Behavioural Models for Route Choice of Passengers in Multimodal Public Transport Networks

PhD Thesis

Marie Karen Askegren Anderson

November, 2013

DTU Transport
Department of Transport

Behavioural Models for Route Choice of Passengers in Multimodal Public Transport Networks

PhD Thesis

Marie Karen Askegren Anderson

Supervisor:
Professor Otto Anker Nielsen
Technical University of Denmark
Department of Transport

Co-supervisor:
Professor Carlo Giacomo Prato
Technical University of Denmark
Department of Transport

November 2013

PhD-2013-03
Department of Transport, Technical University of Denmark

Published by: DTU Transport
Bygningstorvet 116B
DK-2800 Kongens Lyngby

Order at: http://www.transport.dtu.dk
Tel +45 45256500
transport@transport.dtu.dk

ISBN: 978-87-7327-265-7 (Electronic version)

ISBN: 978-87-7327-266-4 (Printed version)

PREFACE

The following thesis concludes my PhD study entitled Behavioural Models for Route Choice of Public Transport Passengers. The PhD study has been carried out at the Department of Transport at the Technical University of Denmark (DTU Transport) under supervision of Professor Otto Anker Nielsen and co-supervised by Associate Professor Carlo Giacomo Prato. The study has been conducted in the period from May 2007 – September 2013.

The PhD study has been a long process and along the way many people have been involved. I am very grateful to my supervisor and co-supervisor who have guided me through the project and helped with both criticism and inspiration. Furthermore I would like to thank my colleagues, especially Hjalmar Christiansen, for their great help with the TU survey, Thomas Kjær Rasmussen for his great help and our valuable discussions, and Niken Prameswari for her help with the data. I would also like to thank Mette Aagaard Knudsen for our many (more or less academic) conversations.

Least but not less I would like to thank my friends and family for their great support and for their continuous believe in me and the project. I owe the greatest thanks to my loving husband Mark and to our dear children Malthe and Alberte; they are the greatest joys and loves of my life.

DTU Transport, Kgs. Lyngby, 2013

Marie Karen Askegren Anderson

ABSTRACT

The subject of this thesis is behavioural models for route choice of passengers in multimodal public transport networks.

While research in sustainable transport has dedicated much attention toward the determinants of choice between car and sustainable travel options, it has devoted less attention toward the route choices of public transport users. Clearly, identifying relevant factors that affect route choice decisions could guide stakeholders (e.g., local governmental agencies and public transport agencies) toward effective improvement of public transport services in metropolitan areas in order to increase their attractiveness with respect to the car. Accordingly, this PhD thesis faces the multi-faceted challenge of modelling route choices of travellers moving in a metropolitan multimodal network. The analysis focuses on revealed preferences data collected for the multimodal network of the Greater Copenhagen Area and solves the multiple facets of the challenge concerning (i) data collection, (ii) data analysis, (iii) choice set generation, and (iv) model estimation.

From the *data perspective*, this thesis overcomes limitations in the collection of actual route choices of public transport users. The literature shows a lot of effort in modelling route choices of car users, which has benefitted from increasingly accurate GPS devices to track vehicles and increasingly precise map-matching algorithms to translate the GPS points into routes on GIS networks. However, the literature shows scarce effort in the estimation of route choice models of public transport users based upon observed choices. Public transport route choice models have not benefitted from the same technological enhancements as car models because of the necessity (i) to collect additional information concerning lines and transfers, and (ii) to overcome technical limitations related to GPS signals not always being retrievable in tunnels that are used by metro and urban rail systems. In this PhD project, a questionnaire to collect details about the actual route choice behaviour in public transport networks was developed and tested in a full scale test. Afterwards the questions were added to the Danish Travel Behaviour Survey that collects daily travel diaries with a questionnaire covering activities and travel of a representative sample of the population. When the travel is by public transport modes, an additional section of the survey with the new questions collects detailed information about access modes, stations, lines, departure and arrival times, trip purposes, transfers, and egress modes. In order to analyse travellers' preferences in the multimodal network, about 6,000 observations from the Greater Copenhagen Area were collected and processed in this study. The characteristics of the collected data are analysed and the actual choices of the public transport passengers are revealed in the thesis. The data were map-matched to the GIS network of the area and quality controlled in a multi-step procedure.

From the *choice set generation perspective*, this thesis generates attractive routes for the origin-destination pair of each traveller. The problem is not trivial when considering the combinatorial

nature of the problem. The dense network of the Greater Copenhagen Area includes metro, trains (regional, suburban, urban and local), and buses (high-frequency, express and regular), and access and egress modes comprise both private (bicycle and car) and public transport modes. Accordingly, the universal realm of possible combinations (i.e., access modes, public transport modes, lines, transfers, egress modes) is large. This thesis proposes a doubly stochastic approach for generating alternative routes that are relevant to travellers, since the method allows accounting for both perceived costs of the network elements and heterogeneity in the preferences of travellers. The coverage of the observed choices with the generated choice sets provides a measure of the behavioural plausibility of the applied path generation technique. Notably, the definition of the coverage for public transport networks is different from the one for automobile users because of the increased dimensionality of the problem, as similarity in multimodal networks may be calculated at both the line level and the link level. The thesis describes testing of the choice set generation algorithm with regard to the number of routes generated as well as its ability to generate the observed routes.

From the *model estimation perspective*, this thesis describes the estimation of route choice models able to account for similarities across alternatives. A simple approach is the formulation of a Path Size Logit in which the different definitions of similarity (i.e., at the line level and at the link level) are alternatively tested. A more elaborated approach is the formulation of a Mixed Path Size Logit. For both approaches, the utility function is specified in order to consider the multidimensional nature of the problem in terms of access/egress characteristics, waiting time, in-vehicle travel time, and transfer characteristics. Moreover, travellers' characteristics and trip purposes enrich the model and provide insight into the preference structures of different travellers with different motivations for travelling, and finally the study indicates that the actual length of the trip has an impact on the preferences of the travellers.

The estimation confirms the expected importance of waiting and transfer times, shows different preferences for bus and train, emphasize the importance of the trip length, shows the effect of specific modes of access and egress, and indicates the relevance of individual characteristics within and across trip purposes. The results suggest the importance of coordination between different public transport modes, the relevance of transfer locations that allow seamless passage from one vehicle to another, and the significance of access and egress modes in terms of parking availability for both automobiles and bicycles. In this specific study, parameters not only allow assessing travellers' preferences that shed light on the necessary improvements in public transport networks for an even higher attractiveness of sustainable travel options, but also allow providing input to the public transport assignment model of the Danish National Transport Model.

The contributions of the thesis are thus to demonstrate a new survey-based data collection technique that can reveal passengers route choices in large and complex multi-modal networks, how such data can be map-matched and choice sets be generated for model estimation, and the results of the estimation of a multimodal route choice model based upon this data. Finally, the thesis describes revealed preferences and behavioural interpretations of the study.

