume that it is considered socially desirable to drive slowly (i.e. being more responsible towards other road users, as well as towards the environment), and therefore state longer travel times than they actually experience. Third, a bias might have resulted from stating average travel times (over the last 20 commuting trips), which are not necessarily relevant for actual scheduling behavior, as the latter is usually based on time-of-day-specific expected travel times. Finally, also strategic behavior might play a role in the answers given in the questionnaire. By overstating travel times, the respondents might hope to influence policy makers such that they take some actions in order to decrease travel times.

To summarize, there is strong evidence of ‘overstating’, however, less evidence of ‘overreacting’, neither in SP nor in RP. The results obtained in this study should thus be taken as a strong note of caution with respect to the use of reported travel time data as a measure of actual or perceived travel times. While we show this for the case of travel times, this conclusion can probably be drawn in a much more general sense.
Long-run vs. short-run perspectives on scheduling: Evidence from a revealed-preference experiment among peak-hour road commuters
5.1 **Introduction**

Peak-hour traffic congestion is a major problem in most urban areas, causing substantial economic costs. Policy measures that aim at relieving congestion have therefore received much attention. In particular, road pricing schemes have been proposed, leading to the implementation of pricing both along corridors - one example are high occupancy toll lanes at various locations in the US - as well as within specific urban areas, such as London, Singapore, Stockholm. The expansion of infrastructure capacity is another, more widely applied policy.

Peak-hour congestion usually arises when a large number of people share the preference to be at the same place at the same time. In such cases, individuals face a trade-off in their scheduling decisions: If they depart at the begin or at the end of the peak, they will encounter modest or no congestion; however, they will arrive at their destination at a time quite different from their preferred arrival time. On the contrary, the longest travel times will be faced by those who choose a departure time that results in an arrival time close to their preferred arrival time. Vickrey (1969) was the first to account for such within-the-day dynamics, using a dynamic equilibrium model of queuing at a bottleneck. Small (1982), building on the works of Becker (1965) and DeSerpa (1971) on the allocation of time to non-work activities, included terms representing the disutility of arriving at work earlier or later than desired in the utility function of commuters. He also provided empirical estimates of the willingness-to-pay values to reduce these so-called schedule delays. His work gave rise to numerous studies that aim at deriving monetary valuations of travel time savings and schedule delays, and more recently, also of travel time reliability (e.g. Noland and Small, 1995). Most of them use disaggregate data on departure time choices collected from stated preference (SP) or revealed preference (RP) experiments, or combine data from both sources (e.g. Small et al., 2005). Random-utility discrete choice models are then usually used to estimate the model parameters (McFadden, 1974).

Even though the estimated values of travel time savings (henceforth, the ‘value of time’) and schedule delays vary widely across studies, they are usually found to be substantial. The value of time usually ranges between 20 and 90% of the gross wage rate, while the costs of arriving one minute early tend to be lower, and the costs of arriving one minute late tend to be higher than the costs of an additional minute of traveling (see for instance Small and Verhoef (2007) and Li et al. (2010) for overviews). These time-related costs thus compose a large share of generalized travel costs, and are key ingredients in the evaluation of policy options. To identify the appropriate values of time and schedule delays, and to use them for policy evaluation, a good understanding of the underlying scheduling behavior of travelers is crucial.

In this study, we introduce a distinction between short-run and long-run scheduling decisions. The distinction reflects that commuters typically have more flexibility in adapting their daily schedules in the longer run than in the shorter run. We expect this to be reflected by different valuations of times and schedule delays.

---

64This chapter is based on Peer, Verhoef, et al. (2011).
in the long run compared to the short run. We therefore propose a framework that decomposes scheduling decisions into long-run choices of what we will call ‘routines’, and short-run choices of departure times. Routines are considered fixed in the short run, but are endogenous in the long run. We also distinguish between long-run and short-run travel time expectations, taking into consideration that more information on possible travel time realizations becomes available in the short run. This distinction is an important one to make. Most current models implicitly assume that successive days are exact replicas, which implies that the long-run choice of routines, and the short-run choices of departure time are identical, and could not be meaningfully disentangled. The same is in fact true for stochastic models when these assume that travel time expectations do not vary between days (hence, when ignoring learning or information-updating). In this case, departure time choices will not be adjusted between days, and again, long-run and short-run behavior coincide. But when travel conditions vary between days, we can distinguish between the two concepts, assuming that long-run choices are based on (time-of-day specific) expectations of travel times, whereas short-run choices also consider deviations from these expectations.

The model proposed in this chapter assumes that in the long run, drivers choose a routine arrival time, which we refer to as short-run preferred arrival time (SRPAT). The SRPAT minimizes the sum of costs from expected travel time losses and deviations of the SRPAT from the exogenously determined long-run preferred arrival time (LRPAT), which is defined as the preferred arrival time that would be relevant if no congestion occurred ever. In the short run, drivers are able to forecast travel times more accurately, and in the light of that may choose an actual arrival time that differs from the SRPAT. The definitions we use for the ‘long run’ and the ‘short run’ are chosen to best suit the available data on travel behavior. These cover a four-month period. Decisions that are made in the face of attributes that are constant across this period are referred to as long-run decisions, while decisions that are based on attributes that are day-specific are referred to as short-run decisions. Such an exogenous boundary between the long run and the short run may seem somewhat arbitrary, but may in fact strengthen our results, as it is likely to lead to a conservative estimate of the difference between long-run and short-run valuations of time and schedule delays.\footnote{Alternative definitions of the long and short-run are certainly feasible. For instance, it is possible to endogenize the structural arrival time preferences and let them depend on job choices as well as residential and work locations choices. Some studies that focus on such choices also determine valuations of travel time (e.g Van Ommeren and Fosgerau, 2009; Van Ommeren et al., 2000). However, the resulting values are usually not directly comparable to the values derived in this study, not in the last place because they usually do not distinguish between time and scheduling-related costs.}

The data used in this study are drawn from a large-scale, revealed-preference peak avoidance experiment among commuters in the Netherlands. Participants were able to obtain a daily reward of 4 Euro if they avoided traveling on a specific, frequently congested highway link during morning peak hours. Only few comparable RP experiments that account for departure time choice behavior have been conducted (e.g Börjesson, 2008; Lam and Small, 2001; Small, 1982), mainly because...
they require a setup with a monetary attribute, a toll or a reward, that varies over the time of the day. Indicators of routine (long-run) travel behavior are usually avoided at all when modeling RP choices (Gärling and Axhausen, 2003).

Unlike the datasets used in most existing studies that use RP data to investigate departure time choices, the current dataset has a panel structure, allowing for the possibility to observe commuters over time and to identify both the long-run and short-run dimensions of their scheduling behavior. Moreover, detailed data on observed departure times and travel times are available. We make use of these data to construct long-run travel time expectations, as well as day-specific travel time expectations for the short-run model. This allows us to largely avoid the use of reported data, which are frequently imprecise and often even biased (see for instance Chapter 4 of this thesis for a striking example of overstated travel times). A distinction between long-run and short-run scheduling behavior would probably be much more difficult to achieve using SP data, as the choice experiment would have to be phrased such that participants follow the intention of the researcher in interpreting the choice situations in a long-run sense or in a short-run sense, respectively. A more generic drawback of SP data is that they may be subject to hypothetical biases. These arise if individuals act differently in a laboratory setting than they would in real life, which has been found to be true also for travel choice behavior, too (Hensher, 2010).

We use discrete choice models to analyze the participants’ choices of routines and departure times. We find that drivers value more permanent (long-run) travel time gains substantially higher than short-run gains, for a given trip with a factor ranging from three to five, presumably because these can be exploited better through the adjustment of routines. After all, an incidental minute of time gain can typically be used less effectively than a structural gain of one minute. This result is consistent with the results obtained by Tseng et al. (2011), who find that the value attached to a permanent one-hour travel time gain is 14.5 Euro/hour, compared to 3 Euro/hour for an incidental travel time gain of one hour. On the other hand, we find a substantially higher value attached to less permanent (short-run) changes of schedule delays, with the difference mounting to an order of two to three. This finding is consistent with the notion that scheduling restrictions are normally more binding in the short run. Börjesson (2009) and Börjesson et al. (2012) speculate along similar lines to explain the differences they find between the valuation of planned and unplanned delays in models that are based on SP data and do not distinguish between short-run and long-run values of time and schedule delays.

Our results strongly suggest that individuals apply some sort of inter-temporal optimization that leads to differences in long-run and short-run preferred arrival times and valuations of time and scheduling. The distinction between short-run and long-run in the modeling of consumer behavior in the study of valuation of time and scheduling is therefore a very important one to make, not only in empirical models but also in theoretical equilibrium models of travel behavior, for example as used to assess the social desirability of infrastructure investments or road pricing. Moreover, we believe that the large dispersion of the values of time and schedule delays that is found in the literature may partly be explained by differences in the relative prominence of long-run and short-run definitions of the model attributes.
The outline of this chapter is as follows. In Section 5.2 we discuss literature related to long- and short-run scheduling decisions, while in Section 5.3 we introduce the respective modeling framework. Section 5.4 gives an overview of the data, including the experimental setup and variable definitions. The main estimation results are presented in Section 5.5, and various robustness checks in Section 5.6. Section 5.7 concludes the chapter.

5.2 Related literature

Most existing research on travel behavior ignores the repetitive nature of travel choices (with the most prominent example being commuting travel), which essentially makes the distinction between long-run travel routine and short-run departure time choices useful. This is especially true for studies that attempt to derive monetary valuations from SP and RP choice experiments. To our knowledge, this research is the first one to estimate distinct long-run and short-run valuations of travel time and schedule delays, and therefore to explicitly acknowledge that the underlying departure time choices are made repetitively. Also the distinction between a long-run and a short-run preferred arrival time is new to the relevant literature. Most studies adopt preferred arrival time definitions that comply with the idea of the LRPAT. So, the preferred arrival time is often defined in the context of hypothetical behavior in uncongested conditions (e.g. Börjesson, 2009), or as the official work start time (e.g. Noland and Small, 1995; Small, 1982). Only few studies define the preferred arrival time as based on routine behavior, corresponding to the definition of the SRPAT. One such example is Tseng et al. (2011). However, they use a preferred departure rather than arrival time.

One stream of literature that does consider the repetitive character of many travel choices consists of studies that model how travelers update their travel time expectations and perceptions over time, for instance as a consequence of previous travel experiences or traffic information. Most of them use weighted average, adaptive expectation, or Bayesian approaches to model the expectation updating processes (e.g. Jotisankasa and Polak, 2006). Also bounded rationality approaches such as reinforcement learning (e.g. Arentze and Timmermans, 2003) are employed in some papers. However, besides not deriving monetary valuations of the choice attributes, most of these papers confine themselves to analyzing the changes in short-run travel choices over time, without considering the long-run implications thereof, such as the formation of travel routines. An exception is the conceptual model developed by Bogers et al. (2005), which explicitly takes into account the relation between (short-run) day-by-day learning and the (long-run) formation of travel habits.

Contrariwise, there are various studies that emphasize the long-run domain of travel behavior. Many of them are based in social psychology, and rely on the theories of planned and repeated behavior (Ajzen, 1991; Ronis et al., 1989). They investigate concepts such as habitual travel and inertia, arguing that these arise from a lack of deliberation as an attempt to reduce cognitive efforts. In this stream of literature, travel pattern that are fairly stable over time are thus interpreted at
being driven by bounded rationality. However, several authors acknowledge the
difficulty of disentangling a lack of deliberation from stable behavioral intentions
such as a stable routine (e.g. Gärling and Axhausen, 2003; Thøgersen, 2006).

Our model attempts to combine the short-run and long-run oriented models,
assuming that drivers act as (expected) utility maximizers in both time dimensions.
It adopts an adaptive expectation approach to model short-run travel time expec-
tations, and models the formation of long-run travel routines as an optimization
across time-of-day specific expected travel times and schedule delays. Even though
we find individual travel patterns to be fairly stable across days, we are able to show
that drivers take into account additional day-specific information in their daily
departure time decisions.

5.3 Modeling framework

Main models

In the conventional scheduling model, introduced by Small (1982), the utility of
departing at a specific time \( t \), \( V_t \), is a linear function of four attributes. It depends
on the monetary attribute, a toll or reward \( R_t \), travel time \( T_t \), and the extent of
earliness or lateness with respect to the preferred arrival time, PAT. The last two
attributes are generally referred to as schedule delay early, SDE\(_t\), and schedule delay
late, SDL\(_t\). The subscript \( t \) indicates dependency on the moment of departure \( t \).
More recently, the conventional scheduling model was extended to account for
stochastic travel times (e.g. Bates et al., 2001; Fosgerau and Karlström, 2010; Noland
and Small, 1995). In the formulation that resembles most closely the deterministic
case\(^{66}\), \( R_t \), \( T_t \), SDE\(_t\), and SDL\(_t\) represent distributions which arise from day-to-
day variations in travel times at a given time of the day \( t \). The attributes of the
utility function are then defined in expected terms, denoted by the operator \( E[.]. \)
In a conventional linear model we may denote the coefficients attached to these
attributes by \( \beta_R \), \( \beta_T \), \( \beta_E \) and \( \beta_L \), respectively:

\[
E[V_t] = \beta_R E[R_t] + \beta_T E[T_t] + \beta_E E[SDE_t] + \beta_L E[SDL_t]
\]

where

\[
E[SDE_t] = E[\max[\text{PAT} - t - T_t, 0]]
\]

\[
E[SDL_t] = E[\max[t + T_t - \text{PAT}, 0]]
\]

The term \( 1/\beta_R \) can then be interpreted as the marginal utility of income, if \( R_t \) is
defined as reward (rather than a cost), while \( -\beta_T \), \( -\beta_E \), \( -\beta_L \) indicate the marginal
utility gained from reductions in travel times and schedule delays, respectively.

\(^{66}\)Sometimes the probability of being late is added to the utility function, allowing for an extra
penalty for lateness. Also measures of dispersion are occasionally added, with the underlying idea
that they capture disutility from travel time variability that is not directly related to scheduling, such
as stress or anxiety.
The value of time (VOT) and the values of schedule delay early (VSDE) and late (VSDL) can thus be expressed as follows:

\[
\begin{align*}
VOT &= -\frac{\beta_T}{\beta_R} \\
VSDE &= -\frac{\beta_E}{\beta_R} \\
VSDL &= -\frac{\beta_L}{\beta_R}
\end{align*}
\] (5.2)

In this study, we adapt the standard scheduling model such that it allows for the possibility to study differences between long-run and short-run scheduling behavior. There is reason to expect that these are inherently different in at least two ways. First, in the long run, drivers are able to adapt their routines. Specifically, in a world without traffic congestion, a driver’s preferred arrival time may for instance be 9:00 in the morning, whereas recurring congestion may induce him to establish a routine with a usual arrival time earlier than 9:00. As a consequence, he may make appointments at work earlier, and given those appointments, the actually preferred arrival time also shifts to an earlier time of the day. That is, if the preferred arrival time under uncongested conditions is given as in the above example, but the driver has organized routines such that he expects to arrive around 8:00 and therefore plans a meeting at 8:15, a delayed arrival at 8:30 brings schedule delay late, not early.

Second, also the availability of information on travel times is likely to differ between the long and the short run. In the long run, only average traffic conditions are known, whereas in the short run, more information on actual travel time realizations becomes available. We expect these differences between long-run and short-run scheduling decisions to be reflected also in the valuation of travel time and schedule delays. In particular, we expect that travel time gains are valued higher in the long run, as these can be exploited better through the re-scheduling of routines. Schedule delays, on the other hand, are likely to be valued higher in the short run, as scheduling restrictions tend to be more binding in the short run.

We will thus distinguish between a long-run and a short-run scheduling model. In the long-run, arrival routines are chosen subject to structural arrival time preferences, whereas in the short run, departure times are chosen subject to the arrival routines that result from maximizing long-run utility. In the terminology that we will use, the structural arrival time preference is referred to as long-run preferred arrival time (LRPAT), and the long-run utility maximizing arrival time as the short-run preferred arrival time (SRPAT). Schedule delays in the long-run model are therefore defined as deviations from the LRPAT, while schedule delays in the short-run model are defined as deviations from the SRPAT.

The long-run model, which models the choice of arrival routines, differs from the conventional scheduling model with stochastic travel times (Eq. 5.1) in various respects. First, long-run travel time expectations denoted by the superscript LR are used. Second, rather than actual departure times \( t \), the long-run model explains the choice of arrival time \( a \) as short-run preferred arrival time, SRPAT. As a consequence, schedule delays, and hence the deviations from the LRPAT, become deterministic. We add a term \( C^{LR}[\sigma_a] \) to the equation in order to account for the costs of travel time variability. They capture the expected cost due to incidental delays that drivers face in the short run. \( C^{LR}[\sigma_a] \) is defined as the standard deviation of those travel
times that result in an arrival time at time $a$, $\sigma_a$. This formulation thus uses the standard deviation of travel times for an arrival time exactly at $a$, ignoring that travelers may adjust actual arrival moments through their short-run decisions.\footnote{\textit{Given the correlation between standard deviations at close arrival moments, this may be considered a mild simplification. Moreover, we will not be estimating }$C^{LR}[\sigma_a] \text{ but instead will impute it, for reasons to be outlined shortly.}} The indirect utility for SRPAT = $a$ can thus be written as:

$$E[V_a]^{LR} = \beta_R E[R_a]^{LR} + \beta_T E[T_a]^{LR} + \beta_E SDE_a + \beta_L SDL_a + C^{LR}[\sigma_a],$$  \hspace{2cm} (5.3)

where $SDE_a = \max[LRPAT - a, 0]$ and $SDL_a = \max[a - LRPAT, 0]$

The optimal arrival routine, the SRPAT, is therefore a deliberate choice, which maximizes the long-run utility function:

$$SRPAT = \arg\max_a E[V_a]^{LR}$$  \hspace{2cm} (5.4)

The short-run utility function closely resembles the conventional scheduling model that accounts for stochastic travel times (Eq. 5.1). Arrival routines are fixed in the short run. The SRPAT thus enters the indirect short-run utility function as the relevant anchor point for defining the schedule delays. Moreover, we use short-run travel time expectations denoted by the superscript $SR$, which account for the availability of more up-to-date information on travel time realizations when the departure moment has to be chosen, than what applies in the long-run model. Note that the short-run model assumes that disutility of travel time variability is entirely captured by its impact on expected schedule delay costs.

$$E[V_t]^{SR} = \beta_R E[R_t]^{SR} + \beta_T E[T_t]^{SR} + \beta_E E[SDE_t]^{SR} + \beta_L E[SDL_t]^{SR},$$  \hspace{2cm} (5.5)

where $E[SDE_t] = E[\max[SRPAT - t, 0]]$ and $E[SDL_t] = E[\max[t + T_t - SRPAT, 0]]$

Besides the basic long-run and short-run models of Eqs. 5.3 and 5.5, we introduce an alternative short-run model, which is used exclusively to test whether it is true that the SRPAT is indeed the relevant anchor point in departure time decisions, as we assumed in Eq. 5.5. If this is true, a move away from the SRPAT towards the LRPAT would increase schedule delay costs. Following our earlier example, an arrival time at 8:30, given a SRPAT at 8:00 and an LRPAT at 9:00, would yield costs of schedule delay late rather than early. We therefore re-formulate the short-run model, changing the definition of the schedule delays such that we can determine whether scheduling costs are at their minimum at the SRPAT or at the LRPAT. To do so, we introduce the adapted schedule delays $SDAE_t$ and $SDAL_t$, which capture arrival moments that are early or late with respect to both measures of SRPAT and LRPAT. In addition, we define the intermediate domains $SDME_t$ and $SDML_t$. $SDME_t$ captures arrival moments that are early with respect to the SRPAT but late with respect to the LRPAT, when $LRPAT < SRPAT$. Likewise, $SDML_t$ captures
5.3. Modeling framework

Arrival moments that are late with respect to the SRPAT but early with respect to the LRPAT, when SRPAT ≤ LRPAT. The corresponding utility in this ‘three-domains’ model is denoted by $V_{t}^{SR-3D}$, and the travel time expectations are defined in the same way as in the short-run model:

$$V_{t}^{SR-3D} = \beta R E[R_{t}]^{SR} + \beta T E[T_{t}]^{SR} + \beta E E[SDAE_{t}]^{SR} + \beta L E[SDAL_{t}]^{SR} + \beta ME E[SDME_{t}]^{SR} + \beta ML E[SDML_{t}]^{SR},$$

(5.6)

where

<table>
<thead>
<tr>
<th>LRPAT &lt; SRPAT</th>
<th>SRPAT ≤ LRPAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SDAE_{t} =$</td>
<td>$SDAE_{t} =$</td>
</tr>
<tr>
<td>$\max[LRPAT - t - T_{t}, 0]$</td>
<td>$\max[SRPAT - t - T_{t}, 0]$</td>
</tr>
<tr>
<td>$SDAL_{t} =$</td>
<td>$SDAL_{t} =$</td>
</tr>
<tr>
<td>$\max[t + T_{t} - SRPAT, 0]$</td>
<td>$\max[t + T_{t} - LRPAT, 0]$</td>
</tr>
<tr>
<td>$SDME_{t} =$</td>
<td>$SDME_{t} =$</td>
</tr>
<tr>
<td>$\min[SRPAT - LRPAT, \max[SRPAT - t - T_{t}, 0]]$</td>
<td>$\min[SRPAT - LRPAT, \max[t + T_{t} - LRPAT, 0]]$</td>
</tr>
<tr>
<td>$SDML_{t} =$</td>
<td>$SDML_{t} =$</td>
</tr>
<tr>
<td>0</td>
<td>$\min[LRPAT - SRPAT, \max[t + T_{t} - SRPAT, 0]]$</td>
</tr>
</tbody>
</table>

Consider the case when SRPAT ≤ LRPAT, as depicted in Figure 16: If $\beta ML$ would be equal to $\beta L$, the model would suggest that the SRPAT is the relevant anchor to describe short-run behavior, and the LRPAT has no influence whatsoever in the short-run model. In contrast, when $\beta ML = -\beta E$, the model suggests that the LRPAT is the relevant anchor point, and the SRPAT has no impact on short-run choices. We may expect $\beta ML$ to be somewhere between these polar cases, and would take a significant negative sign as an indicator of the desirability of an arrival at the SRPAT over an arrival at the LRPAT. A similar argument can be made for the case where LRPAT < SRPAT. If both $\beta ML$ and $\beta ME$ are negative, we can confirm that deviations from the SRPAT towards the LRPAT indeed increase scheduling costs. The three domains model, therefore, helps identifying the degree to which the SRPAT rather than the LRPAT indeed determines short-run scheduling behavior.

Figure 16: Diagrammatic exposition: SRPAT ≤ LRPAT
Empirical model specifications

The models discussed in Section 5.3 can be estimated by employing the additive random-utility model developed by McFadden (1974). It assumes that choices between discrete alternatives are made such that the utility of the decision maker is maximized. Choices are probabilistic as they are not only affected by observed attributes, which are captured in the systematic part of the utility function, but also by unobserved attributes, which are captured by a random term. We will apply this method for the estimation of the long-run as well as the short-run models.

We allow arrival routines to be weekday-specific. Therefore, in the long-run model, each driver $z = 1, \ldots, Z$ chooses between $j = 1, \ldots, J$ discrete arrival routines for each weekday $l = 1, \ldots, 5$. In the short-run model the same drivers choose between $j = 1, \ldots, J$ discrete departure time alternatives on days $k = 1, \ldots, K$. The systematic part of the long-run utility function for arrival routine $a$, $E[V_{zl}]^{LR}$, is therefore re-written as $E[V_{zl}]^{LR}$. Similarly, the systematic part of the short-run utility function for departure time $t$, $E[V_{zk}]^{SR}$, is re-written as $E[V_{zk}]^{SR}$.

The corresponding random terms are denoted by $\epsilon_{zlj}$ and $\epsilon_{zkj}$, and the overall utilities, $U_{zlj}^{LR}$ and $U_{zkj}^{SR}$ are thus given by:

$$U_{zlj}^{LR} = E[V_{zl}^{LR}] + \epsilon_{zlj} \quad \text{and} \quad U_{zkj}^{SR} = E[V_{zk}^{SR}] + \epsilon_{zkj} \quad (5.7)$$

We apply the most basic discrete-choice specification and assume that the random terms are iid with the extreme-value distribution. McFadden (1974) showed that then the choice probabilities for alternative $i$, $P_{zl}^{LR}$ and $P_{zk}^{SR}$ respectively, have the logit form:

$$p_{zl}^{LR} = \frac{\exp(E[V_{zli}]^{LR})}{\sum_{j=1}^{J} \exp(E[V_{zlj}]^{LR})} \quad \text{and} \quad p_{zk}^{SR} = \frac{\exp(E[V_{zki}]^{SR})}{\sum_{j=1}^{J} \exp(E[V_{zkj}]^{SR})} \quad (5.8)$$

From this assumption, the multinomial logit (MNL) model arises. The model parameters are then estimated by maximizing the corresponding log-likelihood function, which is defined as the sum of the logs of the choice probabilities across observations. Due to the fact that arrival routines are weekday-specific, and departure time decisions are made on multiple days, multiple observations per driver are included in the long-run as well as short-run datasets. The MNL model ignores this panel nature of the datasets. This assumption of independent observations is probably less drastic for RP data compared to SP data, as the valuation attached to travel time and schedule delays may differ across days also for a given driver. Still, in order to correct for a possible under-estimation of the standard errors, the panel specification of the sandwich estimator is used (e.g. Menard, 2009). Moreover, we will later test whether the results still hold if the models allow for heterogeneity across drivers. For this purpose, we re-estimate the main models using panel latent-class models. These models assume that drivers
can be sorted into a set of $Q$ classes, with coefficient estimates being class specific. It is unknown to the analyst to which class a particular driver belongs to. The estimation procedure differs from the estimation of MNL models mainly by the fact that not only choice probabilities for the alternatives $j = 1, \ldots, J$ need to be determined, but also the probabilities of being member of class $q = 1, \ldots, Q$ (see McCutcheon (1987) for a general discussion on latent class analysis and Greene and Hensher (2003) for an application of latent class analysis to panel data and discrete choice problems).

**Travel time expectations**

Like the arrival routines, also travel time expectations are defined as weekday-specific. In fact, differences in travel time expectations for different weekdays may be an important reason to also differentiate routines over weekdays. We then define the long-run expectation for attribute $A = \{R, T, SDE, SDL\}$ on weekday $l$ as the average attribute value on that weekday over a time period $k = 1, \ldots, K$. The computations therefore implicitly assume that the driver has information about future travel times when choosing the optimal routines. While it is clearly impossible to know travel time realizations on single days (in particular, in the long run), the average weekday-specific travel time seems to be a good representation of the expectation a regular commuter may have about average travel time realizations, also in the future. We thus define an indicator function $1_{l,k}$, which is equal to 1 if it is true that weekday($k$) = $l$, and 0 otherwise. The long-run expectation for attribute $A$, $E[A_{zlj}]^{LR}$, is then given by

$$E[A_{zlj}]^{LR} = \sum_{k=1}^{K} 1_{l,k} A_{zlj} / \sum_{k=1}^{K} 1_{l,k}. \quad (5.9)$$

In the short run, which means at the moment that the actual departure time decision is taken, more up-to-date information on travel time realizations is available compared to the long run, for instance due to a better knowledge of expected weather conditions. We represent this enhanced information availability by computing a weighted average of the long-run expectation of attribute $A$ and the respective attribute value based on the actual travel time realization on the day of travel $l$.

69 The weight attached to the actual travel time realizations, relative to the long-run travel time expectations, is denoted by $\theta$ and will be estimated. Although it is unrealistic that these actual travel times are known to drivers at the moment of departure, they represent a benchmark for the maximum extent of

---

69 According to the terminology used by Jotisankasa and Polak (2006), this procedure can be categorized as an adaptive expectation approach.

70 We also tested methods (i.e. state-space and semi-parametric estimations) that allowed for the expected travel time to depend on day-specific variables that may affect travel times, such as weather conditions. However, we found that the difference between realized and expected travel times barely decreased as a consequence of adding these variables. This can probably be attributed to the fact that only 75 working days are taken into account here, and that weather conditions were fairly stable over that time period.
information they may have at that moment. Again, we define an indicator function $1_{kk}$, which is equal to 1 if it is true that weekday($k$)=weekday($\tilde{k}$), and 0 otherwise.

$$E[A_{z\tilde{k}j}]^{SR} = \theta A_{z\tilde{k}j} + (1 - \theta) \sum_{k=1}^{K} 1_{kk} A_{zkj} / \sum_{k=1}^{K} 1_{kk}$$

(5.10)

5.4 Data

Experimental setting

We use RP data that were gathered during a large-scale peak avoidance (Spitsmijden in Dutch) experiment in the Netherlands. In this experiment, participants were eligible for a monetary reward of 4 Euro per day if they avoided traveling on a specific highway link during the morning peak (6:30–9:30 a.m.). The highway link has a length of 9.21 km and is frequently congested during morning peak hours. We refer to the link as ‘C1–C2’ as it is defined as the road segment between two cameras ‘C’. Rewards could not be earned on weekends or school vacation days, and also not if a driver had already exceeded the maximum number of days per 2-week period for which he could obtain a reward. This maximum is driver-specific, and is based on a driver’s reference behavior, which is defined as the average number of trips he had undertaken along the C1–C2 link per 2-week period before the start of the experiment.

The peak avoidance experiment lasted for more than a year, from November 2008 until December 2009. Participants could join as well as leave the experiment during this period. Overall, about 5000 commuters participated. More than 15000 commuters had been invited to participate in the experiment after they were observed passing the C1–C2 link, resulting in almost 3000 participants. Another 2000 participants were recruited through lease car companies and from an earlier peak avoidance experiment. The selection of the participants is therefore voluntary and not random. From a survey among drivers who were invited to participate but chose not to participate in the experiment, it was found that participants tend to be older, have a higher education and income, and (not surprisingly) are more flexible in choosing their arrival time at work (see also Appendix A of this thesis). Moreover, their reference travel behavior differs. On average, participants travel more frequently along the C1–C2 link, and their passage times of this link are less clustered in the middle of the peak (7:30-8:30 a.m.). We believe that selection effects will affect the observed sensitivity to rewards, but are unlikely to have a systematic impact on the distinction between short-run and long-run decision making that we are interested in here.

Travel times and passage times of the participants have been directly measured along the C1–C2 link using cameras capable of number plate detection. For our analysis, we would like to use door-to-door measures of travel time instead.

---

71 A more detailed overview of this study can be found in Appendix A of this thesis.
72 Highway A12, between Gouda and Zoetermeer with driving direction towards The Hague.
Chapter 3 of this thesis develop a method to approximate driver, day and time-of-day-specific door-to-door travel times as well as departure times from home and arrival times at work, combining speed measurements from loop detectors along the C1–C2 link and GPS-based speed measurements on the home–C1 and C2–work links. They draw on the common finding that speeds are correlated positively across links (for a given day and time of the day). By using geographically weighted regression (GWR), they take into account that this correlation pattern varies over space and is therefore dependent on the home and work location of a driver. It is the travel time calculations from that model that we will be using in the present analysis.

In the analysis performed in this chapter, we focus on drivers who participated in the experiment between begin September 2009 and end of December 2009 \( (K = 75\) working days), and who were observed to use the C1–C2 link at least occasionally during that time period. From these participants, only those are selected who filled in a survey that included questions about their long-run arrival time preference, and for whom we are able to approximate door-to-door travel times given the limited coverage of the GPS-based GWR model developed in Chapter 3. These restrictions leave use with \( Z = 371\) drivers.

**Operationalization of the LRPAT and SRPAT**

For each driver, measures of the LRPAT and the SRPAT need to be determined. We derive the LRPAT from reported data and the SRPAT from actual behavior. By nature of its definition, the LRPAT refers to a guaranteed congestion-free situation, and can therefore not be derived from actual behavior. Instead, the LRPAT is obtained from a questionnaire conducted among the participants of the experiment. They were asked to state their preferred arrival time at work if they knew for sure that they would not face any congestion during their commuting trip.\(^73\)

We consider the SRPAT to be weekday-specific, as travel times differ substantially across weekdays, and drivers are likely to adapt their arrival routines accordingly (Figure 17.1). The SRPAT is defined as the driver and weekday-specific median arrival time at work. For each driver only those weekdays are considered for which at least three passages of the C1–C2 link were observed. Moreover, only those days are taken into account during which a driver was eligible for a reward, to be sure that the monetary incentive indeed applies. If the number of observations per driver and weekday is an even number, we randomly assign one of the two middle values as median. The intuition behind this procedure is that we want to be able to identify those departure time decisions for which the actual arrival time by definition coincides with the corresponding SRPAT. These choices will be excluded from the short-run model, as they are fully endogenous.

Figure 17.2 shows a scatterplot for all drivers and weekdays for which valid measures of the SRPAT and the LRPAT are available. As expected given that long

\(^73\)Drivers with an LRPAT earlier than 6:30 or later than 9:30 were removed from the dataset as these drivers do not have an incentive to travel during the peak period, so that it is more natural to assume that the scheduling model applies to them.
travel times are unattractive, we find that the distribution of the SRPAT is relatively more dispersed in time than the LRPAT. The figure also shows that drivers with a relatively early LRPAT (the median LRPAT is indicated in the figure) tend to choose a SRPAT that is even earlier than their LRPAT, while drivers with a relatively late LRPAT tend to choose an even later SRPAT. This is confirmed by the histograms of the LRPAT and SRPAT distributions, shown in Figures 17.3 and 17.4, respectively. We find that the density of the LRPAT distribution is highest between 8:00 and 9:00, roughly corresponding to the typical work starting times. The distribution of the SRPAT is much flatter. Since rewards can be gained for passing C1–C2 before 6:30, a very pronounced peak can be observed for arrival times at work between 6:30 and 7:00. A relatively small number of participants has a SRPAT that gives them the opportunity to receive a reward after the end of the peak (9:30). These are strong indications that the choice of the SRPAT is indeed the result of a trade-off between deviations from the LRPAT, average congestion patterns, as well as the distribution of the monetary incentive over time of the day.

\[ \text{Figure 17: Descriptives LRPAT and SRPAT} \]
5.5 Estimation results: Main models

Selection of observations and choice set definitions

For both the long-run and the short-run models, the choice set consists of \( J = 16 \) discrete choice alternatives. Each alternative corresponds to a 15-minute interval. In the long-run model, a driver is able to choose between arrival routines that result in an arrival time at work between 6:15 to 10:00. In the short-run model, the choice set is driver-specific. This is to ensure that despite the differences in home–work distances only the most relevant departure time alternatives are included for each driver, in particular those that yield a trade-off with respect to the reward. Thus, for each driver the choice set contains 12 departure time alternatives that in expected terms result in a passage time of the C1–C2 link during the peak, and therefore do not yield a reward, and 4 alternatives that result, again in expected terms, in an off-peak passage time of the C1–C2 link (2 alternatives before, and 2 after the peak).

Table 17 provides some descriptive statistics on the datasets used for the estimation of the long-run and the short-run models. The number of observations in the long-run model is determined by the number of driver-weekday combinations for which valid measures of the SRPAT are available. On average, the number of weekdays for which a SRPAT can be specified, is about three. The short-run dataset then consists of all observations that were used in determining the SRPAT, except for the ones for which the corresponding arrival times have been identified as SRPAT (resulting in an exclusion of 1306 observations). We find that the remaining number of departure time choices per driver is on average about 16. Finally, Table 17 also shows the differences between the actual arrival times that result from the departure time decisions and the SRPAT. Table 17 demonstrates that more than 60% of the departure time choices result in an arrival time that is less than 15 minutes early or late with respect to the SRPAT. Hence, drivers tend to choose their departure times such that they arrive close to their SRPAT.

5.5 Estimation results: Main models

Table 18 shows the results obtained for the long-run and the short-run model, as well as for the 3-domains model that is used to test whether actual departure time decisions are indeed driven by the SRPAT rather than the LRPAT.

For all models, credible point estimates are obtained. The VOT is between 5.20 and 30.16 Euro/hour, the VSDE between 9.34 and 23.16 Euro/hour and the VSDL between 7.22 and 20.22 Euro/hour.\(^{75}\) Significant differences between the long-run and the short-run estimates of the VOT, VSDE and VSDL confirm our main hypothesis that the higher scheduling flexibility in the long run is reflected by the coefficient estimates. We find that travel time is valued six times higher in the long-run model than in the short-run model, indicating that drivers attach a higher

\(^{75}\)It is usually found that the VSDL is higher than the VSDE. The reason why we obtain the opposite result is most likely due to the fact that a large number of participants in this study have a rather early SRPAT. These participants may be quite reluctant to switch to an even earlier departure time, as the utility derived from being at home rather than at work tends to be larger in the early morning, resulting in a high VSDE. See Tseng and Verhoef (2008) for a model that allows scheduling costs to differ by time of the day.
### Table 17: Descriptives datasets

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr. of days ((K))</td>
<td>75</td>
<td>–</td>
</tr>
<tr>
<td>Nr. of drivers ((Z))</td>
<td>371</td>
<td>–</td>
</tr>
<tr>
<td>Nr. of choice alternatives ((J))</td>
<td>16</td>
<td>–</td>
</tr>
<tr>
<td><strong>Long-run dataset</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr. of SRPAT choices</td>
<td>1158</td>
<td>–</td>
</tr>
<tr>
<td>Avg. nr. of SRPAT choices per driver</td>
<td>3.12</td>
<td>1.42</td>
</tr>
<tr>
<td><strong>Short-run dataset</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total nr. of departure time choices</td>
<td>7271</td>
<td>–</td>
</tr>
<tr>
<td>Nr. of excluded choices due to coincidence of SRPAT and arrival time</td>
<td>1306</td>
<td>–</td>
</tr>
<tr>
<td>Remaining nr. of departure time choices</td>
<td>5965</td>
<td>–</td>
</tr>
<tr>
<td>Avg. remaining nr. of departure time choices per driver</td>
<td>16.08</td>
<td>11.21</td>
</tr>
</tbody>
</table>

**Arrival time deviations from the SRPAT**
- more than 30 minutes early: 7.66% –
- between 15 and 30 minutes early: 11.03% –
- between 15 minutes early and 15 minutes late: 62.35% –
- between 15 and 30 minutes late: 9.64% –
- more than 30 minutes late: 9.32% –

### Table 18: Main estimation results

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Long-Run</th>
<th>Short-Run</th>
<th>Short-run: 3 domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_R)</td>
<td>0.22</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>(\beta_T)</td>
<td>-6.56</td>
<td>-0.69</td>
<td>-1.75</td>
</tr>
<tr>
<td>(\beta_E)</td>
<td>-2.03</td>
<td>-2.89</td>
<td>-3.83</td>
</tr>
<tr>
<td>(\beta_L)</td>
<td>-1.57</td>
<td>-2.70</td>
<td>-2.53</td>
</tr>
<tr>
<td>(\beta_{ME})</td>
<td>–</td>
<td>–</td>
<td>-1.45</td>
</tr>
<tr>
<td>(\beta_{ML})</td>
<td>–</td>
<td>–</td>
<td>-3.23</td>
</tr>
<tr>
<td>(\theta)</td>
<td>–</td>
<td>0.43</td>
<td>0.36</td>
</tr>
<tr>
<td>VOT ((\in/h))</td>
<td>30.16</td>
<td>5.20</td>
<td>10.56</td>
</tr>
<tr>
<td>VSDE ((\in/h))</td>
<td>9.34</td>
<td>21.62</td>
<td>23.16</td>
</tr>
<tr>
<td>VSDL ((\in/h))</td>
<td>7.22</td>
<td>20.22</td>
<td>15.27</td>
</tr>
<tr>
<td>VSDME ((\in/h))</td>
<td>–</td>
<td>–</td>
<td>8.76</td>
</tr>
<tr>
<td>VSDML ((\in/h))</td>
<td>–</td>
<td>–</td>
<td>18.54</td>
</tr>
<tr>
<td>Nr. Obs.</td>
<td>1158</td>
<td>5965</td>
<td>5965</td>
</tr>
<tr>
<td>LogLik.</td>
<td>-2681</td>
<td>-10550</td>
<td>-10410</td>
</tr>
<tr>
<td>Pseudo R(^2)</td>
<td>0.17</td>
<td>0.36</td>
<td>0.37</td>
</tr>
</tbody>
</table>
value to more permanent (repetitive) travel time changes, supposedly because these can be exploited better through the rescheduling of routines. Schedule delays are valued two to three times higher in the short-run model than in the long-run model. This is likely because of the more binding nature of short-run scheduling constraints, which renders the re-planning of activities more costly. Moreover, the fact that \( \theta \) is around 0.4 and significantly different from 0, suggests that in the short run drivers indeed have more information on travel time realizations than in the long-run and take this information into account when deciding on their departure time.

Since all scheduling coefficients are negative and significantly different from 0 in the 3-domains model, we conclude that a net scheduling disutility results from moving away from the SRPAT towards the LRPAT. Therefore, the SRPAT is the valid anchor point of the short-run schedule delays. Specifically, we find that \( \beta_E \) is almost three times larger than \( \beta_{ME} \), indicating that schedule delay early is valued higher if one arrives early with respect to both PAT definitions than in the intermediate domain with earliness with respect to the SRPAT but lateness with respect to the LRPAT. While this finding might indicate that drivers still take into account the LRPAT in their departure time decisions, although to a lesser extent than the SRPAT, it may also reflect non-linearities in the valuation of earliness. Regarding the coefficients for schedule delay late, we find that \( \beta_{ML} \) is approximately 1.3 times as large as \( \beta_L \). The costs of being late with respect to both PAT definitions are therefore almost equal to the costs resulting from lateness with respect to the SRPAT but earliness with respect to the LRPAT. We also find that the VOT is significantly higher in the 3-domains model compared to the short-run model (10.56 vs. 5.20 Euro, respectively). A possible reason is that choices that result in arrival times relatively far away from the SRPAT can be better explained when multiple kinks in the schedule delay cost function are present. And these choices are likely to be determined to a large extent by travel time.

In the original equation of the long-run model (Eq. 5.3) a term for travel time variability was included \( (C^{LR}(\sigma_d)) \). However, this term could not be estimated due to the strong correlation between variability and expected travel time. A consequence may be that the relatively high VOT in the long-run model partly reflects the value attached to reliability, and thus to short-run schedule delays. To check whether the difference between the long-run and the short-run VOT persists if we correct for this, we impute the share of the long-run VOT that may be due to variability. For this purpose, we use the relationship between the valuation of reliability (VOR) and the valuation of schedule delays that was established by Noland and Small (1995) and later generalized by Fosgerau and Karlström (2010). Fosgerau and Karlström (2010) showed that for any (standardized) travel time distribution that is constant over the time of the day, the VOR attached to the regarding standardized travel time distribution \( \Phi \) is given by

\[
VOR = (VSDE+VSDL) H(\Phi, \frac{VSDE}{VSDE+VSDL}),
\]

where \( H(\Phi, \frac{VSDE}{VSDE+VSDL}) \) is the so-called mean lateness factor, which is equal to the average lateness conditional on being late. The costs of variability provided that the
travel time distribution $\Phi$ has standard deviation $\sigma$ are then given by $\text{VOR} \cdot \sigma$. This result remains a good approximation if the mean and the standard deviation of the travel time distribution change over the time of the day while the standardized distribution is constant.