DANSK RESUMÉ

Emnet for denne ph.d.-afhandling er adfærdsmodeller for rutevalg for passagerer i kollektiv trafik. Studiet har omhandlet den flersidede udfordring i at modellere rutevalg for trafikanter der bevæger sig i et multimodalt kollektivt netværk i en storby. Analyserne har fokuseret på RP data indsamlet i Hovedstadsområdet. Ph.d.-studiet løser de mange sider af udfordringen gennem (i) data indsamling, (ii) data analyse, (iii) valgsætsgenerering, samt (iv) modelestimering.

I ph.d.-studiet er udviklet en metode til at indsamle rutevalg for rejsende i kollektiv trafik. Der har tidligere været meget få studier omkring virkelige rutevalg for rejsende i kollektiv transport, da rutevalget for rejsende i et multimodalt kollektivt transportnetværk adskiller sig betydeligt fra rutevalg for bilister. I dette studie udvikles et spørgeskema til at indsamle data om rutevalget ved at spørge ind til tilbringer/frabringer trafik, benyttede stationer, linjer, afgangs- og ankomsttider, turformål og skift. Spørgeskemaet blev testet af studerende og ansatte på DTU og efterfølgende implementeret i den danske Transportvaneundersøgelse hvor der løbende indsamles data og indtil videre er indsamlet over 6.000 rutevalgsobservationer i kollektiv trafik i Hovedstadsområdet.

Endvidere genereres rutevalgsalternativer til brug for rutevalgsmodelestimering. Der defineres valgsæt for de enkelte rejsende mellem den start- og slutlokalitet den rejsende har angivet at have rejst imellem. Det kollektive netværk i Hovedstadsområdet består af mange forskellige kollektive transportmidler og de mange mulige kombinationer af transportmidler giver ofte et stort antal mulige ruter mellem de ønskede punkter. Valgsæt er genereret vha. en stokastisk metode hvor de opfattede omkostninger af attributterne i netværket samt variationer i de rejsende opfattelse af netværket kan varieres og hermed give mange forskellige ruter som output. De genererede ruter valideres vha. de observerede ruter, da metoden til at generere rutevalg bør kunne genskabe den observerede rute.

Endelig estimeres rutevalgsmodeller med de observerede rutevalg og de genererede valgsæt. Modellerne specificeres så mange forskellige rutevalgsattributter bliver undersøgt. Estimeringen bekræfter den forventede vigtighed af ventetid og skiftetid, viser forskellige præferencer for bus og tog, viser vigtigheden af turens længde, valg af transportmiddel til at komme til og fra stationer, og viser at de rejsendes præferencer varierer efter hvilket turformål de rejsende har.

Afhandlingens bidrag ligger i demonstrationen af en ny teknik til indsamling af rutevalg som kan beskrive passagerers rutevalg i et komplekst multimodalt netværk, og resultaterne fra estimeringen af den multimodale rutevalgsmodel baseret på dette data. Sluttelig beskriver afhandlingen de undersøgte præferencer og de adfærdsmæssige fortolkninger af studiet.

TABLE OF CONTENTS

1 INTRODUCTION

The transport systems in metropolitan areas often consist of various transport possibilities facing various challenges with the handling of the daily traffic flow. The choice of private transport modes such as car is traditionally the dominant transport mode choice and the road network suffers from congestion which affects the accessibility and mobility of the travellers and the economic centres. The public transport modes such as bus and train are more sustainable transport mode choices but less travellers use these because of challenges such as long travel time, low frequencies, low accessibility, low regularity, etc. Higher use of sustainable transport modes might be achieved by combining the private and public transport modes in a multimodal transport network. By increasing the use of sustainable transport modes the congestion can be reduced which both private and public transport users will benefit from. Also the environment will benefit from the switch in transport modes by being imposed to less pollution and other external effects.

When a traveller combines several transport modes for a multimodal trip he might be able to achieve higher benefits compared to a trip with only one mode. The opportunities offered to and the choices made by the travellers are important and the route choice of each traveller is an important factor. In order to choose the multimodal trip the chain of transport modes should fulfil the requirements of the traveller and a combination of elements are used to describe the structure of the trip in terms of route choice.

The investigation of the relevant factors that affect the route choice of the traveller is therefore very relevant and can guide stakeholders such as local governmental agencies and public transport agencies toward an effective improvement of the public transport services in metropolitan areas and of conditions for combining the private and public transport networks in a multimodal network. When improving the public transport modes their attractiveness compared to car increase. When improving the conditions for transferring between the private and the public transport networks the attractiveness of the multimodal network increases especially with respect to the car.

Accordingly, this PhD thesis faces the multi-faceted challenge of modelling route choices of travellers moving in a metropolitan multimodal network. The analysis focuses on revealed preferences data collected for the multimodal network of the Greater Copenhagen Area and solves the multiple facets of the challenge concerning (i) data collection, (ii) data analysis, (iii) choice set generation, and (iv) model estimation.

1.1 Research subjects

In short a number of issues motivates the research in this thesis:

- We see a limitation in the collection of actual route choice data for public transport users. Car user route choice models benefit from GPS data and map matching algorithms but the public route choice data are sparser. The collection is not as straightforward because information on transport mode, public transport line used, boarding and alighting location etc. has to be collected in order to describe the route of the traveller.
- Revealing the preferences of the public transport users would provide a better foundation for assignment models and a better understanding of the travellers to guide stakeholders in improving the conditions for the multimodal transport travellers.
- To model the data a set of available choices should be used and the literature shows that the generation technique and the composition of the final route choice set are important for the final results. The choice sets have to describe a range of route alternatives for the traveller to describe both the traveller's preference to the route attributes.
- The literature shows several proposals for models to describe the preferences of the public transport users. To model the public transport route choice the model should be able to account for similarities over alternatives and include a number of factors describing the route choice.
- Various approaches to generate route choice sets exist and these should be investigated in order to develop an appropriate model for estimating the factors relevant for the route choice.

The issues are sought to be solved and fulfilled using the following approaches:

- Collecting detailed information about the actual route choice of public transport passengers in the Greater Copenhagen Area using a special route choice questionnaire developed in this PhD project. The questionnaire has detailed questions about the public transport parts of the trips in a travel diary. The questionnaire method is detailed enough to collect the important route choice data and simple enough to keep costs and respondent burden down in order to collect the route choice observations from a large and representative sample of the Danish population.
- Developing a method to map-match the descriptions of the actual routes from the questionnaires to a GIS network to visualise the trips and to use for the model estimation.
- Describe statistically the route choice, mode choice and demographic data from the Travel Survey and carry out various analyses in order to reveal the interesting aspects of the data.
- Using a method accounting for heterogeneity among travellers perception of the network and preferences for the network attributes to generate alternative routes relevant to travellers by varying the scale parameters of this doubly stochastic method. When taking into consideration the variation in travellers' preferences and perceptions

the routes in the choice sets generated cover a large variation of the routes possible and relevant to the travellers.