To approximate the long-run VOR, we derive for each driver a standardized travel time distribution $\Phi_z$ and mean lateness factor $H(\Phi_z, \text{VSDE}_\text{VSDL})$, using the VSDE and VSDL derived in the short-run model.\footnote{For each driver we form a distribution that contains all $J$ time-of-day-specific standardized travel time distributions based on the long-run travel time expectations. The assumption that travel time distributions are constant over the time of the day is not perfectly fulfilled. However, it was verified that our results hold also if standardized travel times are based either on only peak or off-peak travel time distributions.} These can be interpreted as the values attached to rather incidental variations of travel time and therefore an upper limit to the valuation of travel time variability in the long run. We compute the costs of variability that arise if travel time changes by one hour, in order to be able to compare them directly to the VOT. Therefore, we multiply the VOR by the derivative $\left( \frac{\partial \sigma}{\partial E[T]} \right)_z$, which is computed using a simple OLS regression. On average, $\left( \frac{\partial \sigma}{\partial E[T]} \right)_z$ is found equal to 0.54, confirming earlier research on the positive correlation between mean travel time and travel time variability (see Fosgerau (2010) for a theoretical derivation and Chapter 2 of this thesis for empirical evidence). We then find that the average costs of variability (across drivers) that result when the expected travel time changes by 1 hour is equal to 16.71 Euro. The costs are fairly similar across drivers, with a standard deviation amounting to 1.00 Euro. We can therefore conclude that about half of the long-run VOT as presented in Table 18 should be attributed to the costs resulting from travel time variability. But note that even after removing the costs of variability, the long-run VOT of 13.45 Euro still exceeds the short-run VOT by a factor 2.6.

### 5.6 Estimation results: Robustness checks

**MNL Models**

In this section, various robustness tests of the above results are presented, again using MNL models (as opposed to the latent class models considered in the next section). We implement two tests to ensure that the LRPAT provided by the respondents is indeed a valid measure for the preferred arrival time in the long-run choice, and therefore explains SRPATs better than other measures of the LRPAT. In the first test, we alter the LRPAT for all respondents by $\pm 5, \pm 10, \pm 15, \pm 20, \pm 25, \pm 30$ minutes. Figure 18 shows that indeed the log-likelihood peaks very near the model where the LRPAT is left unchanged, confirming that the LRPAT, although drawn from a questionnaire, indeed seems to represent the drivers’ most desired arrival time from a long-run perspective. In the second robustness check for the long-run model, we divide the observations into two groups: the earliest 50% and the latest 50% of LRPATs. This is to check that the schedule delay cost valuations in the long-run model are not mainly due to early commuters (for schedule delay early),
5.6. Estimation results: Robustness checks

Figure 18: LRPAT ± x minutes

or mainly to late commuters (for schedule delay late). Table 19 shows that lower values of time and schedule delay are obtained for the earlier LRPATs, induced by a higher reward coefficient. This is probably the result of relatively more drivers with an early LRPAT choosing for early off-peak travel routines than drivers with a late LRPAT choosing for late, off-peak travel routines. Despite the differences between the two groups, the result of travel time being valued higher in the long run and schedule delays being valued higher in the short run is still true for both.

We also perform two robustness checks for the short-run model. For the first one, we re-estimate the short-run model as presented in Table 18, with the only difference that travel time expectations are assumed equal to the actual travel time realizations (hence, \( \theta = 1 \)). If this model yields similar results as the standard short-run model, we can conclude that the results of the standard model are not driven by long-run travel time expectations. If the latter was true, differences between long-run arrival routines and actual arrival times would be due to unobserved factors only, rendering the distinction between the long and the short-run model meaningless. The estimation results in Table 19 show that similar values of schedule delay as in the standard short-run model are obtained, indicating that actual travel time realizations are able to explain short-run behavior. The finding that the time coefficient decreases can probably be attributed to the volatile, to some extent unpredictable nature of the actual travel times. This provides an argument for including the long-run travel time expectations as part of the short-run expectations (i.e. \( \theta < 1 \)), assuming the role of a stabilizing element in the formation of short-run travel time expectations. But anyway, the main qualitative differences in relative valuations compared to the long-run model survive.

Finally, we test whether the particular shape of the SRPAT distribution over the time of the day, with a large share of drivers departing before the peak (Figure 17.4), drives the outcome of the short-run model. For this purpose, we
re-estimate the short-run model while applying weights to the observations that are defined as inversely proportional to the density of the SRPAT distribution. The corresponding results are presented again in Table 19. We find that values of schedule delay resulting from this weighted model are close to those of the standard short-run model, while the VOT becomes close to 0. A possible explanation for the latter result could be that the large number of drivers who usually travel before the peak attach a higher weight to travel time compared to the remaining drivers, most of whom travel during the peak. Since the drivers departing before the peak are not weighted heavily in this robustness test, the average VOT drops.

Table 19: Robustness checks

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Long-run LRPAT</th>
<th>Late LRPAT</th>
<th>Short-run ( \theta = 1 )</th>
<th>Inverse Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_R )</td>
<td>0.23 3.61</td>
<td>0.15 2.05</td>
<td>0.15 7.36</td>
<td>0.15 6.33</td>
</tr>
<tr>
<td>( \beta_T )</td>
<td>-5.59 -4.59</td>
<td>-7.18 -5.45</td>
<td>-0.13 -0.50</td>
<td>-0.05 -0.13</td>
</tr>
<tr>
<td>( \beta_E )</td>
<td>-1.80 -5.72</td>
<td>-2.09 -10.67</td>
<td>-2.80 -19.19</td>
<td>-2.70 -20.84</td>
</tr>
<tr>
<td>( \beta_L )</td>
<td>-1.57 -10.20</td>
<td>-1.48 -7.06</td>
<td>-2.65 -21.50</td>
<td>-2.69 -19.48</td>
</tr>
<tr>
<td>( \theta )</td>
<td>- - - -</td>
<td>- - - -</td>
<td>1 - - -</td>
<td>0.47 8.33</td>
</tr>
<tr>
<td>VOT (€/h)</td>
<td>23.97 2.85</td>
<td>47.10 2.02</td>
<td>0.88 0.37</td>
<td>0.36 0.07</td>
</tr>
<tr>
<td>VSDE (€/h)</td>
<td>7.76 3.13</td>
<td>13.74 2.13</td>
<td>19.20 6.85</td>
<td>18.30 5.96</td>
</tr>
<tr>
<td>VSDL (€/h)</td>
<td>6.72 3.47</td>
<td>9.68 2.08</td>
<td>18.14 6.96</td>
<td>18.18 5.93</td>
</tr>
<tr>
<td>Nr. Obs.</td>
<td>579 579</td>
<td>5965 5965</td>
<td>5965 5965</td>
<td>5965 5965</td>
</tr>
<tr>
<td>LogLik.</td>
<td>-1269 -1409</td>
<td>-10592 -10871</td>
<td>-1269 -10871</td>
<td>-1269 -10871</td>
</tr>
<tr>
<td>Pseudo R(^2)</td>
<td>0.21 0.12</td>
<td>0.36 0.34</td>
<td>0.36 0.34</td>
<td></td>
</tr>
</tbody>
</table>

Latent class models

Finally, we estimate panel latent class models to investigate to which extent the results differ from the standard models in Table 18 if we account for unobserved heterogeneity among drivers. Compared to mixed logit models, which are frequently used to capture unobserved heterogeneity, latent class models have the advantage that they do not require distributional assumptions regarding heterogeneity. Moreover, in a latent class setting it is usually not necessary to assume that certain coefficients do not differ across individuals. This assumption is often present in mixed logit models, especially in cases where the variation in explanatory variables is low and correlation between the variables is high, rendering the identification of heterogeneity more difficult. As these characteristics are also present in the datasets used in this study, panel latent class models are used as a simple yet insightful check for the effects of heterogeneity among drivers.

\(^{77}\)The density of the SRPAT distribution is calculated using a normal kernel function with a bandwidth of 0.25.
We re-estimate the standard long- and the short-run model allowing for two latent classes\(^{78}\) in both models (Table 20). In both cases, the log-likelihood of the models improves significantly, implying that the latent class models are able to capture the unobserved heterogeneity among drivers. For the long-run model, we found a large group of drivers (ca. 65% of all drivers) with a reward coefficient close to zero. We therefore fix the reward coefficient to zero for this class, in order to avoid unrealistically high values of time and schedule delay. While it is impossible to determine the monetary valuations of time and schedule delays of this class, it is reassuring that the time coefficient is higher (in absolute size) than the scheduling coefficients in the long-run model, while the opposite is true for the short-run model, confirming the results of the standard model. For the group of drivers who take into account the reward in the long-run model, the values of time and schedule delays are fairly consistent with the results of the standard long-run model in Table 18. The evidence from the short-run model regarding heterogeneity among drivers differs substantially from that of the long-run model. We find similar coefficient estimates for the reward and travel time in both classes, while the scheduling coefficients differ significantly between them. We can identify one group of drivers with very high values of schedule delay (55% of drivers), and another group of drivers whose values of schedule delay are in about the same range as the long-run VSDE and VSDL.

We can therefore conclude that also if we account for (unobserved) heterogeneity between drivers, the consequences of increased flexibility in the long run compared to the short run remain robust. So, the VOT is still found substantially higher in the long run than in the short run. The values of schedule delay seem to be similar between long run and short run for one group of drivers, however, on average they are again significantly higher in the short run.

\(^{78}\) We tested also higher numbers of classes, however, we found the resulting coefficients then to be fairly sensitive to the starting values.

### Table 20: Latent class estimation results

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Long-Run</th>
<th>Short-Run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td>(\beta_R)</td>
<td>0.27</td>
<td>5.53</td>
</tr>
<tr>
<td>(\beta_T)</td>
<td>-10.01</td>
<td>-6.02</td>
</tr>
<tr>
<td>(\beta_E)</td>
<td>-1.47</td>
<td>-6.74</td>
</tr>
<tr>
<td>(\beta_L)</td>
<td>-4.00</td>
<td>-7.92</td>
</tr>
<tr>
<td>(\theta)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Class Prob.</td>
<td>0.35</td>
<td>8.66</td>
</tr>
<tr>
<td>VOT (€/h)</td>
<td>37.44</td>
<td>3.84</td>
</tr>
<tr>
<td>VSDE (€/h)</td>
<td>5.48</td>
<td>3.82</td>
</tr>
<tr>
<td>VSDL (€/h)</td>
<td>14.97</td>
<td>4.19</td>
</tr>
<tr>
<td>Nr. Obs.</td>
<td>1158</td>
<td>–</td>
</tr>
<tr>
<td>LogLik.</td>
<td>-2503</td>
<td>–</td>
</tr>
<tr>
<td>Pseudo R(^2)</td>
<td>0.22</td>
<td>–</td>
</tr>
</tbody>
</table>
5.7 Conclusions

We decompose scheduling decisions of commuters into long-run choices of routines and short-run choices of departure times. Data from a large-scale revealed preference experiment are used. Participants of the experiment were able to gain monetary rewards for not using a specific highway link during the morning rush hour. We find that drivers attach different values to travel time and schedule delays in the long run compared to the short run. Travel times are valued higher in the long run and schedule delays are valued higher in the short run. These findings are consistent with the intuitive notion that commuters are typically more flexible in their scheduling decisions in the longer run than in the shorter run. Travel time gains can thus be exploited better in the long run through the adaptation of routines, while scheduling constraints are more binding in the short run when routines are fixed.

Since the short-run and the long-run shadow prices for a given trip differ by factors ranging from two to five in our basic model, our results may have substantial impacts on optimal choices for transport policies such as pricing and investment. Most importantly, they imply that different values of time and schedule delay should be used in the evaluation of policy options, depending on whether they lead to permanent changes in travel times (e.g. an increase in road capacity) or changes in travel times that occur only under certain circumstances (e.g. incident management). Our results also have implications for the determination of optimal (road) pricing schemes (see Chapter 6 of this thesis). One question that we consider in Chapter 6 concerns the optimal pricing schedule for a congestible facility when preferences take on the form suggested in this study. A particularly interesting aspect of the problem is whether a consistent application of optimal short-run toll schedules, which optimize departure times given fixed routines, lead to a long-run expected toll schedule that optimizes the choice of routines given structural (long-run) arrival time preferences.

This chapter focuses on scheduling decisions concerning morning commute trips. It does not take into account that these are usually embedded in a greater activity pattern, consisting of different types of activities with varying durations, time restrictions and interdependencies in time, space as well as between persons. While there are studies that employ an activity-based rather than a single-trip modeling approach (e.g. Ettema et al., 2007; Timmermans and Zhang, 2009; Zhang et al., 2005), they usually ignore the repetitive nature of many travel choices and do not derive the monetary valuations of the choice attributes. The latter, however, are required for policy evaluation. Future research should therefore aim at estimating these valuations taking into account both that travel choices are repetitive (such as this study does) and that single scheduling decisions are part of a wider activity schedule.
Equilibrium with diverging long-run and short-run scheduling preferences
6.1 Introduction

In this chapter, we introduce a distinction between long-run and short-run scheduling preferences in an equilibrium setting. We assume a standard bottleneck technology that is dynamic in nature. Since the original study by Vickrey (1969), this bottleneck model has become the workhorse model for the analysis of equilibrium and socially optimal timing of usage of congestible facilities (e.g. Arnott et al., 1990, 1993). Applied to traffic congestion, the basic idea is that drivers have common preferences to arrive at a certain place at a certain time, a typical example being the morning rush hour. If the road capacity is not sufficient to accommodate all drivers in such a way that all of them are able to arrive at their preferred moment, a queue will form in front of the bottleneck. In equilibrium, drivers who arrive close to their preferred arrival time will have spent a considerable amount of time queuing in front of the bottleneck, while those who depart early or late in the peak will have faced lower travel times at the cost of arriving earlier or later than desired. The costs resulting from earliness and lateness with respect to the preferred arrival time are commonly referred to as schedule delay costs.

Unlike the standard bottleneck model, the model introduced in this chapter distinguishes between long-run and short-run scheduling decisions. In the long run, commuters decide on their optimal arrival routines, while in the short run, they choose their optimal departure times subject to these arrival routines. Individuals are therefore less constrained in their long-run choices than in their short-run choices. More specifically, this implies that in the long run they are able to optimize their commuting routines, trading off the time-varying average congestion levels over time of the day against deviations from their 'long-run preferred arrival time' (LRPAT). The latter is defined as the preferred arrival time they would have under uncongested conditions, and can be interpreted as a preference that is driven by external factors, which may for instance be biological (such as daylight) or institutional (such as positive temporal agglomeration forces at work). Also the number of working hours and scheduling restrictions arising from activities other than work may affect this long-run preferred arrival time (e.g. Jenelius et al., 2011; Zhang et al., 2005). In the short run, the travel routines that have been chosen in the long run are fixed, and the optimized arrival time from the long-run problem becomes the preferred moment of arrival, which we refer to as 'short-run preferred arrival time' (SRPAT). Daily short-run departure time decisions are thus made in the face of the routines chosen in the long run, taking into account the bottleneck capacity on that particular day and the resulting time-varying congestion levels.

The distinction between short-run and long-run behavior is only relevant when consecutive peaks are not exact replica’s. Otherwise, travelers are likely to end up in a less interesting corner solution where they either equate their LRPAT and SRPAT, or they always choose their departure time such that their arrival time is identical to the SRPAT. To make the distinction useful, we consider a bottleneck with stochastic, day-specific capacity, such that a difference exists between the long-run problem of choosing the SRPAT considering expected travel times, versus the short-run

79 This chapter is based on Peer and Verhoef (2013).
problem of choosing the departure time when the capacity realization is known and the SRPAT is fixed. This reflects that typically more information becomes available in the short run (e.g. Chorus et al., 2006). Our assumption of a bottleneck capacity that varies between days represents situations where changes in road capacity persist over the entire day, for instance as a consequence of severe incidents, lane closures or adverse weather conditions. Comparable representations of capacity fluctuations in a bottleneck setting have also been used in previous studies (e.g. Arnott et al., 1996, 1999; Lindsey, 1995).

Earlier studies of bottleneck congestion did not distinguish between a long- and short-run dimension of scheduling. However, Chapter 5 of this thesis provides empirical evidence that short-run and long-run preferences, and as a consequence also the corresponding scheduling choices, may diverge. The results presented in Chapter 5 suggest that drivers plan their routines to avoid congestion, driving a wedge between the LRPAT and SRPAT. Estimating a scheduling model that distinguishes explicitly between the long run and the short run, it confirms the intuitive notion that the value of travel time is higher in the long run than in the short run, presumably because an incidental time gain can be used less effectively than a structural one. The opposite is true for the values of schedule delays early and late, which may well reflect that scheduling constraints are more binding in the short run than in the long run. They find that the long-run and short-run valuations differ substantially, by factors ranging between 2 to 5. Differences between short-run and long-run shadow prices are also present in the theoretical model introduced in this study.

In this chapter, we do not only characterize the unpriced equilibrium, but also first-best and second-best optima. We show that the first-best optimum can be achieved by levying first-best tolls upon passage of the bottleneck - hence, in the short run - while simultaneously using a long-run pricing instrument to affect the choice of the short-run preferred arrival time. The application of both short-run and long-run pricing instruments may not always be feasible, for instance due to technical or political restrictions. We thus consider also second-best situations where either only short-run or long-run pricing instruments are available. We find that the unpriced equilibrium as well as the first- and second-best optima entail routine arrival times at work (the SRPAts) that differ across drivers.

The welfare implications of dispersed work starting hours have been studied earlier, for instance by Henderson (1981) and Arnott (2007), assuming flow congestion, and by Fosgerau and Karlström (2010), assuming bottleneck congestion. In these chapters, the equilibrium pattern of endogenous preferred arrival times

---

80It is quite straightforward to imagine how a short-run toll can be implemented in practice, namely by charging a (time-of-day- and capacity-dependent) toll at the entry of the bottleneck. The practical implementation of a long-run toll, however, is less straightforward, since the SRPAT can usually not be directly observed nor affected by policy makers. As a consequence, in reality, long-run tolls might have to be levied in a more indirect way. One example would be to use financial incentives to shift day-care and school starting times to off-peak hours, possibly resulting in off-peak travel routine choices of the parents. Also, for some groups of people such as public sector employees, the SRPAT can more easily be observed and therefore also be influenced by policy makers through pricing instruments.
at work is driven by positive agglomeration externalities that increase in the number of individuals who are simultaneously present at work. Our model shows that the consideration of equilibria with dispersed work starting hours does not necessarily require the presence of agglomeration economics, but may also follow from distinguishing between long-run and short-run scheduling decisions.

In the context of our analysis it is natural to define the long run as the time frame where travel routines are chosen. This may be different from other settings, in particular those that involve also residential or employment choices (e.g. De Vany and Saving, 1982; Van Ommeren and Fosgerau, 2009; Van Ommeren et al., 2000). Our focus is motivated by our aim to analyze the distinction between long-run and short-run scheduling decisions in the framework of the standard bottleneck model.

Using a model structure that is close to the standard formulation of the bottleneck model enables us to compare the equilibrium solutions of the model introduced here to the solutions obtained in the standard model.

The structure of this study is as follows. Section 6.2 introduces the bottleneck model that distinguishes between long-run and short-run scheduling decisions. Section 6.3 characterizes the unpriced equilibrium. Section 6.4 provides a discussion on social optima in a bottleneck model that distinguishes between long-run and short-run scheduling preferences. Section 6.5 characterizes the first-best optimum and Sections 6.6 and 6.7 the second best optima if only short-run or long-run tolls are available, respectively. Section 6.8 concludes. Various mathematical proofs are contained in appendix 6.A of this chapter.

6.2 Modeling framework

Introduction

Applied to the morning peak hour, the standard bottleneck model assumes that every day a fixed number of $N$ identical commuters travel from home to work. All of them have the same route, which includes the passage of a single bottleneck with a fixed capacity $s$. The capacity level $s$ therefore denotes the maximum number of vehicles that can pass the bottleneck per unit of time. Drivers pass the bottleneck in the order of their departure time from home. If the departure rate from home exceeds $s$, a so-called ‘vertical’ queue grows, meaning that spill-back effects are ignored. Without loss of generality, the free-flow time required to travel from home to work is normalized to 0. Consequently, home-work travel times consist only of the queuing time in front of the bottleneck.

When deciding on their optimal departure time, commuters trade off costs of travel delays and costs of schedule delays. The former indicate the costs associated with travel time losses, while the latter are defined as the costs of earliness and lateness of the actual arrival time relative to the preferred arrival time. The costs of a trip through the bottleneck then depend linearly on travel times and schedule delays, whereby schedule delay costs assume a piecewise linear function, which allows the costs attached to being one minute early to differ from the costs that result from arriving one minute late. The unit costs associated with travel delays,
schedule delay early and schedule delay late are denoted by $\alpha$, $\beta$ and $\gamma$, respectively. The unit cost parameters as well as the preferred arrival time are often assumed identical across drivers (e.g. Arnott et al., 1990, 1993).

In the bottleneck model that is introduced in this chapter, long-run and short-run scheduling decisions are distinguished. More specifically, long-run decisions reflect choices of travel routines, whereas short-run decisions represent departure time choices. Just as in the standard bottleneck model, drivers trade off travel delay and schedule delay costs both in the long run and in the short run. The distinction requires two adaptations to the standard formulation of the bottleneck model.

First, as argued in the introduction of this chapter, a distinction between short-run and long-run scheduling decisions is not useful if days are exact replicas. Therefore, in our model, the bottleneck capacity can assume $i = 1, \ldots, J$ possible discrete values, each of which is realized with probability $p_i$. The regarding capacity levels are then denoted by $s_i$. Clearly, it must hold that $\sum_{i=1}^{J} p_i = 1$. As in the models of Lindsey (1995) and Arnott et al. (1996, 1999), these capacity levels are day-specific, and therefore do not vary during a given day. We assume that in the long run, commuters only know these probabilities, while in the short run (hence, before deciding on their departure time on a specific day) they know the actual realization of the bottleneck capacity.

Second, we distinguish between two different preferred arrival times: the 'long-run preferred arrival time' (LRPAT) and the 'short-run preferred arrival time' (SRPAT). The LRPAT is exogenously given, and is assumed to be identical across all drivers. The SRPAT, in contrast, is endogenous and may differ across drivers. It represents the preferred arrival routine. We introduce a function $Z(t)$ that describes the cumulative distribution of SRPATs over time of the day. The corresponding density function is denoted by $\dot{Z}(t)$.