- Using the coverage of the observed choices with the generated choice sets to provide a measure of the behavioural plausibility of the applied path generation technique.
- Specifying and estimating Path Size Logit and Mixed Path Size Logit models to describe the route choice preferences of the passengers in the multimodal transport system. The Path Size component takes into consideration the similarities between route alternatives and the mixing of the parameters enables the description of the travellers' different perceptions of the route choice attributes following specified distributions.

1.2 Main contributions

The main contributions of the thesis are:

- The development and validation of a new survey-based data collection technique that can reveal passengers route choices in large and complex multi-modal networks. By adding a list of questions including intelligent search for the via-points on the traveller's route, the method proves to be able to collect data with a precession enabling the exact reconstruction of the route through the public transport network. The thesis presents how this data can be map-matched onto a GIS network to use for analysis, visualisation and modelling purposes.
- The generation of choice sets for model estimation by using a doubly stochastic method accounting for heterogeneity among traveller's perception of and preferences for the network attributes. By varying the scale parameters of the parameters and the error terms a number of alternative routes are generated. For the origin-destination pair of each public transport traveller from the travel survey a choice set is generated and the coverage of the generated choice set is tested with the observed choice of the traveller. This provides a measure of how plausible the choice set generation technique is. The technique proves to reconstruct over 80% of the route for 99% of the observed trips calculated at stop level and for 88% of the trips assessed at link level.
- The results of the estimation of Path Size Logit models and Mixed Path Size Logit models for multimodal route choice in the large scale network based upon the observed trips and generated choice sets. The Path Size component ensures that the overlapping of the route alternatives is considered. The estimation of the Path Size factor models shows that in public transport networks routes which have a high overlap with other routes are more attractive to the travellers since the high overlap embeds a higher robustness to the route. The models include a high number of mode and route choice attributes such as travel time for each transport mode, characteristics of the transfers and transfer locations, etc.
- The mixing in the Mixed Path Size Logit model enables the description of the heterogeneity among travellers and the travellers' perception of the route choice attributes. The results show several public route choice attributes to be following lognormal distributions.

The thesis describes revealed preferences and behavioural interpretations of the study.

1.3 Contents

The following describes the contents of the thesis and provides reading instructions.

Framework

In Chapter 2 the framework of the thesis is presented to the reader. Focus is put on the definition of terms used throughout the thesis. The contents of the network of interest to this thesis are presented and each of the elements is illustrated and described to the reader. The geographical structures of the area are put into context, the public and private transport modes are presented and the fare structure is visualised.

Data collection method

Chapter 3 deals with the collection of public transport route choice data. After a literature review the Danish Travel Survey (TU) is presented and the process of developing and implementing an additional questionnaire to the national survey is explained. The questionnaire contains additional detailed questions on the trip part using public transport modes. The additional questionnaire was tested at a full scale test survey at DTU and the test and results from this test are illustrated. Also a few results from the national survey are presented. The method of map-matching the observations to a GIS network is explained and the results are visualised.

Public route choice data

In Chapter 4 the characteristics of the collected route choice data are analysed and the various factors affecting the actual choices of the public transport passengers are investigated. Two main analyses are carried out. An analysis of the choice between public and private transport is conducted followed by an analysis of the choice between unimodal and multimodal trips. Finally an analysis of the transfers in the network is carried out.

Generation and quality assessments of route choice sets

Chapter 5 deals with the generation of attractive routes for the origin-destination pair of each traveller. A doubly stochastic approach is used for generating alternative routes relevant to the travellers, since the method allows accounting for both perceived costs of the network elements and heterogeneity in the preferences of travellers. The coverage of the observed choices with the generated choice sets is used to provide a measure of the behavioural plausibility of the applied path generation technique.

Estimation of public route choice models

Chapter 6 describes the estimation of route choice models able to account for similarities across alternatives. The simple approach of a formulation of a Path Size Logit with definitions of similarity is tested and also the more elaborated approach with the formulation of a Mixed Path Size Logit is considered. The utility functions are specified with various factors describing access/egress characteristics, waiting time, in-vehicle travel time, and transfer characteristics.

Also the importance of travellers' characteristics, trip purposes and trip length on the preference structure is investigated. The results are presented, compared and discussed.

Conclusion

Chapter 7 includes a conclusion on the thesis and points out the main findings of the thesis. Also recommendations to stakeholders and suggestions for further work are provided.

The PhD project resulted in this thesis and the following papers:

Anderson, M.K.A. (2010a). Characteristics of Trips and Travellers in Private and Public Transportation in the Danish Travel Survey data. In *Selected Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Anderson, M.K.A. (2010b). Development and Assessment of a Data Collection Method for Route Choice in Public Transport. In *Selected Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Anderson, M.K.A. & Rasmussen, T.K. (2010). Matching Observed Public Route Choice Data to a GIS Network. In *Selected Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Larsen, M.K. (2008). Indsamling af Data for Rutevalg i Kollektiv Transport. In *Selected Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark (in Danish).

Larsen, M.K., Nielsen, O.A., Prato, C.G. & Rasmussen, T.K. (2010). Generation and Quality Assessment of Route Choice Sets in Public Transport Networks by means of Data Analysis. In *Proceedings of the European Transport Conference,* Noordwijkerhout, the Netherlands.

2 FRAMEWORK

In this chapter the framework for the thesis is set up. The terminology used in the study is presented for the reader to have an introduction to the field of research.

The description goes over the definitions used regarding the multimodal transport network of the Greater Copenhagen Area. The public transport network of the Greater Copenhagen Area is presented and the public transport modes are described to give the user knowledge of the transport network. Examples of alternative routes are offered and the fare structure in the public transport network is presented.

2.1 Main definitions

In this thesis *multimodal transport* is described by a set of definitions which are important to keep in mind when reading the thesis. The definitions are mainly concerning the travelling of the public transport passengers and describe the various parts of the trips.

A trip is defined as the travel between the origin point and the destination point. A traveller most often has at least two trips during a day (home->work and work->home) and often more. A trip consists of one or several legs each defined by a transport mode.

2.1.1 Transport modes

To describe the transport modes in the multimodal transport network of the Greater Copenhagen Area, we distinguish between private and public transport modes where private transport modes are modes that the traveller and/or his family/household (or similar) have at their disposal.

<u>Private transport modes</u>

The private transport modes included in the study are:

- Walking.
- Bicycle.
- Car.

Unless disabled, walking will always be at the traveller’s disposal, but the travellers have various perceptions of how long distances they are willing to walk and various travel speeds. As part of a multimodal trip, walking is often used to get to the first and from the last vehicle and between transport modes if the trip involves transfers.