The relation between the LRPAT and the SRPAT is established in the long-run model, where drivers choose their SRPAT as a function of their LRPAT. Drivers will choose a SRPAT that differs from their LRPAT if the long-run scheduling costs - due to deviating from the LRPAT- are counterbalanced by lower costs due to shorter travel delays (in the long and/or short run) or schedule delays (in the short run). In reality, this may for instance translate to the situation where a commuter with a LRPAT at 9:00 chooses a routine arrival time at work (his SRPAT) at 7:00, in order to avoid lengthy average travel times, when these are higher at 9:00 than at 7:00. Given the traffic conditions on a specific day, he may choose a departure time that results in an arrival time different from the SRPAT. For example, on a day with low road capacity, she may depart from home such that she arrives at work already at 6:30.

**Short-run scheduling decision**

Short-run decisions are analogous to the decisions represented in the standard bottleneck model. Therefore, when we will later on compare the results of the bottleneck model that distinguishes between long-run and short-run scheduling decisions with the results of the standard bottleneck model, we can refer to the results of the latter by using the results presented in this section.
The short-run costs of passing the bottleneck for a driver with a SRPAT equal to \( t \), \( C_{SR}(t, s_i) \), consist of travel delay (‘queuing’) costs, \( C_{SD}(t, s_i) \), and schedule delay (‘scheduling’) costs, \( C_{SD}(t, s_i) \):

\[
C_{SR}(t, s_i) = C_{SD}(t, s_i) + C_{SD}(t, s_i)
\]  

(6.1)

Note that in contrast to earlier studies where \( t \), besides being the time index, usually denotes the timing of the departure time decisions, \( t \) is used here as a short-hand for the SRPAT. We adopt this notation for all cost, price and toll functions for which \( t \) serves as an argument. This renders it easier to integrate short-run and long-run scheduling decisions using a common notation. Moreover, we add \( s_i \) as function argument – not only to the cost functions but also to departure and arrival rates, and the starting and end time of the queue – in order to emphasize the capacity-dependency of the short-run equilibrium.

While not added as an argument explicitly, short-run costs of course depend on the cumulative distribution of SRPATs, \( Z(t) \); specifically, on the relation between the density function of the SRPATs, \( Z(t) \), and the bottleneck capacity \( s_i \). So, unless the density of SRPATs, \( Z(t) \), is smaller than (or equal to) \( s_i \) for all time instances \( t \) between the earliest SRPAT, \( t_f \), and the latest SRPAT, \( t_f' \), the equilibrium outcome will entail queuing. In the following analysis, we distinguish these two cases: \( Z(t) > s_i \) (Case 1) and \( Z(t) \leq s_i \) (Case 2) (for all \( t_f \leq t \leq t_f' \)). Other cases will only be discussed briefly, as they turn out to be irrelevant in the analysis of our model.

**Case 1: Density of SRPATs exceeds \( s_i \) for all \( t_f \leq t \leq t_f' \)**

If \( Z(t) \) is consistently steeper than the cumulative arrivals at work, \( A(t, s_i) \), and therefore intersects \( A(t, s_i) \) only once (at a moment in time that will be denoted by \( t^* \)), Hendrickson and Kocur (1981) showed that, regardless of the exact shape of \( Z(t) \), the cumulative departures from home, \( D(t, s_i) \), will be such that there is only one time interval where the queue in front of the bottleneck will grow (\( (t_q(s_i), t^*) \)), and another one where the queue will dissipate (\( (t^*, t_q(s_i)) \)); \( t^*_q(s_i) \) thus denotes the start of the peak, and \( t_q(s_i) \) the end of it. Between these two time instances, the bottleneck operates at its maximum capacity. Hence, all drivers except the first and last one departing experience queuing. The following conditions need to be satisfied in equilibrium:

\[
A(t^*, s_i) = Z(t^*)
\]  

(6.2a)

\[
t_q(s_i) - t_q(s_i) = \frac{N}{s_i}
\]  

(6.2b)

\[
D(t_q(s_i), s_i) = A(t_q(s_i), s_i) = Z(t_q(s_i)) = 0
\]  

(6.2c)

\[
D(t^*_q(s_i), s_i) = A(t^*_q(s_i), s_i) = Z(t^*_q(s_i)) = N
\]  

(6.2d)

Eq. 6.2a provides the definition of \( t^* \) as the intersection point of \( Z(t) \) and \( A(t, s_i) \), and Eq. 6.2b defines the duration of the peak as the ratio between \( N \) and \( s_i \). Eq. 6.2c states that at the time the queue starts to form \( t_q(s_i) \), no driver has yet passed the bottleneck, and all drivers have a SRPAT equal or later than \( t_q(s_i) \). At
the other end, Eq. 6.2d states that at the time the queue has disappeared, \( t_q'(s_i) \), all drivers \( N \) must have passed the bottleneck, and none of them has a SRPAT later than \( t_q(s_i) \).

One of the properties that has been shown to hold in the standard model, and that is therefore also valid in the short-run model here, is that in the case when drivers have different preferred arrival times (and are identical otherwise), the equilibrium order of departure is undetermined (Daganzo, 1985; Lindsey, 2004; Smith, 1979). This is a direct consequence of the linear formulation of the cost function.\(^{81}\) In all subsequent analyses, we make the assumption that drivers pass the bottleneck in order of increasing SRPAT. Although this equilibrium is not unique, it is equivalent to other equilibria in terms of costs, both in the aggregate and for every driver individually.

Schedule delays for a driver with a SRPAT equal to \( t \) are then defined as the difference between \( t \) and the actual arrival time. The latter is given by \( t_q(s_i) + Z(t)/s_i \), because the bottleneck is active since \( t_q(s_i) \) and drivers are assumed to arrive in order of their SRPAT. All drivers with a SRPAT between \( t_i \) and \( t_f \) arrive early, while all with a SRPAT between \( t^* \) and \( t_f \) arrive late. Depending on whether a driver arrives early or late, the unit costs of early, while all with a SRPAT between \( t \) and \( t_f \) arrive in order of their SRPAT. All drivers with a SRPAT between \( t \) and the actual arrival time. The latter is given by \( t_q(s_i) \) and drivers are assumed to arrive in order of their SRPAT. All drivers with a SRPAT between \( t \) and the actual arrival time.

\[ C_{SD}^{SR}(t, s_i) = \begin{cases} \hat{D}(t, s_i) & \text{if } t_1 < t \leq t^* \\ \hat{D}(t, s_i) & \text{if } t^* < t < t_f \end{cases} \]

\[ \hat{D}(t, s_i) = \begin{cases} \frac{s_i}{a - \beta} & \text{if } t_1 < t \leq t^* \\ \frac{s_i}{a + \gamma} & \text{if } t^* < t < t_f \end{cases} \]

Travel times are defined as the difference between arrival and departure times.\(^{83}\)

The corresponding cost, \( C_{SR}^{SR}(t) \) for a driver with a SRPAT equal to \( t \), is obtained

\[ \begin{align*}
\hat{D}(t, s_i) &= \frac{s_i}{a - \beta} & \text{if } t_1 < t \leq t^* \\
\hat{D}(t, s_i) &= \frac{s_i}{a + \gamma} & \text{if } t^* < t < t_f 
\end{align*} \]

\[ \text{Eq. 6.3} \]

\[ \text{Eq. 6.4} \]

\(^{81}\)Due to the linearity of the cost function, drivers are indifferent between arrival times in the interval \([t_q(s_i), t^*]\), and in the interval \([t^*, t_q'(s_i)]\).

\(^{82}\)We assume that \( a > \beta \), which is in accordance with empirical findings (e.g. Small, 1982). Without this assumption, cost equality among equal drivers can only be established if a mass departure of drivers takes place at \( t_q(s_i) \).

\(^{83}\)Departure times can be derived by solving the equations \( Z(t) = \hat{D}(t, s_i)(t - t_q(s_i)) \) (if \( t_1 < t \leq t^* \)) and \( Z(t) = N - \hat{D}(t, s_i)(t_q(s_i) - t) \) (if \( t^* < t < t_f \)) with respect to \( t \). It therefore follows that the departure times for a driver with a SRPAT = \( t \) are given by \( t_q(s_i) + \frac{a - \beta}{a} Z(t)/t \) (if \( t_1 < t \leq t^* \)) and \( t_q'(s_i) - \frac{a + \gamma}{a} N - Z(t)/t \) (if \( t^* < t < t_f \)), respectively. Subtracting them from the respective arrival times \( t_q(s_i) + \frac{Z(t)}{t} \) and \( t_q'(s_i) - \frac{N - Z(t)}{t} \) results in the travel times inputed in Eq. 6.5.
by multiplying travel times by parameter $\alpha$:

$$C_{TR}^{SR}(t, s_i) = \begin{cases} \alpha \left( \frac{\beta Z(t)}{s_i} \right) = \beta \frac{Z(t)}{s_i} & \text{if } t_l < t \leq t^* \\ \alpha \left( \gamma \frac{N-Z(t)}{s_i} \right) = \gamma \frac{N-Z(t)}{s_i} & \text{if } t^* < t < t_l' \end{cases} \quad (6.5)$$

From the conditions given in Eq. 6.2b–6.2d, the equilibrium travel delay and queuing costs (Eqs. 6.3 and 6.5) and the fact that the first driver and the last driver must face equal costs in equilibrium\(^8\), we can derive the following equilibrium results for the relative share of drivers who arrive before their SRPAT, $\theta$ (hence, $Z(t^*) = \theta N$), as well as the start and the end time of the queue.

$$\theta = \frac{\gamma}{\beta + \gamma}, \quad t_q(s_i) = t^* - \theta \frac{N}{s_i} \quad \text{and} \quad t_q'(s_i) = t^* + (1 - \theta) \frac{N}{s_i} \quad (6.6)$$

As a next step, we can then derive total travel delay and scheduling costs in equilibrium, which we denote by $TC_{TR}^{SR}(s_i)$ and $TC_{SD}^{SR}(s_i)$:

$$TC_{TR}^{SR}(s_i) = \frac{\delta}{2} \frac{N^2}{s_i}, \quad \text{where} \quad \delta = \frac{\beta \gamma}{\beta + \gamma} \quad (6.7)$$

$$TC_{SD}^{SR}(s_i) = \int_{t_l}^{t_l'} C_{SD}^{SR}(t, s_i) \dot{Z}(t) dt \leq \frac{\delta}{2} \frac{N^2}{s_i} \quad (6.8)$$

We cannot give a closed-form analytical expression for total scheduling costs $TC_{SD}^{SR}(s_i)$, since we have not explicitly defined a distribution of SRPATs, $Z(t)$. However, we can use the setting where all drivers have equal SRPATs as a benchmark. Then, all drivers face equal costs in equilibrium, and total travel delay and scheduling costs are equal. For obvious reasons, scheduling costs are at their maximum in that case. For any setting with dispersed preferred arrival times, total scheduling costs will thus be lower than total travel delay costs.

Figure 19 provides an example of a bottleneck where the bottleneck operates at full capacity throughout the peak, showing cumulative departures and arrivals in equilibrium. Travel times are represented by the horizontal difference between cumulative departures and arrivals; and schedule delays by the horizontal difference between the cumulative arrivals and the cumulative distribution of SRPATs, $Z(t)$.

**Case 2: Density of SRPATs is smaller than (equal to) $s_i$ for all $t_l \leq t \leq t_l'$**

The bottleneck congestion technology implies that total costs become 0 if the density of SRPATs, $\dot{Z}(t)$ is below the bottleneck capacity $s_i$ for all time instances between $t_l$ and $t_l'$. Both queuing and scheduling costs are equal to 0 then, as each driver is able to arrive at her SRPAT without queuing. It must therefore hold that:

$$D(t, s_i) = A(t, s_i) = Z(t) \quad (6.9)$$

$$C_{TR}^{SR}(t, s_i) = C_{SD}^{SR}(t, s_i) = 0$$

---

\(^8\)This follows from the rationale that the driver with a SRPAT equal to $t^*$ must be willing to exchange with both the first and the last driver (see also Footnote 81).
6.2. Modeling framework

Figure 19: Standard bottleneck model

Other cases

Besides the equilibria that entail congestion throughout the entire peak, or no congestion at all, one can also imagine equilibria where the queue does not start with the first driver but only after some drivers have arrived under uncongested conditions. This is the case if $A(t, s_i)$ and $Z(t)$ intersect multiple times. While the start and end of the queue will change in such a setting, the optimal departure rates for drivers who depart and arrive under congested conditions (see Eq. 6.4) still remain valid also in this case (e.g. Newell, 1987).

Long-run scheduling decision

Drivers decide on their travel routine by minimizing overall costs, $EC(t)$.$^85$ Their long-run choices thus determine the distribution of SRPATs, $Z(t)$. The overall costs consist of long-run costs as well as (equilibrium) short-run costs. In accordance with empirical findings of Chapter 5, the long-run values of travel time and schedule delay early and late may differ from the corresponding short-run valuations. The long-run values are denoted by $a\alpha, b\beta, c\gamma$, respectively, where $a, b, c$ thus reflect the ratios of long-run and short-run costs.

In the long run, only the probability distribution of capacity realizations is known, rather than the actual capacity realization as in the short run. Since we assume that drivers perceive the probability distribution of capacities correctly,

---

$^85$ Just as in the short-run cost functions, we do not explicitly add $Z(t)$ as a function argument, in order to keep the notation simple.
long-run travel delay costs are a function of expected short-run travel times. More specifically, long-run travel delay costs, \( C_{LR}^T(t) \), differ from expected short-run travel delay costs \( EC_{SR}^T(t) \) only by parameter \( a \). The overall travel delay costs that determine the choice of the SRPAT, \( C_T(t) \) are then equal to:

\[
C_T(t) = EC_{SR}^T(t) + C_{LR}^T(t) = (1 + a)EC_{SR}^T(t) \tag{6.10}
\]

On the basis of intuition and empirical estimates obtained in Chapter 5 and by Tseng et al. (2011), which show that the costs related to one hour of queuing are higher in the long-run model (taking into account overall costs) than in the short-run model, one might expect \( \alpha > 0 \).

Note that here, as well as in the rest of the chapter, we define the expectation operator attached to any capacity-dependent (and hence, short-run) function \( f(s_i) \), \( Ef \), as the weighted average across the capacity levels \( s_i \) \((i = 1,\ldots,J)\), with the weights being defined as the probabilities of occurrence, \( p_i \):

\[
Ef = \sum_i^J p_i f(s_i) \tag{6.11}
\]

Next, we define the costs related to long-run schedule delays, \( C_{LR}^{SD}(t) \), as the deviations of the SRPAT \( t \) from the LRPAT, evaluated by \( b\beta \) or \( c\gamma \), depending on whether they concern earliness or lateness (with respect to the LRPAT). For the sake of notational convenience we set the LRPAT, which is assumed to be identical across drivers, at 0. Based on the results obtained by Börjesson (2009), Börjesson et al. (2012), and Chapter 5 of this thesis, and the intuition that delays are less costly if they are known far in advance and thus allow for adjustments in one's schedule, it is expected that \( 0 < b < 1 \) and \( 0 < c < 1 \).

\[
C_{LR}^{SD}(t) = \begin{cases} 
-t b \beta & \text{if } t_l < t \leq 0 \\
-t c \gamma & \text{if } 0 < t < t_l'
\end{cases} \tag{6.12}
\]

Finally the overall cost function can be stated as follows:

\[
EC(t) = C_T(t) + C_{SR}^{SD}(t) + C_{LR}^{SD}(t) \tag{6.13}
\]

**Further assumptions**

In order to maintain a simple model structure, we assume in the subsequent analysis of the unpriced equilibrium and the social optima that only two different realizations of capacity levels are possible. The lower capacity state, denoted by \( s_{\text{min}} \), occurs with probability \( p \), and the higher state, \( s_{\text{max}} \), occurs with probability \( 1 - p \). Moreover, the analyses below assume that \( b \) equals \( c \). Drivers thus attach the same value to long-run schedule delays relative to short-run schedule delays for both earliness and lateness, which is a rather realistic assumption judging by the estimates presented in Chapter 5. We denote this common scheduling parameter by \( g \). Moreover, we focus on the parameter range of \( p < g < 1 \), which leads to
the most insightful solutions, and is consistent with empirical findings of \( g < 1 \). Outside this range, mostly corner solutions arise. Finally, for reasons that will become clear in the next section (see in particular Footnote 88), the parameters are set such that the following inequality holds: 

\[ a < \frac{g}{p} - 1 < a_{s_{\text{max}}}^{s_{\text{min}}} \]

Some figures will be added as an illustration of the analytical results in the following sections. These assume the following parameter values: 

\[ N = 1000, s_{\text{min}} = 10/\text{min}, s_{\text{max}} = 20/\text{min}, \alpha = 10 \text{ Euro/h}, \beta = 5 \text{ Euro/h}, \gamma = 15 \text{ Euro/h}, a = 0.4, g = 0.8, p = 0.5. \]

The long-run costs of travel delay are thus 14 Euro/h and the long-run schedule delay costs 12 Euro/h. The duration of the peak \( (N/s_i) \) is 100 min in the \( s_{\text{min}} \) state, and 50 min in the \( s_{\text{max}} \) state (if the bottleneck operates at its maximum capacity throughout). The short-run unit cost parameters have been chosen such that the usual relation \( \beta < \alpha < \gamma \) holds (e.g. Small, 1982). For the long-run unit cost parameters, the values are specified in a rather conservative way, understating the differences between long-run and short-run values by factors 2–5 that were found in Chapter 5. If the differences were assumed larger in the theoretical model, again corner solutions would be obtained for many instances.

### 6.3 Unpriced equilibrium

In the unpriced equilibrium (denoted by superscript \( E \)), each driver chooses the SRPAT in an attempt to minimize the sum of expected short-run and long-run costs of traveling through the bottleneck, \( EC_E(t) \) (Eq. 6.13). Since drivers are identical in their valuations of travel time and schedule delays, and have the same LRPAT, they must face equal values of \( EC_E(t) \). The cost equality condition is therefore satisfied if the derivative of the expected costs in Eq. 6.13 with respect to the SRPAT \( t \), \( dEC_E(t)/dt \), equals 0. From the resulting differential equation we can obtain an expression for the equilibrium density of SRPATs, \( \dot{Z}_E(t) \) (see Appendix 6.A.1 for the derivations). We find that in equilibrium, queuing only occurs in the \( s_{\text{min}} \) state, and is absent in the \( s_{\text{max}} \) state. The equilibrium density of SRPATs in equilibrium, \( \dot{Z}_E(t) \), is then given by:

\[ \dot{Z}_E(t) = \frac{g - p}{ap} s_{\text{min}} \]  

Eq. 6.14 shows that \( \dot{Z}_E(t) \) is constant, implying that the SRPATs are uniformly distributed. Moreover, \( \dot{Z}_E(t) \) is proportional to \( s_{\text{min}} \), which is a natural result as \( s_{\text{min}} \) determines short-run scheduling and queuing costs, while these costs are equal to 0 in the \( s_{\text{max}} \) state (where no queuing takes place). A similar reasoning holds for the finding that \( \dot{Z}_E(t) \) is a decreasing function of \( p \): The higher the probability that the \( s_{\text{min}} \) state occurs, the flatter and therefore closer to \( s_{\text{min}} \) \( \dot{Z}_E(t) \) will be. Moreover, \( \dot{Z}_E(t) \) is increasing in \( g \) and decreasing in \( a \). Naturally, relatively high-long-run scheduling costs (i.e. a high \( g \)) lead to a steeper \( \dot{Z}_E(t) \), as deviations

---

86 For instance, if \( g \) was larger than 1, the social optimum would imply that each driver has a SRPAT equal to his LRPAT. On the other side, if \( g \) was smaller than \( p \), the social optimum would entail a density of SRPATs equal to the low capacity state, and therefore no queuing even in the absence of tolling. The underlying argumentation can be found in derivation of the first-best optimum.

87 Note that \( a < \frac{g}{p} - 1 < a_{s_{\text{max}}}^{s_{\text{min}}} \) holds for this parameter combination: \( 0.4 < 0.6 < 0.8 \).
from the LRPAT become more costly. Higher long-run travel delay costs (i.e. a high \(a\)), on the other hand, render a relatively flat \(Z^E(t)\) necessary in order to compensate for the high travel delay costs for drivers with a SRPAT close to their LRPAT.

For the cost equality condition to be satisfied, \(Z^E(t)\) must then intersect cumulative arrivals at the LRPAT (i.e. \(t^* = 0\)) both in the \(s_{\text{min}}\) as well as in the \(s_{\text{max}}\) state, resulting in \(\theta N\) drivers who have a SRPAT that is earlier than their LRPAT, while the remaining \((1 - \theta)N\) drivers have a SRPAT that is later than their LRPAT (see again Appendix 6.A.1 for the derivations). This means that drivers with a SRPAT that is earlier than their LRPAT always arrive early (in the \(s_{\text{min}}\) state) or on time (in the \(s_{\text{max}}\) state). This result is closely related to the finding in the standard bottleneck framework that \(\theta N\) drivers arrive before their preferred arrival time (Eq. 6.6). It follows directly from \(Z^E(0) = \theta N\) and the linearity of \(Z^E(t)\) that the timing of the earliest SRPAT, \(t^E_1\), and the latest SRPAT, \(t^E_{1'}\), must be equal to the inverse of \(Z^E(t)\) times \(-\theta N\) and \((1 - \theta)N\), respectively.

\[
\begin{align*}
  t^E_1 &= \frac{-\theta N ap}{(g - p)s_{\text{min}}} \\
  t^E_{1'} &= (1 - \theta)N \frac{ap}{(g - p)s_{\text{min}}} 
\end{align*}
\] (6.15)

It is straightforward to show that drivers do not have an incentive to shift their SRPAT when these equilibrium solutions for the density of the SRPATs and the timing of the first and last SRPAT prevail. For instance, the driver with the earliest SRPAT (who also departs first and hence does not face queuing) has no incentive to switch to an earlier SRPAT (a shift size denoted by \(\Delta\)), as this would cause him to lose \(g \beta \Delta\) from moving away from the LRPAT, while gaining only \(p \beta \Delta\) for decreasing the short run schedule delay costs. Since we assumed \(g\) to be larger than \(p\), losses would prevail. If he moved his SRPAT to a later moment in time, his costs cannot decrease due to the cost equality condition under which \(Z^E(t)\) has been derived.