Most travellers in the Greater Copenhagen Area (76.4% according to the Danish Travel Survey, TU) have a bicycle at their disposal but the availability of the bicycle often depends on the location of the traveller. The bicycle will often be located at the home of the traveller and is therefore only available at other activity locations if brought from home on a previous trip.

The household might have availability of a car but if the household has one car only and multiple family members owning a driver's license the car might not be available to all of them. Also the car availability is depending on other trips during the day, for example can a traveller only use the car from a train station to home if the traveller (or perhaps another user of the car) parked the car at the station on an earlier trip. Chapter 4 looks further into the car availability in Denmark by analysing the TU survey data.

Public transport modes

The public transport modes in the Greater Copenhagen Area are:

- Buses (high-frequency, express, regular).
- Trains (intercity, regional, suburban, urban and local).
- Metro.

The different types of buses serve all locations of the Greater Copenhagen Area. Trains serve the Central Business District (CBD) of Copenhagen and lead in radial lines from Copenhagen to the other cities in the area. The metro serves the cities of Copenhagen and Frederiksberg, and the Kastrup Airport (see section 0 for a description of the public transport modes types).

The modes are available for all travellers but whether or not the mode is considered as an alternative for the traveller depends on the geographic locations of origin and destination and the route alternatives for the traveller. In the city of Copenhagen all of the above modes (except for local trains) are available for the travellers (within a given distance) and away from the urban areas most often only buses are available close to the origin.

2.1.2 Trip legs

Each of the above mentioned transport modes represents a trip leg. A trip leg is defined as the use of one specific transport mode and starts at the point of boarding the transport mode and ends at the point of alighting the transport mode. A trip leg is described by:

- Point of boarding.
- Transport mode used.
- Whether the traveller is a driver or passenger.
- Time travelled (incl. waiting time at transfers).
- Distance travelled.
- Point of alighting.

Since the transport vehicle defines the trip leg the legs can have very various lengths. A train leg can be very long taking the traveller from the most northern to the most southern point in the Greater Copenhagen Area, but it can also be short from one station to the next. Walking legs are most often short legs because of the relatively low travel speed. In multimodal transport networks walking are used for access, egress and transfer legs in combination with one or more (longer) trip legs using a public transport mode and for access and egress to/from bicycle and car.

Access and egress legs

In this thesis we define the access and egress legs as the trip legs leading to the first and from the last public transport mode. By this definition a private transport mode will always be used at the access and egress legs. Also a trip can have several access modes if the traveller walks to the bicycle and then bike to the first public transport stop. The access legs end and the egress legs start at the points where the road/path and public networks are connected to each other.

2.1.3 Trips

We define a trip as the travel between the geographic locations of the origin and the destination. The trip between the two points of location consists of all movements in time and location involved in taking the traveller from origin to destination. In this thesis a trip is described, among others, by:

- Departure time.
- Location of origin.
- Location of destination.
- Arrival time.
- Trip purpose (at destination).

The actual route of a traveller travelling through the network is described by the trip legs between origin and destination. A trip can consist of one or several trip legs. When several trip legs are used the order of the trip legs is described by the chain of transport modes.

This thesis mainly deals with multimodal trips and the definition and characteristics of such are described in section 2.2.

2.1.4 Main transport mode/primary mode

When assessing all trips legs on a trip from origin to destination made by the traveller the main transport mode can be identified. In this thesis we define the main transport mode as the mode used for the longest trip leg measured in distance. The access and egress legs can also be very long measured in time (depending on the travel speed and the waiting time at the stop/station) but most often the longest leg measured in distance will be made by a vehicular transport mode.

2.1.5 Journey

A journey is defined as a combination of a series of trips leading from the origin through a number of destinations and back again. The origin is the primary base location of the traveller, often home, and the destination is the main destination of the journey, often work for commuters etc. The term is not used much in this thesis since the focus is about the trips.

2.2 Typical characteristics of multimodal transport networks

The multimodal transport network is characterised by being a combination of two sub-networks: the private and the public transport network. The private transport network can consist of several networks, a path network for walking and cycling and a road network for the car. The public transport network consists of the road network for buses, one or several networks for rail (metro, train, local rails, etc.) and schedules or frequencies for the public transport modes. The private transport network is a continuous network whereas the public transport network is discontinuous meaning that the travellers have to wait at stops and transfer between the various public transport modes used. When using two separate networks, connections need to be added to create one network where it is possible to travel from the origin to the destination.

In the travel from origin to destination, specific conditions for the order of the transport modes can be defined. The fastest transport services will often be used for the longest part of the trip (train, fast buses) and the slower modes are used for access and egress trip parts. Private vehicles have restrictions for use in the mode chain since they are often home-based and only possible to use between modes or at the activity-end if, for example, the traveller has placed a bicycle at the train station or brought one along in the train.

The multimodal transport networks are often very large networks in terms of number of links and nodes and also offer a large number of alternative routes. Various bus services serve the same stops, some trains are faster than others and the composition of routes and route alternatives can change much depending on the time of day (if the schedules of each public transport mode used have different frequencies over the day) or be almost identical.

Accounting for similarities among routes is more challenging than in the car road network alone since similarities can be accounted on both the stop and the link level, which is discussed in Chapter 6.

In principle a multimodal trip can be divided in three trip parts: access – main transportation – egress. Some behavioural characteristics describing these trip parts can be found in Hoogendoorn-Lanser (2005). When travelling through the network many attributes of the network and the trip parts are important for the traveller’s choice of route.

The following list of behavioural conditions important for the route choice in multimodal transport networks is inspired by a focus group interview on public transport route choice behaviour from 2003 of students from the Technical University of Denmark (see Appendix 1 for the full description of the focus group interviews), the in-depth investigation of behavioural attributes important in a multimodal transport networks by Hoogendoorn-Lanser (2005), and the investigation by Litman (2008) of the factors affecting travel time cost in public transport trips.

Complete trip

- Only limited detours in terms of distance, time, or number of transfers is accepted by travellers.
- Travellers have different preferences for the characteristics of the transport services for example for seating comfort, noise, driving pattern:
    - Trains are more comfortable than buses: trains do not stop as often as buses, not mixed with other traffic so appears safer when operated, not as many hard breaking manoeuvres as in buses.

    - Buses are more comfortable than train: possible to sit more private (two-person-seats in the bus, four-person-seats in the train facing each other), easier to get contact with the bus chauffeur than the train driver.
- Travellers have different perceptions of the elements of time: walking, waiting, in-vehicle, and transfer time.
- Most travellers prefer public transport services that operate at frequent intervals (the waiting time minimises if the headway is low), that routes are direct, that the scheduled public transport trips are operated, that the public transport services operate on time, that transfers are maintained (that buses do not leave before scheduled arrival when transferring from train to bus, if the train is delayed, ideally, the bus should wait for transferring travellers).