Figure 20 shows an example of an equilibrium situation, with queueing in the \(s_{\text{min}}\) and no queuing in the \(s_{\text{max}}\) state.

Finally, we can specify the expected total costs faced by the commuters in equilibrium, \(ETC^E\). We can split them up into costs related to overall travel times (both long- and short-run), \(ETC_T\), short-run schedule delay costs, \(ETC_{SR}\), and long-run schedule delay costs, \(TC_{SD}^{ELR}\) (see Appendix 6.A.2 for the analytical derivations):

\[
\begin{align*}
  \delta N^2 \frac{p}{2} \frac{1 + a}{s_{\text{min}}} + \delta \frac{N^2}{2} \frac{p}{s_{\text{min}}} \left(1 + a - \frac{ap}{(g - p)s_{\text{min}}} \right) + \delta \frac{N^2}{2} \frac{p}{s_{\text{min}}} \left(1 + a \right) = \delta N^2 \frac{p}{2} \frac{1 + a}{s_{\text{min}}} \right) + \frac{\delta N^2}{2} \frac{p}{s_{\text{min}}} \left(1 + a \right)
\end{align*}
\] (6.16)

\(^{88}\) Note that here as well as in the further analysis, we focus on the case when the parameters of the model are such that \(s_{\text{min}} < Z^E(t) < s_{\text{max}}\). It is easy to show that this inequality holds if the parameters are chosen such that \(a < \frac{g}{p} - 1 < a s_{\text{max}} s_{\text{min}}\).
6.4 Social optima: An Introduction

We find that half of total costs are due to travel delay costs. Also this outcome is closely related to the regarding expression in the standard bottleneck model, for which the same result holds if all drivers have the same preferred arrival time. The regarding total costs are then given by $\delta N^2/(2s_i)$ (see Eqs. 6.7 and 6.8). It is straightforward to show that $ETC^E$ converges to the solution obtained in the standard bottleneck framework if $a$ goes to 0, and $p$ goes to 1. The proportionality of the cost function with respect to $p$ can be attributed to the finding that drivers incur costs only if $s_{min}$ is realized, which happens with probability $p$. The factor $(1 + a)$, on the other hand, is due to the additional long-run travel delay costs, which are not present in the standard bottleneck model. The second half of total costs consists of scheduling costs. The relative shares being attributed to short-run and long-run scheduling costs are dependent on the inverse of $Z^E(t)$. Clearly, the steeper $Z^E(t)$, the smaller are the long-run schedule delay costs.

6.4 Social optima: An Introduction

The unpriced equilibrium as derived in the previous section is not an efficient outcome, because it entails queuing when capacity is low, and the average private costs that drivers face for traveling through the bottleneck are only half the marginal social costs that they cause. These average private costs can be derived by dividing...
total costs in equilibrium ($E_{TC}^E$, see Eq. 6.16) by the number of drivers $N$, while the marginal cost are defined as the derivative of $E_{TC}^E$ with respect to $N$:

$$\delta N \frac{P}{s_{\text{min}}}(1 + a) < \delta 2N \frac{P}{s_{\text{min}}}(1 + a)$$

(6.17)

The difference between marginal social costs and private costs is referred to as marginal external congestion cost, which arises because drivers do not internalize the costs they impose on other drivers by contributing to overall congestion. The social (first-best) optimum can be achieved if all drivers face the marginal social costs that result from their scheduling decisions. Pigou (1920) was the first to show that the social optimum can be decentralized by applying tolls that are equal to the marginal external congestion cost. In the standard bottleneck model (with $Z(t) > s_i$ for all $t_1 \leq t \leq t_1'$), this can be achieved by levying a time-varying toll, which follows exactly the pattern of travel delay costs in the unpriced equilibrium. The first-best toll is thus 0 for the first and the last driver, and largest $(\beta \theta N / s_i)$ for the driver who arrives exactly at his preferred arrival time (the $\theta N$th driver). In the standard bottleneck model, drivers will then arrive at the same time as in the unpriced equilibrium; however, without facing any queuing delay, while the bottleneck will operate at its capacity throughout the peak. Travel delays are thus always a deadweight loss in the bottleneck model, as they can be reduced without increasing scheduling costs.

In our model that distinguishes between long-run and short-run scheduling decisions, two types of pricing instruments are conceivable: long-run tolls and short-run tolls. Both can vary freely over the time of the day. The former would vary with the choice of the SRPAT, and the latter with the choice of departure time. We will consider the first-best situation where both instruments are available, as well as second-best optima, where only one of these two pricing instruments is available, for instance for political or technical reasons.

Short-run tolls are levied at the bottleneck. They depend on both the realized bottleneck capacity as well as the departure time chosen by a specific driver (for the realized capacity state), and are denoted by $\tau_{SR}(t, s_i)$. We differentiate between two different forms of short-run pricing. The first one are the conventional 'first-best short-run tolls', which are by definition equal to the marginal external costs. These tolls are only relevant if congestion would occur without the application of the toll; otherwise they are equal to 0. First-best short-run tolls can be determined without the regulator knowing about the underlying long-run choice process that gives rise to the distribution of SRPATs. The reason is that just as queuing costs, first-best short-run tolls are independent of $Z(t)$ as long as it holds that $Z(t) > s_i$ for all time instances between $t_1$ and $t_1'$.

In addition to the first-best short-run tolls, we define a second form of short-run tolling, which we refer to as 'complementary short-run tolls'. We use this label to refer to tolls that are levied on days where the capacity of the bottleneck is high enough such that no queuing would occur in the absence of tolls. We will show that under specific conditions it is welfare-improving to levy such tolls in addition to
first-best short-run tolls, since they can be used to affect the long-run choice of the SRPAT such that schedule delay costs are minimized. To set the complementary short-run tolls optimally, the regulator must therefore be aware of the long-run choice problem of the drivers.

In contrast to the short-run tolls, long-run tolls are independent of the bottleneck capacity, and only depend on a driver’s SRPAT. They are denoted by \( \tau^{LR}(t) \). The interpretation behind such a long-run pricing instrument is that the regulator levies a tax on the choice of the routine work starting time (i.e. the SRPAT).  

Finally, the expected price function \( EP(t) \) can be defined. For the first-best optimum it consists of the overall costs \( EC(t) \) (Eq. 6.13) as well as expected short-run and long-run tolls:

\[
EP(t) = EC(t) + E\tau^{SR}(t) + \tau^{LR}(t)
\]  

(6.18)

This price function can be adjusted easily for the second-best optima, leaving out one of the pricing instruments. Similar to the cost equality condition in the unpriced equilibrium and for the same reasons (drivers share a common LRPAT and attach identical values to reductions in travel delays and schedule delays), expected prices must be equal across drivers if first- and second-best optima are decentralized. We will show that the social optima derived in the following sections again imply a uniform distribution of SRPATs. Consequently, the relative share of drivers who have a SRPAT earlier than their LRPAT must again be equal to \( \theta \), and the timing of the earliest SRPAT and the latest SRPAT are given by 

\[
t_l = -\frac{\theta N}{\dot{Z}(t)}
\]

and 

\[
t_l' = (1-\theta) N / \dot{Z}(t),
\]

respectively.

6.5 First-best optimum

Specification of the optimum

In this section, we will first characterize the social optimum (denoted by superscript \( F \)), and then derive the tolls required to achieve it. Since it is feasible to levy first-best short-run tolls, the optimum entails no queuing. It is easy to see that scheduling costs are minimized if the density of SRPATs, \( \dot{Z}^F(t) \), is equal to the higher capacity state \( s_{max} \). Starting from that, a \( \dot{Z}^F(t) \) below \( s_{max} \) would induce a decrease in aggregate short-run schedule delays (evaluated at \( p\beta \) and \( p\gamma \)) per unit of time adjustment of the SRPAT) and an increase in long-run schedule delays (evaluated at \( g\beta \) and \( g\gamma \)). These changes in short-run and long-run aggregate schedule delays are equally big, but since we assumed that \( p < g \), the value of the decrease in short-run scheduling costs does not outweigh the increase in long-run scheduling costs. At the same time, an increase in \( \dot{Z}^F(t) \) above \( s_{max} \) would induce a decrease in aggregate long-run schedule delays (again evaluated at \( g\beta \))

\footnote{We do not worry here about the realism of such a tax (see also Footnote 80); what is of interest to us is the question of how it would be set if it were available.}

\footnote{This is true if \( s_{min} \leq \dot{Z}^F(t) \). But any decrease of \( \dot{Z}^F(t) \) below \( s_{min} \) is inefficient for obvious reasons, inducing unnecessarily high long-run scheduling costs, while not decreasing short-run schedule delay any further.}
and $g\gamma$), and an increase in aggregate short-run schedule delays of the same size (now evaluated at $\beta$, because short-run schedule delays would then result for both capacity states). Since $g < 1$ is assumed, the decrease in long-run scheduling costs does not outweigh the increase in short-run scheduling costs. The socially optimal density of SRPATs is therefore equal to $s_{\text{max}}$, and therefore higher than in the unpriced equilibrium:

$$\dot{Z}_F(t) = s_{\text{max}}$$  \hfill (6.19)

A higher concentration of SRPATs in the optimum than in the no-toll equilibrium may seem counterintuative, given the standard notion that optimal pricing would lead to a more dispersed traffic pattern over the day. The intuition behind the results is that the stronger concentration of SRPATs is combined with an elimination of queuing. The optimal concentration of SRPATs therefore results from a trade-off between scheduling costs only; the free-market concentration adds a desire to avoid the peak because of travel delay costs on top of these schedule delay components.

The expected total (social) costs corresponding to the first-best optimum, $ETC_F$, consisting of short-run and long-run scheduling costs, $ETC^{\text{SR}_{\text{SD}}}_F$ and $TC^{\text{LR}_{\text{SD}}}_F$, are then given in Eq. 6.20. They can be derived in a similar way as the total costs in the unpriced equilibrium (see, Appendix 6.A.2), with the only difference being that in the unpriced equilibrium queuing costs are 0, and $\dot{Z}(t)$ is equal to $s_{\text{max}}$ (instead of $\dot{Z}_E(t)$).

$$\frac{\delta N^2}{2} p \left( \frac{1}{s_{\text{min}}} - \frac{1}{s_{\text{max}}} \right) + \frac{\delta N^2}{2} \frac{g}{s_{\text{max}}} = \delta \frac{N^2}{2} \left( \frac{p}{s_{\text{min}}} + \frac{g - p}{s_{\text{max}}} \right)$$  \hfill (6.20)

Short-run scheduling costs are increasing in $p$. This is again an intuitive outcome since short-run scheduling costs only arise in the $s_{\text{min}}$ state, which occurs with probability $p$. Moreover, short-run scheduling costs decrease in $s_{\text{min}}$ and increase in $s_{\text{max}}$. This outcome is not unexpected either. If $s_{\text{min}}$ increases and thus becomes closer to $s_{\text{max}}$, each driver is able to arrive closer to his SRPAT (distributed with density $s_{\text{max}}$), decreasing short-run scheduling costs. An increase in $s_{\text{max}}$ leads to exactly the opposite result. It can furthermore be shown that the costs derived for the first-best optimum again approach the corresponding costs for the standard bottleneck case, if parameters are set accordingly. So, if the long-run scheduling parameter $g$ is assumed equal to 0, and $\dot{Z}_F(t)$ is set equal to infinity (indicating that all drivers have equal SRPATs), long-run scheduling costs approach 0 and total scheduling costs approach the scheduling costs found in the standard model for the case that all drivers have an identical preferred arrival time: $\delta N^2/2s_i$ (Eq. 6.8).

Total (social) costs in the first-best case are less than half of total costs in the unpriced equilibrium (for the parameter ranges considered here). Besides the

---

91 It can be easily shown that the social optimum entails lower costs than the unpriced equilibrium: $ETC^F < ETC^E \Leftrightarrow \frac{g}{p} - 1 < (1 + 2a) \frac{s_{\text{max}}}{s_{\text{min}}}$. Since we assumed $\frac{g}{p} - 1 < a \frac{s_{\text{max}}}{s_{\text{min}}}$ (see Footnote 88) and $a > 0$, $\frac{g}{p} - 1 < (1 + 2a) \frac{s_{\text{max}}}{s_{\text{min}}}$ holds too.
elimination of both short-run and long-run queuing costs, also the sum of short and long-run scheduling costs is lower in the first-best optimum than in the unpriced equilibrium. The reason for this result, which is different from the solution found for the standard bottleneck case (where scheduling costs are equal in the unpriced and the first-best optimum), is that the dispersion of SRPATs in the unpriced equilibrium is higher than socially optimal ($\dot{Z}_E(t) < \dot{Z}_F(t)$).

**Tolls**

If $s_{\text{max}}$ applies, first-best short-run tolls will be equal to 0 for all time instances. Since $Z_F(t)$ is equal to $s_{\text{max}}$, also without toll no queuing would occur in this case. The first-best short-run tolls for the $s_{\text{min}}$ state are set such that the queuing that occurs under unpriced conditions is eliminated. As in the standard bottleneck model, this entails that the tolls are equal to short-run travel delay costs (Eq. 6.5). Consequently, the departure rate becomes equal to the capacity of the bottleneck (in this case $s_{\text{min}}$) and all queuing disappears:

$$
\tau_F^{\text{SR}}(t, s_{\text{min}}) = \begin{cases} 
\beta Z_F(t) &= \text{if } t_i < t \leq 0 \\
\gamma N - Z_F(t) &= \text{if } 0 < t < t_{\text{f}}' 
\end{cases}
$$

(6.21)

However, if only first-best short-run tolls were levied, drivers would not choose for the socially optimal density of SRPATs, i.e. $\dot{Z}_F(t) = s_{\text{max}}$, since the expected price of traveling through the bottleneck would then differ across drivers, rendering the equilibrium under tolls as in Eq. 6.21 inefficient (see Appendix 6.A.3 for the derivations). In particular, if $\dot{Z}_F(t)$ were equal to $s_{\text{max}}$, the driver with the SRPAT equal to LRPAT would face the lowest expected price, while the drivers with the earliest and latest SRPAT, respectively, would face the highest one. A long-run toll is thus applied in addition to the short-run toll to bridge this gap, and hence to reach full efficiency. The optimal long-run toll, $\tau_F^{\text{LR}}(t)$, assumes the following shape (note that $g$ is assumed to be larger than $p$, meaning that $\tau_F^{\text{LR}}(t)$ will always be positive):

$$
\tau_F^{\text{LR}}(t) = \begin{cases} 
(g - p)\beta Z_F(t) &= \text{if } t_i < t \leq 0 \\
(g - p)\gamma N - Z_F(t) &= \text{if } 0 < t < t_{\text{f}}' 
\end{cases}
$$

(6.22)

Figure 21 provides the graphical intuition for why both short-run and long-run pricing instruments are required to reach the full optimum. Each instrument by itself is insufficient to equalize the price across across drivers. Only if both of them are used, the social optimum can be decentralized as the sum of scheduling costs and tolls is equal for all drivers, both in the short and the long run.

### 6.6 Second-best situation: Short-run toll only

As discussed above, we consider two second-best optima, both of which are characterized by the availability of only one type of pricing instrument: either...
Chapter 6. LR vs. SR in equilibrium

Figure 21: First-best tolls and schedule delay costs

a long-run or a short-run pricing instrument. In contrast to other possible second-best situations that are frequently considered in the literature, we assume that the available instrument is not restricted in its form as long as its short- or long-run character, respectively, is not altered (e.g. the long-run toll cannot become capacity-specific).

92 An overview of alternative second-best optima, for instance involving the restriction that the toll cannot be varied freely over time of the day, can be found in Small and Verhoef (2007, pp. 137–148).

Specification of the optimum

We first consider the second-best situation, denoted by $S$, when only short-run tolls are available. Again these can be used to fully eliminate queuing. As we argued for the first-best optimum, the remaining costs, the sum of (expected) short-run and long-run scheduling costs, is minimized if the SRPATs are distributed with density $s_{\text{max}}$ (Eq. 6.19). We will show that $Z^S = s_{\text{max}}$ can be achieved by introducing a complementary short-run toll in the $s_{\text{max}}$ state, leading to the same welfare level as in the first-best optimum (Eq. 6.20):

$$Z^S(t) = Z^F(t) = s_{\text{max}}$$

$$\text{ETC}^S = \text{ETC}^F$$

Tolls

As in the first-best optimum, first-best short-run tolls are levied in the $s_{\text{min}}$ state (Eq. 6.21). Moreover, a complementary short-run toll is introduced in the $s_{\text{max}}$ state in order to affect the choice of the SRPATs such that $Z^S(t)$ becomes equal to $s_{\text{max}}$, maximizing social welfare. The complementary short-run toll must be set such that the expected price in this second-best situation becomes equal to the expected price that drivers face in the first-best optimum. This can be attained by replacing the long-run toll of the first-best equilibrium by an appropriate combination of short-run tolls. Since the short-run toll in the $s_{\text{min}}$ state should be set exactly such that it eliminates queuing while keeping the departure rate at $s_{\text{min}}$, only the toll
in the $s_{\text{max}}$ state can be used for this purpose: Because the $s_{\text{max}}$ state occurs with probability $1 - p$, $\tau^S(t, s_{\text{max}})$ should be equal to the contribution of the long-run toll in the first-best optimum, $\tau^{F, LR}(t)$, divided by $1 - p$:

$$\tau^S(t, s_{\text{min}}) = \tau^{F, SR}(t, s_{\text{min}})$$

$$\tau^S(t, s_{\text{max}}) = \frac{1}{1 - p} \tau^{F, LR}(t)$$  (6.24, 6.25)

Recalling from Eq. 6.22 that the slopes of $\tau^{F, LR}$ are $(g - p)\beta$ and $-(g - p)\gamma$, the toll in Eq. 6.25 will be consistent with the short-run optimum ensuing that every driver arrives at her SRPAT, as long as $g < 1$. This is true by assumption, reflecting that the unit cost of schedule delay cannot be smaller in the short run, when there is less flexibility, than in the long run.

Figure 22 gives a graphical overview of the tolls and schedule delay costs in this second-best optimum. Long-run and short-run schedule delay costs are the same as in Figure 21. This holds true also for the sum of schedule delays and tolls. However, unlike in Figure 21, the sum of tolls consists only of the weighted average of short-run tolls rather than both long-run and short-run tolls.

**Figure 22: Tolls and schedule delays if only short-run tolls are available**

---

### 6.7 Second-best situation: Long-run toll only

**Specification of the optimum**

If long-run tolls are the only pricing instrument available to the regulator (a second-best situation which we denote by $L$), queuing can only then be eliminated fully, if the density of SRPATs, $\dot{Z}(t)^L$, is equal or smaller than $s_{\text{min}}$. In this case, also short-run scheduling costs are equal to 0, as each driver can arrive at his SRPAT. However, the downside is that this low density of SRPATs induces high total long-run scheduling costs. A higher $\dot{Z}(t)^L$, on the other hand, decreases long-run scheduling costs, but leads to queuing, and to higher short-run scheduling costs. In fact, we find that for the parameter ranges under consideration, these trade-offs yield two
possible local second-best optima. The parameter values determine which of these is the most efficient one. The first optimum has \( \dot{Z}(t)^L \) equal to \( s_{min} \) (Case 1), while the second one has \( \dot{Z}(t)^L \) equal to \( s_{max} \) (Case 2). Both possible optima are corner solutions, with the density of SRPATs in the unpriced equilibrium, \( \dot{Z}^E(t) \), being located between.

The existence of two local optima arises from the discontinuity in total travel delay cost at \( \dot{Z}(t) = s_{min} \), where all queuing disappears while travel delay cost in state \( s_{min} \) would be independent of \( \dot{Z}(t) \) as long as it exceeds \( s_{min} \). For any constant \( \dot{Z}(t) \) between \( s_{min} \) and \( s_{max} \), it is easy to see that total schedule delay cost increase in \( \dot{Z}(t) \), as they decrease in \( N \dot{Z}(t) \) by a marginal amount \( N \frac{1}{2} (g - p)(\theta \beta + (1 - \theta) \gamma) \). This suggests \( \dot{Z}(t) = s_{max} \) would be optimal. But it is the said discontinuity that may make \( \dot{Z}(t) = s_{min} \) optimal as well.

**Case 1: \( \dot{Z}(t)^L = s_{min} \)**

In this case, drivers do not face any queuing nor short-run delays. Total costs \( TC^L \) thus consist only of total long-run schedule delay costs \( TC_{SD}^L \) that arise for \( \dot{Z}(t)^L = s_{min} \). These can be obtained by integrating long-run scheduling costs (Eq. 6.12) across all drivers, starting from the driver with the earliest SRPAT, \( t_i = -\theta N/s_{min} \), and ending at the driver with the latest \( t_f = (1 - \theta) N/s_{min} \).

\[
TC_{SD}^L = T C^L = \delta \frac{N^2}{2} \frac{g}{s_{min}}
\]  

(6.26)

**Case 2: \( \dot{Z}(t)^L = s_{max} \)**

If \( \dot{Z}(t)^L \) is equal to \( s_{max} \), drivers face queuing as well as short-run schedule delays if \( s_{min} \) but not if \( s_{max} \) is realized. Expected total costs, \( ETC^L \), are thus composed of expected (short-run and long-run) queuing costs, \( ETC_{LR}^L \), short-run schedule delay costs, \( ETC_{SD}^{LR} \), and long-run schedule delay costs \( TC_{SD}^{LR} \). Note that the expected total costs differ from the total costs in the first-best optimum (Eq. 6.20) only by the queuing costs, which are absent in the first-best optimum:

\[
\begin{align*}
ETC_{LR}^L &= \frac{\delta N^2}{2} \frac{g}{s_{max}} + \frac{\delta N^2}{2} \left( \frac{1}{s_{min}} - \frac{1}{s_{max}} \right) \\
TC_{SD}^{LR} &= \delta \frac{N^2}{2} \frac{g - p}{s_{max}} + p g \frac{N^2}{2} \frac{2 + a}{s_{min}} + \frac{\delta N^2}{2} \frac{g - p}{s_{max}} + p g \frac{N^2}{2} \frac{2 + a}{s_{min}}
\end{align*}
\]

(6.27)

Clearly, if the parameters have been chosen such that \( TC^L \) is lower for Case 1, the second-best optimum entails that \( \dot{Z}(t)^L \) equals \( s_{min} \). In the opposite case, hence when total costs are lower for Case 2, \( \dot{Z}(t)^L = s_{max} \) applies. From the comparison of cases we find that Eq. 6.26 is smaller than Eq. 6.27 so that the long-run policy seems to eliminate queuing when \( a \) is sufficiently large (i.e. the long-run costs of travel delays should be sufficiently high), \( g \) sufficiently small (i.e. the long-run schedule delay costs should be sufficiently low), and \( p \) sufficiently large. The exact condition is

\[
\frac{g - p}{s_{max}} + p g \frac{N^2}{2} \frac{2 + a}{s_{min}} \leq \frac{(2 + a)s_{max} - s_{min}}{s_{max} - s_{min}}.
\]
6.8 Conclusions

Tolls

Finally, we can derive the long-run pricing instruments required to achieve the second-best distribution of SRPATs.

Case 1: \( \dot{Z}(t)^L \) equals \( s_{\text{min}} \)

If \( \dot{Z}(t)^L \) equals \( s_{\text{min}} \), drivers only face long-run schedule delay costs. The corresponding toll \( t^L(t) \), which ensures that all drivers face equal (expected) prices, must therefore be highest for the driver who has a SRPAT equal to the LRPAT, and thus long-run schedule delay costs of 0. They decrease linearly towards the first and the last driver at a rate equal to the increase in long-run schedule delay costs: \( g\beta \) and \( g\gamma \), respectively.

\[
\tau^L(t) = \begin{cases} 
  g\beta \frac{\dot{Z}(t)}{s_{\text{min}}} = g\beta \left( \frac{\theta N}{s_{\text{min}}} + t \right) & \text{if } t_L < t \leq 0 \\
  g\gamma \frac{N - \dot{Z}(t)}{s_{\text{min}}} = g\gamma \left( \frac{(1-\theta)N}{s_{\text{min}}} - t \right) & \text{if } 0 < t < t' 
\end{cases}
\] (6.28)

Case 2: \( \dot{Z}(t)^L \) equals \( s_{\text{max}} \)

For the case that \( \dot{Z}(t)^L \) is equal to \( s_{\text{max}} \), the expected costs, \( EC(t) \), are highest for the driver who has a SRPAT equal to the LRPAT. The corresponding toll, \( t^L(t) \), which again ensures that all drivers face equal prices, is therefore lowest for this driver. Note that this unusual result of a toll that is highest for the drivers that depart first and last, respectively, is driven by the fact that in the model that distinguishes long-run and short-run scheduling decisions, long-run travel delay costs are evaluated at \( a \alpha \) rather than \( a \), rendering it relatively more costly to have long travel times (and hence to have a SRPAT close to the LRPAT). The tolls can be derived by taking the derivative of the cost function, based on the assumption that \( \dot{Z}(t)^L \) equals \( s_{\text{max}} \) (see Eq. 6.31 in Appendix 6.A.1). The toll that provides for price equality among drivers must then be equal to \((-1)\) times the derivative:

\[
\tau^L(t) = \begin{cases} 
  t\beta \left( g - p - \frac{ap s_{\text{max}}}{s_{\text{min}}} \right) & \text{if } t_L < t \leq 0 \\
  t\gamma \left( \frac{ap s_{\text{max}}}{s_{\text{min}}} - g + p \right) & \text{if } 0 < t < t' 
\end{cases}
\] (6.29)

Note that from the assumptions made on the parameter ranges, it follows that \( \tau^L(t) \) is always positive for all \( t_L \leq t \leq t' \).

6.8 Conclusions

In this chapter, we develop a bottleneck model that distinguishes between long-run decisions on travel routines and short-run decisions on departure times, with an application to the morning commute. The bottleneck capacity varies between days, and can either assume a high or low capacity state. We assume that in the long run only the distribution of capacities is known by the drivers, whereas in the short run they are informed about the exact realization of the bottleneck capacity.
on a specific day. Our model incorporates the intuitive notion that, in the face of congestion, people may change their schedules such that the desired arrival time at work deviates from what would be the most desired moment if congestion would not exist.

We show that in the unpriced equilibrium, routine arrival times at work, which we refer to as short-run preferred arrival times (SRPATs) and which are chosen by the drivers in the long run, are uniformly distributed in time, and therefore different from the long-run preferred arrival time (LRPAT), which is identical for all drivers by assumption. Congestion occurs only in the low capacity state, whereas it is absent in the high capacity state.

We also characterize first- and second-best optima, the latter being defined by a limited availability of pricing instruments. We examine how these can be decentralized by applying short-run and long-run tolls. While short-run tolls are used to affect departure time choices, long-run instruments are used to affect the choice of the routine arrival time at work. Both instruments have in common that they can vary by time of the day. However, while short-run tolls depend on the bottleneck capacity realized on a specific day, long-run tolls are capacity-independent.

We show that just as in the unpriced equilibrium, the first-best optimum implies a uniform distribution of SRPATs. However, the extent of dispersion is lower than in the unpriced equilibrium. This is surprising, as conventional wisdom tells us that a greater dispersion of desired arrival times (work start times), would be desirable if congestion exists. The first-best optimum can be reached by simultaneously applying first-best short-run and long-run tolls. First-best short-run tolls as standalone pricing instrument are thus insufficient for reaching the socially efficient outcome. We find that the same level of welfare as in the first-best optimum can be attained if only short-run tolls are feasible. However, this second-best situation requires that, in addition to the first-best short-run toll that is applied in the low-capacity state, tolls are levied also on days when the high-capacity state is realized, and thus on days when – even without tolling – no congestion would occur; we refer to these tolls as complementary short-run tolls. Moreover, we investigate the case when only long-run tolls are feasible, and find that the social optimum can no longer be reached under this restriction.

Also, in the second-best optimum, it may be desirable to achieve a greater rather than smaller concentration of desired arrival times. The intuition is that a marginal change in the concentration of desired arrival times usually does not reduce travel delay costs (except for the discontinuity where all queues suddenly disappear in our model, for a density of desired arrival times equal to the lowest capacity $s_{\text{min}}$). For higher densities, travel cost fall if the density is increased, which is due to the benefit from having SRPATs closer to the LRPAT exceeding the probability-weighted short-run schedule delay costs (if the ratio of long-run and short-run schedule delay values exceeds the probability that the lower capacity state occurs). This long-run schedule delay cost, associated with changing daily schedules and desired arrival times in the face of congestion, is not accounted for in typical analyses that propose spreading of work start times.
6.A Proofs

6.A.1 Derivation of $\dot{Z}^E(t)$

As stated in Section 6.3, drivers choose their SRPAT $t$ by minimizing expected costs $EC^E(t)$. Since drivers are identical regarding their LRPAT and their valuations of time and schedule delays, costs must be equal across drivers. Clearly, the costs depend on whether queuing takes place in both capacity states, or only in the $s_{\text{min}}$ state. We find that in equilibrium the latter is true, and queuing is therefore absent in the $s_{\text{max}}$ state. The costs function, $EC^E(t)$, can then be determined using the results obtained for queuing costs (Eqs. 6.5 and 6.10), short-run schedule delay costs (Eq. 6.3) and long-run schedule delay costs (Eq. 6.12). We first derive $\dot{Z}^E(t)$ for the case of a driver who faces schedule delays early both in the short run (i.e. in the $s_{\text{min}}$ state) and in the long run. Later we will argue, that in equilibrium a driver will either face earliness both in the short and the long run, or lateness both in the short and the long run. Hence, no combinations of earliness in one time dimension and lateness in the other time dimension are part of the equilibrium solution. The expected costs for earliness in both time dimensions are then given by:

$$EC^E(t) = (1 + a)p\beta \frac{Z^E(t)}{s_{\text{min}}} + p\beta \left( t - t_q(s_{\text{min}}) - \frac{Z^E(t)}{s_{\text{min}}} \right) + g\beta(-t)$$

(6.30)

The equilibrium starting time of the peak in the $s_{\text{min}}$ state, $t_q(s_{\text{min}}) = t^* - \theta N/s_{\text{min}}$ (Eq. 6.6), is a function of $Z^E(t)$, since $t^*$ defines the moment when

$93$ It is straightforward to show that no queuing in either state cannot be an equilibrium solution. In the absence of queuing costs, all drivers would have an incentive to minimize their long-run scheduling costs by choosing a SRPAT equal to their LRPAT, and then depart at their SRPAT=LRPAT, in turn, inducing queuing.

$94$ If queuing occurred in both states, the resulting equilibrium density of SRPATs, $\dot{Z}^E(t)$, that is consistent with the cost equality condition shows to be negative $\left( -\frac{(1-d)s_{\text{min}}s_{\text{max}}}{\alpha p\beta s_{\text{max}}^2 s_{\text{min}}^2 s_{\text{min}}} < 0 \right)$. Since for obvious reasons $\dot{Z}^E(t)$ cannot be negative, we discard this solution, and focus on the case when queuing only occurs in the $s_{\text{min}}$ state. $\dot{Z}^E(t)$ for the case of queuing in both capacity states can be derived by performing the same computations as for the case that queuing occurs only in the $s_{\text{min}}$ state, but then adding queuing and short-run schedule delay costs for the $s_{\text{max}}$ state.
Chapter 6. LR vs. SR in equilibrium

A(t, s_{\text{min}}) and Z^E(t) intersect. So, an explicit expression for \( t_q(s_{\text{min}}) \) in Eq. 6.30, as a function of \( Z^E(t) \), is only feasible if the functional form of \( Z^E(t) \) is known. Given that the cost function is linear, it is a natural guess that linearity would also hold for \( Z^E(t) \), and \( \dot{Z}^E(t) \) would thus be a constant. If that is the case, \( t^* \) must be equal to the LRPAT (i.e. 0) for the cost equality condition across drivers to hold. This can most easily be demonstrated by comparing the costs of the driver with the earliest and the one with the latest SRPAT. Both of them will face equal short-run scheduling costs as a consequence of behaving optimally in their short-run scheduling problem. Since they both do not face queuing costs, the only costs they encounter besides the short-run scheduling costs are long-run scheduling costs, which thus have to be equal across these two drivers. Given the assumption that \( Z^E(t) \) is linear as well as our assumption that \( g := b = c \), this is only the case if \( t^* = 0 \) and \( Z(0) = \theta N \). We can thus substitute \(-\theta N/s_{\text{min}}\) for \( t_q(s_{\text{min}}) \) in Eq. 6.30 and take the derivative with respect to \( t \). Setting the derivative equal to 0, the equilibrium density of SRPATs (Eq. 6.14) can be derived:

\[
\frac{dEC^E(t)}{dt} = p\beta \dot{Z}^E(t) + p\beta \left( 1 - \frac{\dot{Z}^E(t)}{s_{\text{min}}} \right) + ap\beta \frac{Z^E(t)}{s_{\text{min}}} - g\beta \quad (6.31)
\]

Indeed it shows that \( \dot{Z}^E(t) \) is a constant, and \( Z^E(t) \) therefore linear. The same \( Z^E(t) \) as given in Eq. 6.31 is obtained if the cost function is defined such that it implies lateness both in the short as well as in the long run. The SRPATs are therefore uniformly distributed in equilibrium.

### 6.A.2 Derivation of \( ETCE \)

The total expected costs in the unpriced equilibrium, \( ETCE \), consist of the sum of (expected) short-run and long-run queuing costs, \( ETCE_{\text{SR}} \), as well as (expected) short-run and long-run scheduling costs, denoted by \( ETCE_{SD} \) and \( TC_{SD}^{LR} \), respectively. Each of these cost elements can be derived by integrating the corresponding driver- (or more precisely, SRPAT-) specific costs across all drivers, starting from the driver with the earliest SRPAT, \( t^E_E \), to the driver with the latest SRPAT, \( t^E_L \) (Eq. 6.15). The density of SRPATs, \( \dot{Z}^E(t) \) has been derived in Eq. 6.14.

Based on the definitions of the expected short-run and long-run queuing costs in equilibrium (see Eqs. 6.5 and 6.10, respectively) and the finding that no queuing occurs if the \( s_{\text{max}} \) state is realized, total expected queuing costs can be derived in the following way:

\[
ETCE^E = (1 + a) \int_{t^E_E}^{t^E_L} EC^E_{\text{SR}}(t) \dot{Z}^E(t) dt =
\]

\[
(1 + a)p \left( \int_{t^E_E}^{t^E_L} C^E_{\text{SR}}(t, s_{\text{min}}) dt \right) = \delta \frac{N^2}{2} \frac{1 + a}{s_{\text{min}}}
\]
Similarly, expected short-run schedule delay costs (Eq. 6.3) can be computed, again taking into account only delay costs in the $s_{\text{min}}$ state:

$$ETC_{SD}^{E,SR} = \int_{t^E}^{t^F} EC_{SD}^{SR}(t) \dot{Z}(t) dt = p \int_{t^E}^{t^F} C_{SD}^{SR}(t,s_{\text{min}}) \dot{Z}(t) dt = \delta N^2 \frac{p}{s_{\text{min}}^2} \left( 1 + a \frac{ap}{(g-p)s_{\text{min}}} \right)$$  (6.33)

Finally, it follows from Eq. 6.12 that long-run schedule delay costs are equal to:

$$TC_{SD}^{E,LR} = \int_{t^E}^{t^F} C_{SD}^{LR}(t) \dot{Z}(t) dt = \delta N^2 \frac{ap}{s_{\text{min}}(g-p)s_{\text{min}}}$$  (6.34)

### 6.A.3 Derivation $\tau_{F,LR}(t)$

If only first-best short-run tolls were implemented, the price function, $EP_{F}(t)$, would be given by the following equation, using the earlier derived results for the costs of short-run and long-run schedule delays (Eq. 6.3 and 6.12) and first-best short-run tolls (Eq. 6.21) (for the case that $t \leq 0$ (hence, SRPAT$\leq$LRPAT)).

$$EP_{F}(t) = \underbrace{p\beta \left( t - t_q(s_{\text{min}}) - \frac{Z(t)}{s_{\text{min}}} \right)}_{EC_{SD}(t)} + \underbrace{g\beta(-t)}_{C_{SD}^{LR}(t)} + \underbrace{p\beta \frac{Z(t)}{s_{\text{min}}}}_{E_{SR}}$$  (6.35)

As argued in Section 6.A.1, $t_q(s_{\text{min}})$ must be equal to $-\theta N/s_{\text{min}}$ if the distribution of SRPATs is uniform, which is true also in the first-best optimum, where $Z(t)$ is equal to $s_{\text{max}}$ (Eq. 6.19). We can then substitute this expression for $t_q(s_{\text{min}})$ in Eq. 6.35. Moreover, it is easy to see that the $p\beta Z(t)s_{\text{min}}$ terms cancel out. The derivative of the price function with respect to $t$, $dEP_{F}(t)/dt$, is thus equal to $(p-g)\beta$. In order to provide for price equality among travelers in the optimum, the long-run toll, $\tau_{F,LR}(t)$, must therefore be set such that its derivative is equal to $(g-p)\beta$. The toll itself, $\tau_{F,LR}(t)$, must then equal $(g-p)\beta Z(t)/s_{\text{max}}$, since it starts from 0 for the driver with the earliest SRPAT. Similarly, it can be shown that for $t > 0$, $\dot{\tau}_{F,LR}(t)$ must be equal to $(p-g)\gamma$, and $\tau_{F,LR}(t)$ to $(g-p)\gamma Z(t)/s_{\text{max}}$. 
Conclusions
7.1 Summary

This dissertation sheds light on various aspects of travelers' scheduling behavior in the presence of recurrent and non-recurrent congestion. In particular, the role of travel time information is emphasized, as it determines whether travelers consider a specific delay as recurrent or non-recurrent, and, as a consequence, to which extent travelers are able to adjust their scheduling choices accordingly. A good understanding of these choices at the level of the individual traveler is crucial for predicting travelers' responses to transport policies, and thus for an accurate evaluation of these policies.

All analyses contained in this thesis take into account recurrent congestion, which is the source of two different types of costs: Costs due to travel time losses and schedule delay costs. The latter result if drivers, when facing recurrent congestion, decide for an arrival time that differs from their preferred arrival time. While the costs caused by travel time losses are taken into account in all analyses (Chapters 2–6), schedule delay costs are considered only in Chapters 3–6. Costs related to non-recurrent congestion are taken into account in all chapters, either as a function of the standard deviation of travel times (Chapter 2), or of schedule delays (Chapters 3–6).

Chapter 2 establishes a method to compute the costs of non-recurrent congestion (variability) as a function of the costs that result from travel time losses caused by recurrent congestion (mean delay). The relation between variability (expressed in terms of the standard deviation) and mean delay shows to be close to linear. Assuming that the unit valuations attached to travel time and variability gains are constant, the costs of non-recurrent congestion can be defined as a mark-up to the costs associated with the travel time losses due to recurrent congestion. This is an important result as it implies that the costs of non-recurrent congestion can be approximated using a fairly simple rule of thumb.

Chapter 2 distinguishes between two measures of variability, which differ in their assumption on how well drivers are informed about day-specific characteristics that potentially affect travel times (e.g. weather conditions, day of the week, season). 'Rough information' implies that travelers consider all deviations from the average (time-of-day-specific) travel time as non-recurrent delays, while 'fine information' implies that only deviations from the day- and time-of-day-specific travel time expectation are regarded as non-recurrent delays. The costs of non-recurrent congestion are fairly sensitive with respect to which type of information drivers have. For instance, assuming a reliability ratio of 0.8 and a link length of 10 km, an increase in the costs of mean delay of 10 Euro is associated with an increase of 3.2 Euro in the costs of non-recurrent congestion if travelers have only rough information, as opposed to an increase of 2 Euro if travelers have fine information. The relation between variability and mean delay depends also on link-specific characteristics, most importantly link length. Ceteris paribus, the costs of variability increase in link length, however, less than proportionally. Finally, the relation between variability and mean delay is also affected by the relative

---

95The reliability ratio is defined as the value of reliability (VOR) divided by the value of time (VOT).
predominance of free-flow, flow-congested and hyper-congested traffic states at a certain time of day. For a given mean delay and time of the day, the costs of variability are highest if free-flow conditions apply on most days, and lowest if hyper-congested conditions tend to prevail.

While Chapter 2 focuses on the aggregate (network link) level, Chapters 3–6 analyze disaggregate scheduling decisions of individual travelers. Chapters 3–5 contain empirical analyses, using revealed preference (RP) data from a real-life peak-avoidance experiment conducted in the Netherlands. Participants of the experiment were eligible for a monetary reward if they avoided using a specific highway link during the morning peak. In Chapter 4, the RP data are combined with data from a stated preference (SP) experiment. Finally, Chapter 6 comprises a theoretical bottleneck model with endogenous scheduling decisions. It does not use any empirical data as input.

In order to estimate scheduling models employing the random utility framework, travel time measurements must be available for the chosen as well as the unchosen departure time alternatives. Chapter 3 argues that for obtaining unbiased valuations of travel time, it is crucial to use door-to-door travel times in RP-based scheduling models. The main reasoning is that travel times tend to be correlated positively across links. So, if due to a lack of data only travel times on subparts of the trip are considered, or fixed travel times are assumed on some links, differences between peak and off-peak travel times may be underestimated, leading to an overestimated value of time. Moreover, non-recurrent congestion can only be taken into account if the analyst knows to which extent travel times differ across days. This renders the use of travel time data from surveys or traffic assignment models challenging as these tend to be deterministic.

For the case of the peak avoidance experiment that is covered in Chapters 3–5, continuous measurements are only available for the highway link along which the peak-avoidance experiment took place. In addition to these, sparse GPS measurements for the remaining relevant parts of the network are available. Using these two data sources, Chapter 3 implements a geographically weighted regression (GWR) model to approximate day-, time-of-day- and location-specific door-to-door travel times, which then allow for deriving measures of expected travel times and travel time variability. The analysis shows that biased valuations of time result if very rough travel time approximations are used.

Chapter 4 focuses on travel time perceptions. Participants of the peak avoidance experiment were asked to report their average travel times for their morning commute. These reported travel times were then contrasted with actual, observed travel times. It is found that on average participants overstate travel times by a factor of ca. 1.5 compared to actual travel times. Chapter 4 tests the hypothesis brought forward by Brownstone and Small (2005) stating that if drivers overestimate delays in reality, they should react to stated delays (e.g. in an SP experiment) as if they were overestimated too. For this purpose, a joint model including RP and SP observations from the same drivers is estimated. If the hypothesis was true,
drivers with a higher ratio of reported to actual travel times are expected to have an SP-based value of time that is relatively low compared to their RP-based value of time. This effect shows to be relatively minor in the context of the peak avoidance experiment considered here. Moreover, no evidence is found that drivers act in reality as if they would overestimate travel times to a similar extent as they overstate them. It can therefore be concluded that there is strong evidence of ‘overstating’, however, much less evidence of ‘overreacting’, neither in SP nor in RP. The results obtained in Chapter 4 thus serve as a note of caution regarding the use of reported travel time data.

Chapters 5 and 6 introduce a distinction between long-run choices of travel routines and short-run choices of (daily) departure times, taking into account that travel times typically vary across days. In the long-run model, drivers choose their routine arrival time subject to the recurrent congestion pattern and their long-run preferred arrival time. The latter is considered exogenous, and is defined as the preferred moment of arrival if congestion was always absent. In the short run, drivers have better information about travel time realizations on a given day than in the long run, when they are only aware of the recurrent congestion pattern. Given this short-run information, they may decide to choose a departure time that results in an arrival time different from their routine arrival time. Chapter 5 models this distinction in an empirical setting, whereas Chapter 6 adopts the theoretical framework of bottleneck congestion to model the differences between long-run and short-run scheduling preferences (Arnott et al., 1990; Vickrey, 1969).