Transfers

- Travellers have both a minimum and a maximum for what they will accept in terms of transfer time, transfer-waiting time and transfer-walking time. The minimum is to assure that the next connection can be reached in time and the maximum is to avoid too long total travel time.
- Travellers prefer transfers at high-order bus stops and train stations over low-order stops, both because of frequency and because of the shelter from rain and wind.
- Travellers prefer to transfer at vibrant locations with shops, and that the waiting areas are clean, attractive, well-lit and accessible.
- Travellers do not use unnecessary transfers such as transferring to a lower order or equal train if the current train is stopping at the station preferred by the traveller.

Access/egress mode

- Most travellers choose the bus stop or train station closest to the origin because a large uncertainty factor is related to the walking (risk to miss a planned public transport run if walking takes longer than planned). At the destination, the departure stop (and thereby the walking distance to destination) is not as important because an increased walking time will not impose delays to the chain of modes.
- Most travellers do not use the bus as access to train for short distances. It is faster to walk or bicycle than to walk to a public transport stop, wait for the public transport service to arrive, ride the public transport service, and walk to the destination.
- Most travellers do not use the bicycle or the car as access modes for short distances to the public transport service.

Some of these are of course assumptions and can be different from the above if the traveller has other agendas. For access/egress if a public transport season ticket holder walks a short distance and observes the bus approaching he might board this even for a short distance to save travel time. Also the traveller might use the bicycle as the transfer mode for a short distance if the bicycle is brought along in the train.

The DTU focus group interviews on public transport route choice behaviour identified the most important aspects for the travellers to be:

- The total travel time.
- High frequency of public transport modes.
- A limited increase in travel time is accepted if transfers can be avoided or the waiting time can be reduced.

2.2.1 Definitions regarding multimodal trips in the public transport network

The following section presents and describes definitions of multimodal transport especially used the analysis in Chapter 4.

A *trip* is the combination of *trip legs* (including transfers) used for travelling between two locations (e.g., from home to work, from work to the supermarket, or from the supermarket to home). Many previous analyses made on the TU data (Christensen, 2000, Christensen, 2001, Christensen and Jensen, 2008), have been based on the *journeys*. In this survey, it was important to have the information on the exact trip since, for example, feeder modes can be very different when travelling from home to work and from work to home. By this definition the majority of the trips is generated from and attracted to home. In fact, after work people leave for their home, after shopping they return home, and after sport activities in the evening they go back home.

When travelling through the network travellers use one or more transport modes. For each part of a trip a new transport mode used is defined as a leg. The transport mode used for the longest leg in distance is defined as the *primary mode* (also referred to as the *main transport mode*).

A public transport trip will always be a multimodal trip consisting of several legs and both public and private modes (walking, bicycle, car or combination). A multimodal trip starts and ends with a walking leg. The first walking leg often brings the traveller to a bus stop or a train station or if several access modes to the public transport system are used the walk can bring the traveller to the car or bicycle.

<u>Chapter 4 analysis</u>

In the analysis in Chapter 4.3 *multiple-leg* public transport is defined as more than one transport mode vehicle used for a trip between origin and destination. *Multimodal* transport is defined as the use of at least one public and one private transport mode for a trip between origin and destination. In the network, the motorised transport modes are categorised into the following groups: car driver (car or van), car passenger (car, van or taxi), bus, suburban train, other train (IC, regional and local), metro, and other (truck, tractor, tourist bus, ferry, boat, airplane). The bicycle is an often used transport mode in the Greater Copenhagen Area, both as feeder mode and primary mode, and defined as a transport mode in terms of multimodal travelling as well. Walking is used for short parts of all trips, leading to and from each transport mode (access, egress and transfers). Travellers walk from their starting point to the first mode (e.g., from home to bicycle) and between modes (e.g., from bicycle to bus, etc.). If walking was defined as a transport mode, all trips (except pure walking trips) would be multimodal, and walking is therefore not included as a transport mode in this context. The observed trips are divided into the following groups for analysis purposes:

- *Unimodal*: one vehicular mode (bicycle, train, car driver, car passenger, etc.). Trips with use of several private modes (ex. bicycle as a feeder mode to car passenger) are also in this group.
- *Multiple-leg public*: use of two or more public transport modes (two buses, two trains or a bus-train combination).
- *Multimodal public and private*: use of at least one private vehicular transport mode and one public transport mode (for example, a car-train combination).

In the analyses referring only to public transport, a few trips (3% of the total number of public transport trips in the sample) with public transport mode used for the second longest and private used for the longest part of the trip (measured in distance) were also defined as using public transport as the primary mode. Examples of such trips are:

- Leisure trips from/to visit family/friends where the traveller is passenger in another visitor's car.
- Commuter trips using the car from home to a station, parking the car, and using train to work (and opposite).
- Using the bicycle from home to the metro.

Distance

In the initial analyses when using distances between two points (e.g., from the origin to the destination of the trip, from either home or work to the nearest train station), the straight line distance between the two points was calculated. Considerations were made whether travellers assess the distance between two points as a straight line or as the distance in the transport network. When using straight line distance, there is a problem in the fact that short and long distances may not be exactly comparable since people travelling long distances often use the primary road network for a great part of the trip, making it possible to travel in rather straight lines, while for the short trips people use the secondary road network with more detours. Also, the fact that Denmark is a country consisting of islands and having many fjords could cause the trip distance of the long distance trips to be different from the straight line distance.

However, to find the *distance travelled* the route actually travelled needs to be identified and the routes should be reconstructed. For trips using public transport the *exact distance travelled* in the vehicles can be reconstructed by using the route choice data collected for use for this thesis (see Chapter 3 for a description of the data collection and Chapter 4 for reconstruction of the observed public transport routes). For trips using private vehicles, the route choice is not known and many assumptions would have to be made to reconstruct routes. The routes can be reconstructed using a shortest path calculation in a network, but many other factors than distance affect the choice of route and it was not the scope of this thesis to calculate the route choice for the private trips.

Considering the above, it was chosen to use the *straight line distance* for the initial analyses in Chapter 4 (comparing private and public transportation). The coordinates of the TU data are an exact measure and the straight line distance is comparable from one trip to another. In Chapter 5 and 6 analysing the public transport trips the exact distance derived by matching the observed routes to a public transport GIS network is used.

2.3 Characteristics of the Greater Copenhagen Area multimodal transport network

2.3.1 The geographical area

The area of interest in this PhD thesis is the Greater Copenhagen Area including and surrounding the Danish capital Copenhagen (København) pointed out in Figure 2-1. Copenhagen is located in the eastern part of Denmark close to the sea border to Sweden. Two million people live in this area, which is the most densely populated area in Denmark.