Chapter 5 finds that the long-run and short-run values of travel time and schedule delays (on a given day) diverge strongly, namely by factors between 2 and 5. So, changes in travel time that are considered recurrent (hence, long-run changes) are valued significantly higher than changes that are non-recurrent (hence, short-run changes), most likely because recurrent changes in travel time can be exploited better through the re-scheduling of routines. Conversely, schedule delays are valued higher if they are non-recurrent. An intuitive explanation is that incidental changes induce higher costs because travelers find it difficult to adapt their schedules in the short run.

Chapter 6 focuses on the equilibrium implications of the distinction between long-run and short-run scheduling decisions. Under the assumption that the long-run preferred arrival time is equal across drivers, the user equilibrium implies that short-run preferred arrival times are dispersed in time. The social optimum can then be achieved by applying long-run tolls and first-best short-run tolls, simultaneously. In the second-best situation with short-run tolls being the only available pricing instrument, the same welfare level as in the social optimum can be achieved by applying so-called complementary short-run tolls. Finally, if only long-run tolls can be levied, the social optimum is not reachable any longer.

### 7.2 Policy recommendations

This thesis provides convincing evidence that schedule delay costs from both recurrent and non-recurrent congestion are substantial, stressing that these should
be included in appraisals of transport policies in addition to the benefits (or costs) associated with changes in travel times. Otherwise, cost-benefit analyses might be significantly biased.

Among the studies included in this dissertation, the results obtained in Chapter 2 can most easily be transferred to policy appraisals. They indicate that the costs associated with non-recurrent congestion can be computed as a mark-up to the costs related to recurrent congestion. To utilize these results in practice, it is therefore sufficient if the (expected) travel time losses, which are often readily available from traffic assignment models, and the main characteristics of the link affected by the policy are known. Moreover, a good estimate of the reliability ratio must be available. However, the simplicity of this approach is accompanied by various caveats. Since the analysis in Chapter 2 has been conducted at the aggregate (network link) level, it does not take into account the underlying choice dynamics of scheduling decisions made by individual travelers. In particular, it ignores the scheduling implications of recurrent congestion and only implicitly takes into consideration schedule delays due to non-recurrent congestion by using the standard deviation as a measure of variability. While such criticism does not hold for the disaggregate scheduling models covered in Chapters 3–6, the latter have the downside that they require more sophisticated model structures as well more extensive data input, and are therefore more difficult to implement in policy appraisals.

The results presented in this dissertation strongly emphasize the link between recurrent and non-recurrent congestion on the one hand and travel time information on the other hand. Chapter 2 demonstrates that the costs of non-recurrent congestion are significantly higher if drivers have no information about day-specific factors than if they do have information on factors such as weather conditions or the day of the week. This result is exclusively driven by the fact that the extent of non-recurrent congestion is lower if one is better informed. However, Chapter 5 suggests a second mechanism that may drive a wedge between the costs of variability faced by travelers who are well-informed and those who are not: Costs caused by 1 minute of recurrent (hence, predictable) delay are significantly lower than the costs due to a non-recurrent (hence, unpredictable) delay of 1 minute. This means that the costs of a given delay are lower if drivers were informed about this delay and could therefore adjust their schedules accordingly than in the situation where the delay came unexpectedly. Note that the opposite outcome was found to hold with respect to changes in travel times: An expected travel time gain is valued higher than an unexpected one, supposedly because it can be used more efficiently. From these insights, two major policy recommendations can be derived. First, the valuation of schedule delays as well as of travel time should depend on whether a specific policy induces permanent changes in travel time (e.g. a change in road capacity) or changes in travel times that occur only under certain circumstances (e.g. incident management). Second, in order to correctly estimate the costs and benefits corresponding to a specific transport policy, it is vital that analysts take into account the availability and use of travel time information among the relevant group of travelers.
Finally, it is important to acknowledge that universal values of travel time, schedule delays, and variability do not exist. Instead, they are strongly context-specific. It is therefore essential that the assumptions based on which the valuations have been derived are consistent with the characteristics of the setting to which they applied. For instance, if the value of reliability is based on the assumption that travelers have 'fine' information on travel times, also the extent of variability should be defined such that it only includes those delays that are unpredictable even if one has 'fine' information.

### 7.3 Some suggestions for further research

The selection of promising topics for future research that are discussed below do by no means constitute an exhaustive list of possible research possibilities. Instead, they give suggestions on how the scope of the research contained in this dissertation can be extended and linked to other (in particular, recently emerging) streams in (transport) economics literature.

This thesis has a strong focus on scheduling decisions concerning the morning commute. An important direction of future research is to integrate these decisions in a framework that considers multiple activities with varying durations, time restrictions and interdependencies in time, space as well as between persons. Earlier studies that employ such an activity-based modeling framework for instance take into account the relation between the timing of the morning and the evening commute (e.g. De Jong et al., 2003; Ettema et al., 2007; Jenelius et al., 2011; Zhang et al., 2005), or the interdependencies between the schedules of members of the same household (e.g. Miller and Roorda, 2003; Timmermans and Zhang, 2009). Most of these studies, however, do not derive monetary valuations of travel time and schedule delays, which are required for policy evaluation. Van Acker et al. (2010) argue that valuations obtained from models that consider scheduling decisions as part of wider activity patterns become increasingly relevant as flexible, demand-oriented transportation policies (such as the promotion of telecommuting or congestion pricing) more and more frequently replace long-term, supply-oriented policies (such as road capacity extensions).

Chapters 5 and 6 of this dissertation introduce a distinction between the long-run choice of the travel routine and the short-run choice of the departure time. While the boundary between these decisions is regarded as exogenous in these analyses, in reality it is likely to be the result of inter-temporal optimization processes that involve learning and the updating of travel time expectations. Various theoretical frameworks that model such dynamics have been developed, however, often without validating the framework empirically (e.g. Arentze and Timmermans, 2005; Chen and Mahmassani, 2004; Jha et al., 1998), or employing a fairly descriptive modeling approach towards available data (e.g. Thøgersen, 2006; Zhu et al., 2012). Future research should thus aim at setting up choice experiments that are able to empirically validate these models and to explain adaptations in travel behavior over time.
Most transport economists agree that the economic costs of non-recurrent delays are mainly caused by schedule delays rather than variability \textit{per se} (‘planning costs’) and should therefore also be modeled that way. However, a consensus has not yet been reached on whether recurrent and non-recurrent schedule delays are valued equally, and thus whether the valuations derived for recurrent delays (mainly from SP experiments) can be used to compute the value of reliability, as suggested by Fosgerau and Karlström (2010). Chapter 5 as well as Börjesson (2009) and Börjesson et al. (2012) provide evidence that recurrent and non-recurrent delays might be valued differently. Further research should therefore be undertaken in order to develop a standard methodology for the computation of the value of reliability. The same is true for research on the extent of variability. It is especially important to investigate how variability depends on link length, the road type as well as the underlying network structure such as the existence of parallel routes.

Further research on travel time variability is not only desirable in the domain of empirics, but also in the domain of theoretical equilibrium models with endogenous scheduling decisions that account for both recurrent and non-recurrent congestion (e.g. Fosgerau and Lindsey, 2011). It is particularly interesting to investigate further how the assumptions made regarding the process that generates non-recurrent delays affects the user equilibrium as well as welfare-maximizing road-pricing schemes.

While numerous studies analyze scheduling decisions in passenger transport, only few studies deal with scheduling decisions in freight transport. One reason is the lack of quantitative data. However, also the scheduling decisions as such tend to be relatively more complex for freight transport, as they usually involve various parties, and are part of a supply chain that is likely to be highly sector- or even product-specific (e.g. Kouwenhoven et al., 2005; McKinnon et al., 2009). In order to accurately evaluate policies that affect freight transport, more research concerning the valuation of schedule delays and travel time variability in freight transport needs to be undertaken.

Hence, in order to correctly quantify the costs and benefits resulting from transport-related investments, it is crucial for future research to gain understanding of the underlying decision processes at the level of the individual traveler, including daily activity patterns, the formation of travel time expectations, and the valuation of recurrent and non-recurrent schedule delays. These topics should not only be studied in the context of passenger transport but also in the context of freight transport, for which only few scheduling models have been developed so far. To develop new insights how these decision processes at the individual level affect aggregate congestion levels, the use of equilibrium models such as the bottleneck model (Vickrey, 1969) is vital.
APPENDIX A

The *Spitsmijden* experiment
Appendix A. The Spitsmijden experiment

A.1 Introduction

In Chapters 3, 4 and 5, we use data from a large-scale revealed preference experiment on peak avoidance (it is referred to as Spitsmijden experiment in Dutch). Participants of the experiment were able to obtain a monetary reward for not commuting along a specific highway link during morning peak hours. While each of the chapters that uses data from this experiment discusses those aspects of the experiment that are relevant for the particular chapter, this appendix provides further background on the experimental setting from a more general point of view. Parts of this appendix are also included in the scientific report on this experiment (Knockaert et al., 2012).

A.2 Experimental setting

Background

The peak avoidance experiment took place in the Netherlands and was conducted along the A12 highway between Gouda and Zoetermeer, along a link of 9.21 km length. As this link is confined by two cameras (‘C’), we refer to it as ‘C1–C2’. The peak avoidance experiment was implemented to mitigate the impacts of roadworks on this link, which suffered from recurrent congestion during morning peak hours already prior to the start of the roadworks.

The target of the project was to build up a participation record enabling a daily reduction of 1000 peak hour trips. As participants generally do not travel on a daily basis, several thousand participants were needed to meet the project target. Participants were recruited using automated license plate observations. Cars with an average travel frequency of more than one peak hour trip per two weeks were selected and invitations sent out to license plate registration addresses (including lease companies but limited to passenger cars only). It was decided not to recruit all participants at the initial project start but to steadily increase the participation base until the target was met. The last participants entered the project in September 2009, nearly one year after the project started out with the first batch of participants in November 2008. Overall 15560 direct mail invitations were sent out, resulting in 2911 confirmed participants. More invitations were distributed through lease car companies, and over 500 participants were transferred to this project after another, closely related travel behavior experiment. Overall, the total number of participants was close to 5000. The approximate number of participants during the time period considered in this dissertation is 2000.

Monetary incentive

The reward was intended as a positive monetary stimulus to avoid morning peak hours. To be able to reward a change in travel behavior rather than travel behavior itself, travel behavior during the trial was compared to the participant’s reference travel behavior, as revealed before participation. Based on travel time observations on the C1–C2 link, congested traffic conditions were identified to generally prevail
between 6.30am and 9.30am (during weekdays). Using automated license plate observations over an extended period (30 days or more), the reference travel behavior during this time period was established and invitations were sent out. Upon registration for the trial, the participants had to provide information on all license plates of private cars in their household, accept the project conditions, and confirm the observed aggregate household reference travel behavior (expressed in trips per two weeks) that will be used as reward base. This approach allowed to control for potential fraud by switching vehicles within a household. As a result, the whole household actually participated in the trial.

The reward was calculated every two weeks as the net difference of the reference behavior minus the number of observed peak hour trips over that two week period. Each avoided peak hour trip rewarded the participant with 4 Euro. In case the observed travel frequency over the two weeks was higher than the reference behavior, no reward was registered (and neither did the surplus spill over into subsequent two week periods). In principle, a participant was free to select a travel alternative of his/her choice to avoid the A12 morning congestion and build up reward credit. However, it was not possible for participants to receive rewards for substituting backroad travel for congested motorways. In order to control for the negative safety externalities related to backroad travel, automated observation points were installed on the main alternative roads and these observations were added to A12 motorway travel records in determining rewards. As opposed to backroad travel, it was allowed to substitute travel on other main roads for A12 motorway travel in order to receive a reward. The main travel alternatives available to participants therefore included choice of departure time, route choice (main roads only), transport mode choice, and finally the option to reduce overall travel frequency (for instance by working from home). School holiday periods and official holidays were excluded from the reward calendar (both for determination of the reference behavior and the evaluation of the reward credit).

### A.3 Data sources

A distinction can be made between stated and observed data.

#### Stated data

The stated data are derived from three different questionnaires:

1. **Participant Survey.** The participant survey was filled in by almost all participants at the start of the experiment. It mainly contains questions on socio-economic characteristics and travel habits.

2. **Non-Response Survey.** This survey is comparable to the participant survey in its content. It was conducted among commuters who had been invited to participate in the experiment but declined. 148 non-participants filled in this survey.
3. **Reliability Survey.** This survey was conducted in autumn 2009, after two pilot tests had been performed. 780 participants took part in the reliability survey, yielding a response rate of ca. 30%. The survey focused on information behavior and reactions to delays as well as perceptions of travel time and reliability. Finally, it also comprised a stated preference (SP) choice experiment with 10 choices to be filled in by each respondent. Data from this choice experiment are used in Chapter 4 of this thesis.

**Observed data**

Travel times data are taken from three different sources: Cameras, loop detectors and GPS measurements. The first two are mainly available for the C1–C2 link, whereas (sparse) GPS data are available for different start and end locations (see Figure 9).

1. **Camera data.** The main observation technology used in this experiment are cameras with automated license plate recognition. An array of cameras was deployed along both the A12 motorway and alternative roads. They are used to determine whether a driver avoids traveling during peak hours on a given day. In the analyses undertaken in this dissertation, the camera data are mainly used to measure the scheduling behavior of the participants.

2. **Loop detector data.** Induction loops are installed on the A12 motorway for all lanes and spaced apart in regular intervals of 1 kilometer or less. The loops register both intensities and local speeds. Using a trajectory interpolation methodology, one can infer travel times for a longer motorway stretch. While the induction loop methodology has the drawback of delivering less precise travel times than the camera-based measurements (mainly due to the interpolation involved) in this specific case the availability of loop based observations proved to be better than of camera registration data. In most analyses we therefore use travel times derived from the induction loops in determining C1–C2 travel times and validate them against the camera-observed data.

3. **GPS data.** In a reward experiment preceding the one covered here, 230 drivers were equipped with customized smartphones that included GPS receivers. Those transmitted information about the location of the phones to a central database over the period 11/2008–4/2009 (and at a high frequency of approximately 1 measurement per second). In their raw form, GPS measurements are a series of data points with a location, speed and time stamp attached to them. From these data points, speeds and distances can be calculated. Chapter 3 develops a method based on loop detector data and GPS observations that allows for approximating driver-, day- and time-of-day-specific door-to-door travel times, which are then used to determine expected travel times for chosen as well as unchosen departure time alternatives. Also departure times from home, and arrival times at work can then be derived.
For our analyses we define a commuting trip to start when the speed at which the GPS receiver moves exceeds 5 km/h, and to end when the speed drops to less than 5 km/h for at least 10 minutes. To determine the average speed, also the distance covered between the start and the end time needs to be computed. It is defined as the sum of the distances between the GPS observations belonging to the same trip.

Participants

This section discusses some characteristics of the participants of the experiment, in particular focusing on differences to non-participants. The data are drawn from the participant survey and the non-response survey. Participants are on average older, have attained a higher education level and have a significantly higher income (more than 51% of the participants earn more than 3500 Euro, whereas only 26% of the non-participants earn more than 3500 Euro). A possible explanation might be that jobs that require high education levels and pay high wages tend to offer more flexibility, rendering it in turn easier to participate in the experiment. This explanation is supported by the finding that 70% of the non-participants cannot shift their work starting time to a later time of the day, compared to 30% of the participants. Furthermore, 41% of the non-participants cannot start working right away if they arrive early. The corresponding percentage for participants is only 21%. Participants also tend to travel more frequently along the C1–C2 link than non-participants. 67% of the participants travel 4-5% per week, compared to 50% of the non-participants. Among the participants, a disproportionally large number of drivers faces free-flow travel times longer than 45 minutes. A possible explanation may be that commuters with long travel times tend to travel early in the peak also in the absence of financial incentives to avoid the peak. A shift towards an even earlier travel time (with a passage time of the C1–C2 link before 6:30) is therefore less costly, and therefore makes the participation in the peak avoidance experiment more attractive.

A.4 Datasets

While drawing data from the same experiment, the specifications of the datasets used in the analyses of Chapters 3–5 differ in various respects, as demonstrated by Table 21. First, in Chapter 4 we consider a relatively long time span, not only comprising the months during which rewards were distributed (2009/09–2009/12), but also post-measurements from January 2010. As a consequence, we are able to analyze counterfactual behavior of the participants (without financial incentive) and reduce the correlations between travel time and reward inherent in the dataset. The inclusion of these data seemed less relevant in Chapters 3 and 5 where comparatively simpler choice models were estimated. The number of choice alternatives varies only slightly across chapters (16–17). We take into account days without marginal rewards during the period 2009/09–2009/12 in Chapters 3 and 4. These are equal to zero for a given driver if he has already reached the maximum
number of weekly rewards he can get based on his pre-experimental reference behavior. Days without marginal reward are excluded in Chapter 5, as it is unclear to which extent drivers are able to anticipate them in the long run.

### Table 21: Overview datasets used in Chapters 3–5

<table>
<thead>
<tr>
<th>Choices</th>
<th>Chapter 3</th>
<th>Chapter 4</th>
<th>Chapter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-measurements (unrewarded)</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Nr. of (departure time) choices</td>
<td>9530</td>
<td>2997</td>
<td>5965</td>
</tr>
<tr>
<td>Nr. of choice alternatives</td>
<td>16</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Days without marginal reward</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Expectations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current travel times</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Average travel times</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Personal past experience</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Weekday-specific travel times</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Travel time expectations are defined slightly different in each of the three chapters. In Chapter 3, expectations are based on a weighted average of current travel times (on the day of travel) and average travel times in the period 2009/09–2009/12. This approach has the advantage that this type of expectation can be computed for all days and all drivers (given that door-to-door travel times are available). In line with the aim of Chapter 3, this definition allows us to test different definitions of the expectations, using a high number of departure time choices (9530). In Chapter 4, in contrast, a weighted average between current travel times and the person-specific experience was used to define travel time expectations. The main underlying rationale was that this definition is fairly consistent with the definition of the travel time ratio, which measures the relation between reported and actual travel times and is based on the person-specific history. This specification has the downside that the number of available observations drops sharply, as (by assumption) at least 9 past trips are necessary in order to derive an indication of past experience. Finally, in the departure time choice models of Chapter 5 (hence, the short-run model), a weighted average of current travel times and (average) weekday-specific travel times is taken into account. This is because in Chapter 5 we derive routine travel behavior, which shows to be weekday-specific. We therefore take into account weekday-specific travel time patterns both in the long-run and in the short-run models contained in Chapter 5.

### A.5 Information behavior

This section provides a brief overview of how participants form travel time expectations, how they inform themselves about delays and how they react to delays.

---

97 Note that the relative weight is estimated rather than assumed.
All data are retrieved from the ‘reliability survey’. First, respondents were asked about the determinants of their travel time expectations. They were asked to judge to which extent travel times during an average working day, weather conditions, seasonal patterns, school holidays, road-works, incidents, or any other factors affect their travel time expectations. Figure 23 shows that most drivers take these factors into account ‘strongly’ or ‘very strongly’. The most important factor are school holidays with almost 80% of drivers stating that they consider them a ‘strong’ or ‘very strong’ determinant of travel times, followed by the weekday.

Several factors presented in Figure 23 and their average effects on travel times can be known far in advance (e.g. season, weekday, school vacation days). However, in order to gather information about factors that are typically not known far in advance (e.g. road works), or only become known on the day of travel (e.g. incidents), one needs to use sources of information that provide more up-to-date information on traffic conditions. Figure 24 shows that only 8% of the respondents never use such additional sources of information, while 61% of the respondents always or often use them.
Figure 24: How often do you inform yourself about actual traffic conditions?

Once drivers know about delays occurring along the roads they use during their commuting trips, they have different options on how to react. They change their departure time, work from home, postpone appointments at work, switch to another route, or to another mode of transportation. The choice between these reaction strategies does not only depend on their availability to a specific driver, but also on how much in advance the driver is informed about the delay. Two distinct cases of information timing are considered. In the first case, drivers obtain the information about an (expected) delay the day before the delay actually occurs. In the second case, drivers obtain the information on the same day on which the delay occurs. Hence, in the first case drivers still have the possibility to shift to an earlier departure time, while this is less feasible in the second case where the delay already exists at the time that the driver is informed about it.

For different sizes of delays, respondents were asked to indicate what their most likely behavioral response will be (several choices were possible for each delay size). The differences between the responses for the two cases turn out to be quite small (Figure 25). But as expected, the relative importance of departure time shifts is higher in the first case, while the relative importance of route switching is higher in the second case. The share of drivers who do not change their departure time is close to 80% for a delay of 10 minutes regardless of the time they were informed about it. However, the corresponding percentage drops to 13% (information one day in advance) and 23% (information on the same day) for a delay of 30 minutes.

The above evidence on the participants’ information behavior is consistent with the finding in Chapter 5 that drivers take into account weekday-specific travel times in the expectation formation, as well as other factors, which only become available in the short-run. Also, Figure 25 showed that changes in departure time are an important reaction strategy towards delays.
A.6 Travel time distributions

In this section, reported and observed travel time distributions on the C1–C2 link are compared. The reported distributions are retrieved from the ‘reliability survey’ and the observed travel times from the loop detector data. For the reported distributions, participants were asked to state the travel time distribution on the C1–C2 link for three time instances: a passage time at camera C1 at 7:30, 8:30 and 9:30. For this purpose the respondents were asked to assign 20 trip travel times into 5-minute intervals. Specifically, they had to indicate on many trips they expect to face a travel time between 0 and 5 minutes, between 5 and 10 minutes, etc. A total of 9 such time slots were defined, with the last one identifying the number of trips with a travel time longer than 40 minutes. If respondents were not experienced in driving during those time instances, they were asked to guess the travel time distribution based on their experience of driving during other times of the day along the same link.
We distinguish between two types of observed travel time distributions: Individual observed and generic observed. Individual observed travel time distributions are based on travel times observed for a specific driver with a passage time at C1 within a 30-minute interval enclosing the specific time instance (7.30, 8.30, 9.30, respectively) as midpoint. Travel times between September 1, 2009 and the date of filling out the questionnaire are hereby taken into account. Generic travel time distributions are derived from loop data measurements on the C1–C2 road link during the 20 working days preceding the date on which a specific person has filled out the questionnaire. Since the time interval during which participants of the Spitsmijden project were able to complete the questionnaire is quite short, the generic distributions are fairly similar across respondents. Figure 26 shows the observed as well as the reported travel time distributions using bar diagrams, while Figure 27 compares the cumulative probability distributions from Figure 26, averaged across the three time instances.