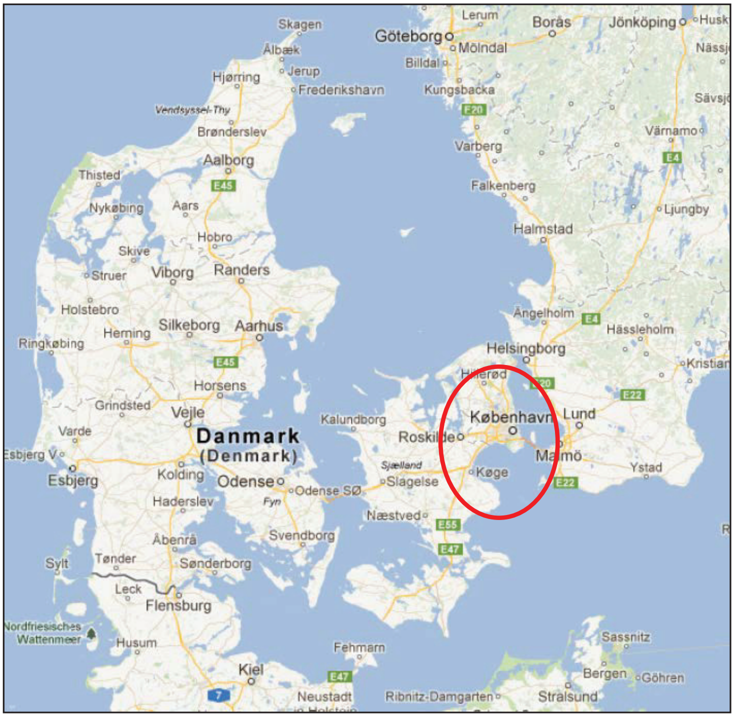

Figure 2-1: The Greater Copenhagen Area (Google maps).

The Greater Copenhagen Area consists of the city of Copenhagen and Frederiksberg in the attraction centre of the region and the five greater cities Køge, Roskilde, Frederikssund, Hillerød and Helsingør. Via the Fingerplan directives (Fingerplanen, 2007) the planning and investments in development focus on the fingers leading from Copenhagen to the five cities served by public transport on rails and motorways or other greater road connections. Figure 2-2 presents the greatest roads and the rail network in the Greater Copenhagen Area. As seen the rails and major roads lead to the five cities and the built-up areas to some extend follow the alignments of the roads and rails.

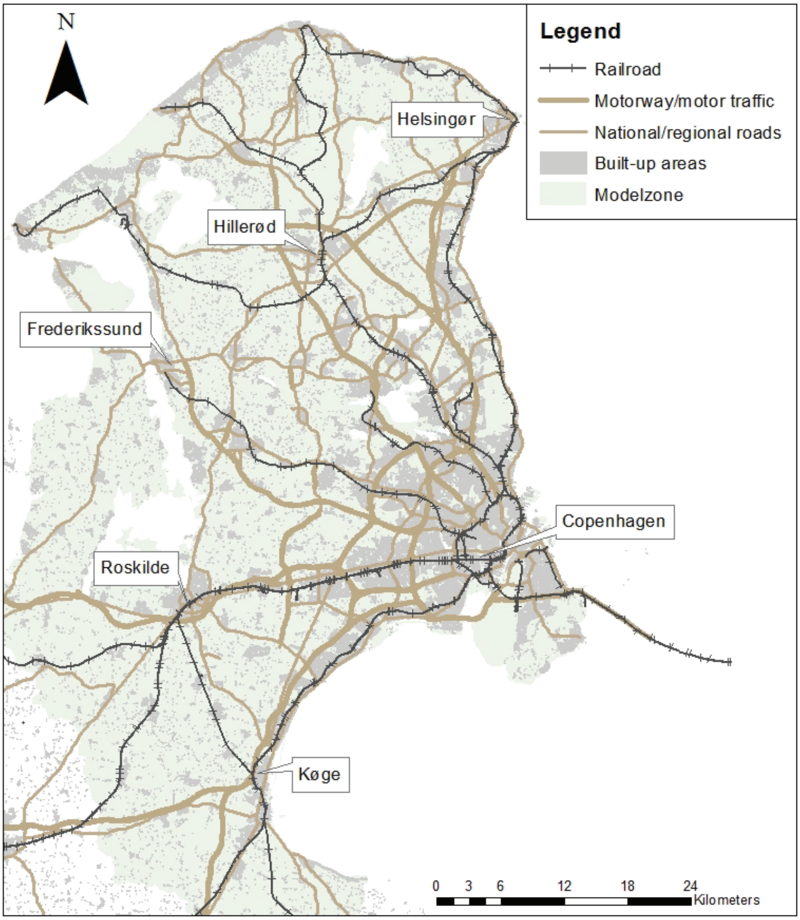

Figure 2-2: The network of greater roads and railroads in the Greater Copenhagen Area.

2.3.2 Public transport modes

The public transport network in the Greater Copenhagen Area consists of seven major transport mode groups:

- Bus.
- Express bus and suburban bus (E- and S-bus).
- Local, high frequent bus (A-bus).
- Suburban train (S-train).
- Regional and Intercity train.
- Metro.
- Local train.
- Ferry/harbour bus.

The public transport network in this area is dominated by the regional trains leading north and west of Copenhagen, and the urban and suburban train lines (S-trains) serving the CBD of Copenhagen and leading in five radial lines to the other cities in the Greater Copenhagen Area (see Figure 2-3). The S-buses and some high frequent buses serve the S-train stations and are primarily driving in rings around Copenhagen. A part of the metro has underground rail network and serves the CBD of Copenhagen with three separate alignments. A-buses are local, high-frequency buses serving the inner central part of Copenhagen. The remaining buses are a variety of buses in Copenhagen, in the suburbs and in the rural areas. The local trains are serving the cities near Copenhagen and the rural areas close to and between these cities.

The various mode types have very different service level in terms of frequencies, see Table 2-1.

Table 2-1: Frequency of public transport modes types in the Greater Copenhagen Area [headway in minutes].

Public transport Mode type	Frequency
A-bus	Approx. 3 min in day times
E bus	Approx. 10 min during rush hours
Bus	10-60 min
S-train	10 min
Metro	2-4 min in day time
Regional and IC-train	20-120 min (some lines only one departure per day)
Local train	30 min
Harbour bus	30 min

The regional trains have very high headways and are primarily used for longer trips. The comfort is the trains are high but in peak hours seating might be difficult. S-trains are very popular since they serve the inner parts of Copenhagen and leads to the cities around. The S-trains leave in fixed minutes so a traveller familiar with the trains always knows when the next is departing. In the city of Copenhagen seating is often difficult. The local trains often have very good connections to the buses in the so called R-network. This means that buses and trains are synchronised so there is connections between the two every half hour. Also the regularity of the local trains is higher than for the trains in general.

For some train station OD pairs, both metro and S-train can be used. In these cases the metro will often be the chosen mode because of the higher frequency. Similar cases exist for S-train and regional train and in these cases the S-train is often chosen because of the higher frequency. The higher frequency means (in average) lower waiting time and higher regularity. Chapter 4 looks further into the choice between metro and S-train for overlapping OD pairs.

The public transport network of the Greater Copenhagen Area is shown in Figure 2-3. The Modelzone layer represents the area of interest and the train stations are coloured according to their type. From the location of the stations it is possible to see that the stop patterns are very different for the different train types. Metro stations are located very close, S-train and local train stations are longer apart and finally the regional train stations are the most separated. For both S-train and Regional/IC-trains some runs also skip train stations.

The areas served are also varying for the different public transport service types:

- The CBD of Copenhagen is served by the *metro and A-buses.*
- The S-*trains* serve the “fingers” to the other great cities in the area.
- The *S-buses* connects the S-train lines by driving in rings around the city of Copenhagen.
- The *local trains* mainly serve the northern part of the region.
- The *regional and IC-trains* serve the legs leading from the area to other parts of Denmark.
- The *other buses* serve all the remaining areas often also serving a train station.

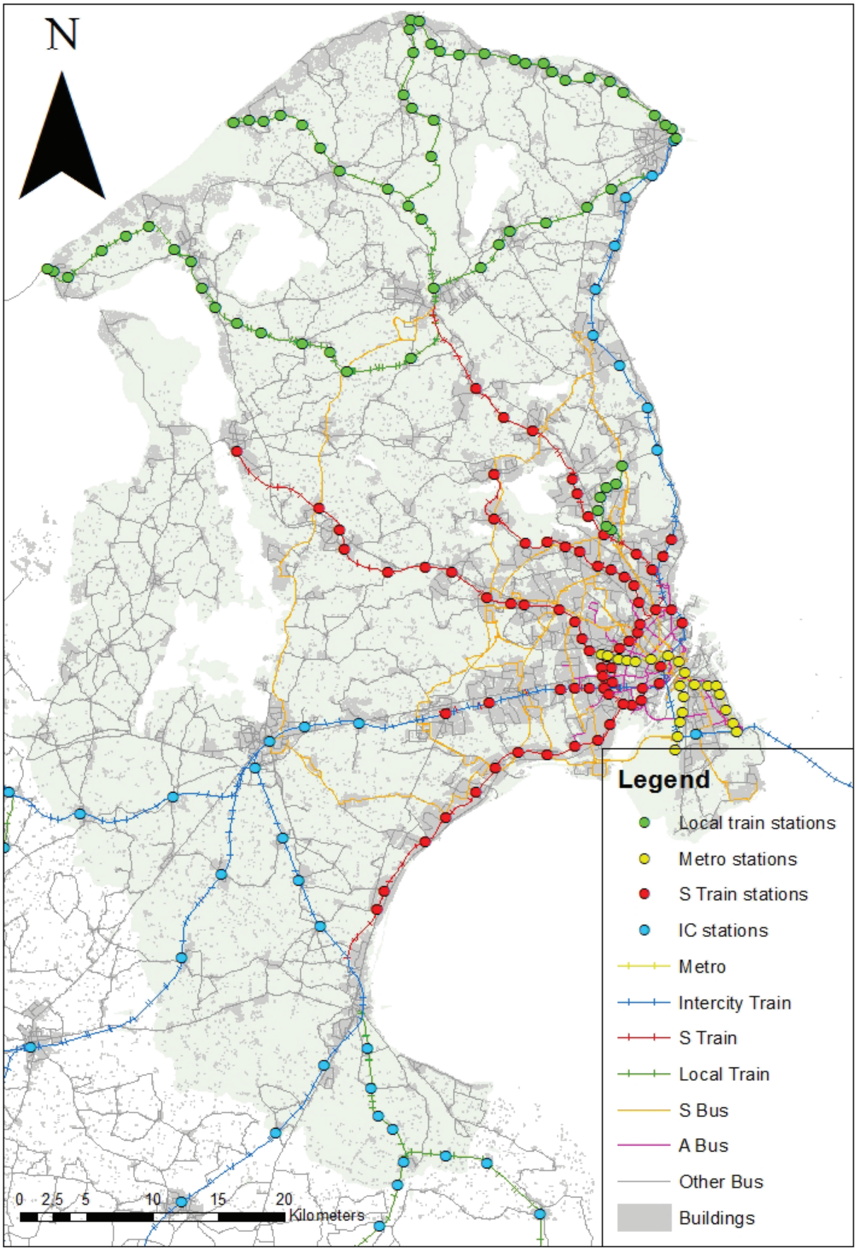

Figure 2-3: Train and bus lines in the Greater Copenhagen Area.

In Figure 2-4 is shown a zoom of the public transport network closer to the City of Copenhagen. The high frequent A-buses serve the city in a grid system which provides easy access to the bus system for most travellers in the city.

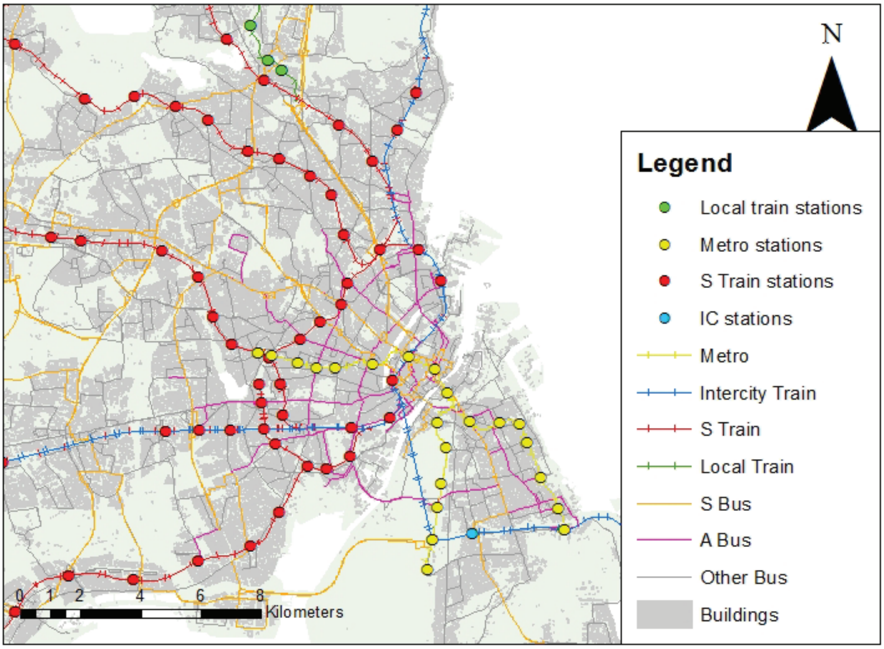

Figure 2-4: Train and bus lines in and around the city of Copenhagen.