We find significant differences between the reported and the observed distributions, which are widely consistent with the findings of Chapter 4, namely that reported travel times are significantly longer than observed travel times. This finding is consistent for all three time instances. In addition, we can also see a
tendency of drivers to overreport the probabilities of events with small likelihoods. This is the case for very long travel times and to a lesser extent also for very short travel times. These findings resemble the usual outcomes found in studies based on cumulative prospect theory, which relates objective probabilities to subjective probabilities (Quiggin, 1982; Schmeidler, 1989; Tversky and Kahneman, 1992).

Moreover, we find that a substantial share (13.70%) of respondents do not take into account any travel time variability. These respondents state that for all three time instances, all 20 trip times are within one of the 5-minute intervals. It is unknown whether this outcome represents their real perception of the travel time distribution, whether they misunderstood the question, or whether they had the intention to save time when filling out the questionnaire.
Bibliography


Cosslett, S. (1977). The trip timing decision for travel to work by automobile. In D. McFadden et al. (Eds.), *Demand model estimation and validation, the urban travel demand forecasting project*. The University of California, Berkely and Irvine, CA: The Institute of Transportation Studies.


Samenvatting (Dutch Summary)

Kosten van files: de invloed van onzekere reistijden, informatie en gedrag


Files leiden dus tot kosten. Sommige van deze kosten, bijvoorbeeld brandstoff-en onderhoudskosten, zijn gemakkelijk in geld uit te drukken en zijn ook direct voor rekening van de gebruiker. Daarentegen zijn de kosten van reistijd en vertragingen niet gemakkelijk in geld uit te drukken. Dit proefschrift kwantificeert diverse soorten kosten van files. De kosten van reistijd, onbetrouwbaarheid van reistijd en het eerder of later dan de gewenste of verwachte tijd aankomen op de bestemming (schedule delays) worden berekend via een tijdstipkeuzemodel waarbij reizigers in werkelijke en hypothetische situaties hun optimale vertrektijd kiezen. Op basis van een beter begrip van het tijdstipkeuzegedrag van de individuele reiziger kunnen kosten en baten van transportbeleid – bijvoorbeeld investeringen in infrastructuur, het beschikbaar maken van (reis)informatie en het beprijzen van autogebruik – nauwkeuriger worden geraamd.

In het onderzoek naar tijdstipkeuzegedrag maakt dit proefschrift een belangrijk onderscheid tussen enerzijds dagelijkse en anderzijds incidentele files. Of files als dagelijks of incidenteel worden ervaren, hangt voornamelijk af van de voor de reiziger beschikbare (reis)informatie. Door verschillen in deze beschikbaarheid van informatie kan dezelfde vertraging door verschillende reizigers anders worden ervaren. Dagelijkse files in de ochtend- en of avondspits zijn voor de reiziger in hoge mate voorspelbaar. Incidentele files zijn minder goed tot niet voorspelbaar en zorgen bij de reiziger - in tegenstelling tot de dagelijkse files – voor onzekerheid over reistijden. Zowel aanbodfluctuaties (bijvoorbeeld afgesloten rijstroken door verkeersongelukken) als vraagfluctuaties (bijvoorbeeld extra verkeer door een evenement) kunnen incidentele files veroorzaken. Een ander belangrijk verschil tussen beide typen files is dat dagelijkse files resulteren in vertragingen ten opzichte van de gewenste aankomsttijd, terwijl incidentele files resulteren in vertragingen ten opzichte van de verwachte aankomsttijd.

Incidentele files impliceren een verdeling van reistijden. Een verdeling van reistijden leidt tot reistijdvariatie oftewel onbetrouwbaarheid van reistijden. De
standaardafwijking en variantie van de reistijdverdeling zijn vaak gebruikte maatstaven voor deze reistijdvariatie. Hoofdstuk 2 toont aan dat deze onbetrouwbaarheid van reistijden een functie is van de gemiddelde vertraging, waarbij de gemiddelde vertraging een indicator is voor de kosten van de dagelijkse files. De relatie tussen reistijdvariatie en gemiddelde vertraging is bij benadering lineair. Daaruit volgt dat de kosten van incidentele files kunnen worden gedefinieerd als de kosten van reistijdverlies veroorzaakt door dagelijkse files plus een mark-up. Dit resultaat is belangrijk omdat het laat zien dat de kosten voor incidentele files kunnen worden benaderd aan de hand van simpele en bestaande indicatoren van dagelijkse files. Verder toont Hoofdstuk 2 aan dat de hoogte van deze mark-up afhankelijk is van trajectspecifieke kenmerken en veronderstellingen over het gebruik van reisinformatie door reizigers. Dit hoofdstuk maakt bij het definiëren van de reistijdvariatie onderscheid tussen twee niveaus van reisinformatie: ‘globale’ en ‘nauwkeurige’ informatie. Bij ‘globale’ informatie beschouwen reizigers alle afwijkingen van de gemiddelde vertrektijd specifieke reistijd als incidentele vertraging. In het geval van ‘nauwkeurige’ informatie worden enkel de afwijkingen van de gemiddelde dag- en vertrektijd specifieke reistijd – dus rekening houdend met dagspecifieke kenmerken, zoals: dag van de week, weersomstandigheden en vakantieperiode – als incidentele vertraging aangemerkt. De resultaten tonen aan dat onder de veronderstelling van ‘globale’ informatie de kosten van reistijdvariatie substantieel hoger zijn ten opzichte van de veronderstelling van ‘nauwkeurige’ informatie.

In tegenstelling tot Hoofdstuk 2 analyseren de Hoofdstukken 3, 4, 5 en 6 het tijdstipkeuzegedrag van de individuele reiziger. Empirisch onderzoek gebaseerd op het Spitsmijden experiment wordt behandeld in de Hoofdstukken 3, 4, en 5. Hoofdstuk 6 analyseert theoretisch het effect van tijdstipkeuzegedrag van reizigers op de korte en lange termijn op basis van het standaard bottleneckmodel.

Voor het schatten van tijdstipkeuzemodellen (in een random utility model) op basis van daadwerkelijk geobserveerde keuzes (revealed preference data) moet de reistijd van zowel de gekozen vertrektijd als van alle andere alternatieve vertrektijden beschikbaar zijn. In het Spitsmijden experiment wordt alleen de reistijd op een snelwegtraject continue gemeten. Hoofdstuk 3 betoogt dat de beschikbaarheid van deur-to-deur reistijden noodzakelijk is voor het verkrijgen van zuivere schattingen van de monetaire waarde van reistijd. Reistijden, met name het reistijdverschil tussen binnen en buiten de spits, kunnen worden onderschat doordat reistijden constant worden verondersteld of zelfs in het geheel worden genegeerd voor trajecten waarvan de reistijd onbekend is. Dit kan leiden tot een overschatting van de waarde van reistijd.

Incidentele files kunnen alleen worden geanalyseerd als de onderzoeker kennis heeft over de spreiding van reistijden over de dag en tussen dagen onderling. Standaard verkeersmodellen bevatten deze noodzakelijke informatie over de spreiding van reistijden meestal niet. Individuele continue reistijdinformatie is

\[98^{In dit experiment maken respondenten aanspraak op een geldelijke beloning als ze in de ochtendspits niet op bepaalde snelwegtrajecten rijden, met andere woorden als ze de spits mijden. Zie Appendix A voor een overzicht van dit experiment.\]
in het Spitsmijden experiment enkel beschikbaar voor het op de snelweg Gouda-Zoetermeer (A12) afgelegde gedeelte van de reis. Voor de overige trajecten binnen het gebied is op beperkte schaal individuele GPS data beschikbaar. Hoofdstuk 3 maakt gebruik van beide soorten informatie en schat via een geographically weighted regression model de deur-tot-deur reistijden variërend over de dag van de week, het tijdstip van de dag en de vertrek- en aankomstlocatie. De resultaten tonen aan dat het gebruik van zeer globale benaderingen van de reistijd kan leiden tot onzuivere schattingen van de waarde van reistijd.

De percepties van reistijden staan centraal in Hoofdstuk 4. In het eerder genoemde Spitsmijden experiment heeft elke respondent zijn of haar gemiddelde reistijd in de ochtend gerapporteerd. Vervolgens worden deze gerapporteerde reistijden met de daadwerkelijke, geobserveerde reistijden vergeleken. Deze vergelijking toont aan dat respondenten hun gemiddelde reistijd ten opzichte van de werkelijke reistijd met een factor van ongeveer 1.5 overschatten. Deze gemiddelde overschatting is groter voor ritten over kortere afstanden en ritten met hogere gemiddelde snelheden.


Hoofdstukken 5 en 6 introduceren het onderscheid tussen langetermijnkeuzes van reisroutines en kortetermijnkeuzes van vertrektijden. De geprefereerde aankomsttijd en reistijdverwachtingen zijn de twee dimensies waarin de verschillen tussen korte- en langetermijnkeuzes naar voren komen. De geprefereerde aankomsttijd op de lange termijn wordt exogeen verondersteld en is gelijk aan de geprefereerde aankomsttijd bij het ontbreken van dagelijkse files. Daarentegen is de geprefereerde aankomsttijd op de korte termijn endogeen. De bestuurder heeft rekening houdend met de dagelijkse files een langetermijnpreferentie voor een bepaalde aankomsttijd. Doordat beschikbare (reis)informatie verschillend is voor de korte en lange termijn, kunnen de verwachtingen van de reistijd ook verschillen. Voor de lange termijn heeft de bestuurder enkel informatie over de gemiddelde vertrektijdspecifieke reistijd. De beschikbare informatie wordt nauwkeuriger naarmate het tijdstip van vertrek dichterbij komt.

Hoofdstuk 5 toont, gebaseerd op een geschat tijdstipkeuzemodel, aan dat de monetaire waarde van reistijd voor de lange termijn substantieel hoger is
vergeleken met de monetaire waarde van reistijd voor de korte termijn. Een mogelijke verklaring hiervoor is dat reizigers reistijdveranderingen voor de lange termijn beter kunnen inpassen in hun reisschema’s. Het tegengestelde geldt voor de monetaire waarde van eerder of later dan de geprefereerde tijd aankomen (schedule delays): als het te vroeg of te laat aankomen wordt veroorzaakt door incidentele files is de monetaire waarde duidelijk hoger. Een verklaring hiervoor is dat bestuurders hogere kosten ervaren bij incidentele veranderingen in de schedule delays omdat zij in dat geval hun tijdstipkeuze gedrag moeilijker kunnen aan passen.

Het effect van het onderscheid tussen korte en lange termijn tijdstipkeuze gedrag op de evenwichtscondities in het bottleneckmodel staat centraal in Hoofdstuk 6. Onder de aanname dat de geprefereerde aankomsttijd op de lange termijn gelijk is voor alle bestuurders toont dit hoofdstuk aan dat de geprefereerde aankomsttijd op de korte termijn moet variëren tussen de reizigers. Hieruit volgt dat het sociaal optimum kan worden bereikt door tegelijkertijd een langetermijnheffing en een first-best kortetermijnheffing in te voeren. De laatstgenoemde heffing is gericht op de vertrektijdkeuze, terwijl de langetermijnheffing zich richt op de reisroutines. Als enkel kortetermijnheffingen mogelijk zijn, kan het sociaal optimum bereikt worden door zogenaamde quasi first-best kortetermijnheffingen in te voeren. In het geval dat enkel langetermijnheffingen mogelijk zijn, kan het sociaal optimum echter niet worden bereikt.

**Beleidsimplicaties en suggesties voor verder onderzoek**

Dit proefschrift laat zien dat de kosten die samenhangen met onbetrouwbaarheid van reistijden, ongeacht of deze veroorzaakt worden door dagelijkse of incidentele files, substantieel zijn. Voor een juiste raming van de kosten en baten van transportbeleid is het daarom van belang om naast de kosten die samenhangen met een verandering in reistijd ook de kosten van onbetrouwbaarheid te kwantificeren. De resultaten uit Hoofdstuk 2 tonen aan dat de kosten van incidentele files kunnen worden gedefinieerd als de kosten van reistijdverlies veroorzaakt door dagelijkse files plus een mark-up. Om deze methode toe te passen is enkel informatie nodig over de reistijdverandering veroorzaakt door dagelijkse files en de belangrijkste kenmerken van de route. Deze methode is daarom ook de meest eenvoudige methode om effecten van incidentele files mee te nemen in kosten-batenanalyses. Echter, het negeren van het onderliggende tijdstipkeuze gedrag van reizigers kan leiden tot een onjuiste raming van de kosten van files.

Om de kosten en baten van transportbeleid juist te kunnen schatten, is het noodzakelijk dat bekend is in hoeverre bestuurders de beschikbare reistijdeninformatie gebruiken in hun beslissingen omtrent reisschema’s en reisroutines. Uit de Hoofdstukken 2 en 5 blijkt dat het gebruik van reisinformatie van invloed is op de kosten van dagelijkse en incidentele files. Hoofdstuk 2 toont aan dat goed geïnformeerde bestuurders files minder vaak als incidenteel ervaren. De omvang en de kosten van incidentele files dalen dus naarmate er meer reisinformatie beschikbaar is.
Hoofdstuk 5 toont aan dat de schedule delay kosten per minuut lager zijn naarmate reizigers beter geïnformeerd zijn en daardoor beter hun reisschema kunnen aanpassen. Het tegengestelde effect treedt op bij veranderingen in reistijden: bestuurders hebben een hogere waardering voor een verwachte reistijdswinst in vergelijking met een onverwachte reistijdswinst. Een mogelijke verklaring hiervoor is dat bestuurders een verwachte reistijdswinst efficiënter kunnen benutten. Beide hoofdstukken benadrukken dus het belang van reisinformatie voor bestuurders. Hieruit volgen twee belangrijke beleidsimplicaties. Ten eerste, de monetaire waarderingen van reistijd en onbetrouwbaarheid van reistijden moeten gedifferentieerd worden naar beleid dat gericht is op permanente (bijvoorbeeld het verbreden van een weg) en/of incidentele (bijvoorbeeld incident management bij ongevallen) veranderingen in reistijd. Ten tweede, voor een juiste raming van de kosten en baten van transportbeleid is het noodzakelijk om de beschikbaarheid en het gebruik van reisinformatie door de reizigers te identificeren en te analyseren.


Er is nog geen consensus over de vraag of schedule delays voor dagelijkse en incidentele files hetzelfde worden gewaardeerd. Een logische vervolgvraag is of de waarde van schedule delays veroorzaakt door dagelijkse files gebruikt kan worden om de waarde van onbetrouwbaarheid van reistijden te berekenen zoals wordt gesuggereerd door Fosgerau and Karlström (2010). Het onderzoek in zowel Hoofdstuk 5 als in Börjesson (2009) en in Börjesson et al. (2012) laat zien dat de schedule delays veroorzaakt door dagelijkse en incidentele files mogelijk verschillend wordt gewaardeerd. In vervolgonderzoek kan aandacht worden besteed aan het ontwikkelen van een standaardmethode om de waarde van onbetrouwbaarheid van reistijden te berekenen.

In het meeste onderzoek naar tijdstipkeuzegedrag staat personenverkeer centraal en wordt vrachtverkeer genegeerd. Aangezien vrachtverkeer ook een belangrijk onderdeel is van de mobiliteit van een samenleving, is een beter begrip nodig van het tijdstipkeuzegedrag van het vrachtverkeer. Dit tijdstipkeuzegedrag is meer complex omdat er vaak meerdere partijen bij betrokken zijn en de beslissingen van en over vrachtverkeer een integraal onderdeel uitmaken van de productie- en distributielogistiek.

De analyses in de Hoofdstukken 5 en 6 veronderstellen dat korte- en langetermijnbeslissingen onafhankelijk van elkaar zijn. Een meer accurate en uitdagende
veronderstelling is dat de korte- en langetermijnbeslissingen allebei onderdeel zijn van een intertemporeel keuzeproces. Dit proces wordt gekenmerkt door verwachtingen en leereffecten. Met name de empirische validatie van zo'n dynamisch keuzeproces - en de effecten voor monetaire waardering van onder andere de reistijd - kan een belangrijk aspect zijn van vervolgonderzoek. Naast deze empirische validatie is het (door)ontwikkelen van theoretische verkeersevenwichtsmodellen, zoals bijvoorbeeld het bottleneckmodel, belangrijk. De bestaande modellen moeten verder worden verfijnd om zo ook het recursieve effect van tijdstipkeuzegedrag op zowel dagelijkse en incidentele files te kunnen analyseren.
The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

<table>
<thead>
<tr>
<th>Book</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>507</td>
<td>Essays on Firm Dynamics, Competition and Productivity</td>
<td>U. Kilinc</td>
</tr>
<tr>
<td>508</td>
<td>R&amp;D, Innovation and the Policy Mix</td>
<td>M.J.L. de Heide</td>
</tr>
<tr>
<td>509</td>
<td>The Impact and Performance of Industrial Sites: Evidence from the Netherlands</td>
<td>F. de Vor</td>
</tr>
<tr>
<td>510</td>
<td>Do ut Des: Incentives, Reciprocity, and Organizational Performance</td>
<td>J.A. Non</td>
</tr>
<tr>
<td>511</td>
<td>Empirical Studies on Credit Risk</td>
<td>S.J.J. Koniijn</td>
</tr>
<tr>
<td>512</td>
<td>Enhanced Cooperation in Corporate Taxation</td>
<td>H. Vrijburg</td>
</tr>
<tr>
<td>513</td>
<td>Behavioural Models of Technological Change</td>
<td>P. Zeppini</td>
</tr>
<tr>
<td>514</td>
<td>It's Communication, Stupid! Essays on Communication, Reputation and (Committee) Decision-Making</td>
<td>P.H. Steffens</td>
</tr>
<tr>
<td>515</td>
<td>Essays on Executive Compensation - Managerial Incentives and Disincentives</td>
<td>K.C. Yu</td>
</tr>
<tr>
<td>516</td>
<td>Of Needles and Haystacks: Novel Techniques for Data-Rich Economic Forecasting</td>
<td>P. Exterkate</td>
</tr>
<tr>
<td>517</td>
<td>Political Economics in the Laboratory</td>
<td>M. Tyszler</td>
</tr>
<tr>
<td>518</td>
<td>Aggregate Productivity Growth under the Microscope</td>
<td>Z. Wolf</td>
</tr>
<tr>
<td>520</td>
<td>The cost of travel time variability for air and car travelers</td>
<td>P.R. Koster</td>
</tr>
<tr>
<td>521</td>
<td>Essays of nonparametric econometrics of stochastic volatility</td>
<td>Y. Zuo</td>
</tr>
<tr>
<td>522</td>
<td>Rare Event Simulation Techniques for Stochastic Design Problems in Markovian Setting</td>
<td>B. Kaynar</td>
</tr>
<tr>
<td>523</td>
<td>Developments in Measuring and Modeling Financial Volatility</td>
<td>P. Janus</td>
</tr>
<tr>
<td>524</td>
<td>Essays on the Economics of Housing Subsidies</td>
<td>E.P.W. Schilder</td>
</tr>
<tr>
<td>525</td>
<td>Bifurcations of Indifference Points in Discrete Time Optimal Control Problems</td>
<td>S. M Moghayer</td>
</tr>
<tr>
<td>526</td>
<td>Exploiting Common Features in Macroeconomic and Financial Data</td>
<td>C. Çakmakli</td>
</tr>
<tr>
<td>527</td>
<td>Experimenting with new combinations of old ideas</td>
<td>J. Linde</td>
</tr>
<tr>
<td>528</td>
<td>Bounded rationality and heterogeneous expectations in macroeconomics</td>
<td>D. Massaro</td>
</tr>
<tr>
<td>529</td>
<td>Groups in Economics</td>
<td>J. Gillet</td>
</tr>
<tr>
<td>530</td>
<td>Evaluating Econometric Models and Expert Intuition</td>
<td>R. Legerstee</td>
</tr>
<tr>
<td>531</td>
<td>Essays on the Political Economy of Finance</td>
<td>M.R.C. Bersem</td>
</tr>
</tbody>
</table>
532 T. WILLEMS, Essays on Optimal Experimentation  
533 Z. GAO, Essays on Empirical Likelihood in Economics  
534 J. SWART, Natural Resources and the Environment: Implications for Economic Development and International Relations  
535 A. KOTHIYAL, Subjective Probability and Ambiguity  
536 B. VOOGT, Essays on Consumer Search and Dynamic Committees  
537 T. DE HAAN, Strategic Communication: Theory and Experiment  
538 T. BUSER, Essays in Behavioural Economics  
539 J.A. ROSENO MONCAYO, On the importance of families and public policies for child development outcomes  
540 E. ERDOGAN CIFTCI, Health Perceptions and Labor Force Participation of Older Workers  
541 T. WANG, Essays on Empirical Market Microstructure  
542 T. BAO, Experiments on Heterogeneous Expectations and Switching Behavior  
543 S.D. LANSDORP, On Risks and Opportunities in Financial Markets  
544 N. MOES, Cooperative decision making in river water allocation problems  
545 P. STAKENAS, Fractional integration and cointegration in financial time series  
546 M. SCARTH, Essays on Monte Carlo Methods for State Space Models  
547 J. ZENHORST, Macroeconomic Perspectives on the Equity Premium Puzzle  
548 B. PELLOUX, the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies  
549 N. YANG, Markov-Perfect Industry Dynamics: Theory, Computation, and Applications  
550 R.R. VAN VELDHUIZEN, Essays in Experimental Economics  
551 X. ZHANG, Modeling Time Variation in Systemic Risk  
552 H.R.A. KOSTER, The internal structure of cities: the economics of agglomeration, amenities and accessibility.  
554 J.L. MÖHLMANN, Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products  
555 S.M. HOOGENDOORN, Diversity and Team Performance: A Series of Field Experiments  
556 C.L. BEHRENS, Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition  
557 G. SMRKOLJ, Dynamic Models of Research and Development