2.3.3 Train station transfers

Figure 2-5 shows a schematic overview of the S-train lines, metro lines and regional train lines in the Greater Copenhagen Area. Nørreport St. is the only station served by all train types connecting the urban and suburban train systems to the metro rail network.

Figure 2-5: The S-train line network (schematic – not identical to the actual topology) (from http://mapsof.net/map/copenhagen-stog-metro-districts 2013-09-10).

Other train stations are served by two train types and make it possible for the travellers to transfer between the train systems. The following train stations offer transfers between two or more train systems:

- Nørreport st. (Regional and IC-train, S-train, Metro).
- København H, Central St. (Regional and IC-train, S-train).
- Køge st. (Regional train and S-train).
- Flintholm st. (S-train and Metro).
- Hellerup st. (Regional and IC-trains and S-train).
- Ørestad st. (Regional trains and Metro).
- Kastrup Airport st. (Regional and IC-trains and Metro).

At a number of stations it is possible to transfer between two or more S-train lines crossing each other:

- Ny Ellebjerg st. (line F crosses A and E).
- Danshøj st. (line F crosses B and B+).
- Flintholm st. (line F crosses C and H).
- Ryparken (Line F crosses line A and B+).
- Hellerup (line F and C cross line E and B).

With this number of stations making it possible to transfer within the train network or between two train networks the number of alternative routes can be high. If the traveller prefers one train type over another it is possible to transfer to this line for a number of origins and destinations. This makes the public transport network of the Greater Copenhagen Area very convenient for use in examining the route choice preferences for passengers in public transport.

In the train networks several types of choices are available.

S-train example

When travelling on S-train line B from a station between Høje Taastrup st. and Hvidovre st. to Hellerup st. several options are available:

- The traveller can stay on the B line boarded (24 min from Hvidovre st.).
- The traveller can transfer to the F line at Danshøj st. (19 min from Hvidovre st., incl. 2 min of transfer).

The second option is the fastest but also includes a transfer. The choice between the two routes therefore depends on the travellers' preferences for in vehicle time versus the inconveniences of transferring.

S-train to metro example

When travelling on line C or F from a train station between Frederikssund st. and Jyllingevej st. to Nørreport st. the traveller has several options:

- The traveller can stay in the S-train (19 min from Jyllingevej st.).
- The traveller can transfer to a metro line at Vanløse st. (16 min from Jyllingevej st. incl. 4 min of transfer).

These are just two examples of alternative routes within the train network. For the bus network and the bus and train networks combined, there is a high number of alternatives for the traveller depending on his origin and destination location points. When considering the entire multimodal transport network also including access and egress to the public transport networks and choice of boarding/alighting stops/stations the network provides a large variety of alternative routes making the multimodal transport network of the Greater Copenhagen Area suitable for the data collection, analysis, and route choice model estimations carried out in this thesis.

2.3.4 Fare structure

The public transport network of the Greater Copenhagen Area has a fixed fare structure. A ticket (or public transport seasonal card) can be used in all public transport modes listed in Table 2-1. This means that the transport mode types do not compete on fares. The fare for a given trip is determined by the location of the points of origin and destination. All zones are numbered and the fare depends on the number of zone rings between the origin and the destination. The zone structure and numbering of the zones is seen in Figure 2-6.

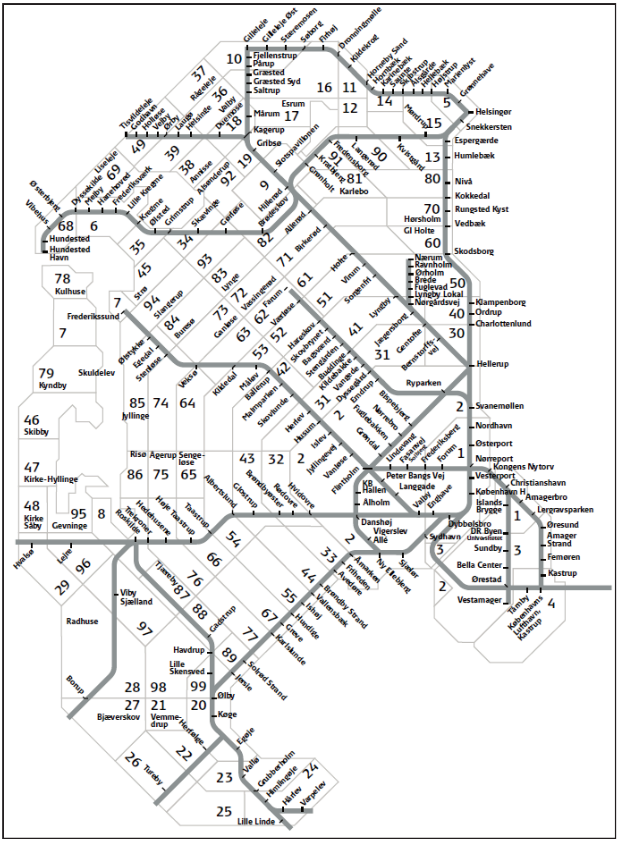

Figure 2-6: The fare structure of the Greater Copenhagen Area - from www.moviatrafik.dk 2013-09-10.

Public transport season tickets

Most commuters travelling by public transport have a monthly public transport season ticket where they pay in advance for the number of zones they travel via. On the card is registered the exact zones the traveller has paid for and it is not allowed for him to travel via other zones without paying extra. The cost of adding extra zones for a single trip equals the fare paid per zone for travellers buying a single ticket. According to the TU survey, 58.3% of the travellers travelling within the Greater Copenhagen Area are using public transport season tickets as fare payment. Of all travellers (private and public), 21.8% own a season card (66.5% of the public transport travellers).

Public transport tickets

The most expensive payment method is to buy a ticket every time a new trip is started. The traveller using tickets pays for the number of zones travelled. The price of the trip is calculated by use of the zone farthest away from the start zone and not necessarily the destination zone. Examples of the fare structure calculation are shown in Figure 2-7. According to the TU survey, 4.1% of the travellers within the Greater Copenhagen Area used public transport tickets as fare payment.

Public transport multiple-ride tickets

In the Greater Copenhagen Area it is also possible to use public transport by paying with a multiple-ride ticket. The multiple-ride ticket covers payment for ten rides and is bought in advance. It is possible to buy multiple-ride ticket from two to nine zones. Nine zones are the maximum you can pay for so the price is constant when travelling through nine zones and more. According to the TU Survey 32.8% of the travellers within the Greater Copenhagen Area used public multiple-ride tickets as fare payment.

Figure 2-7 shows the fare structure calculated from the point of origin. The fare for the trip is calculated by locating the point travelled via which is farthest away from the origin point. This is often, but not necessarily, the destination zone. The trip fare depends on the colour (according to distance) and increases in price in the following order (red is the zone of origin for the trip): red, blue, yellow, brown, purple, orange, green, pink, and grey.

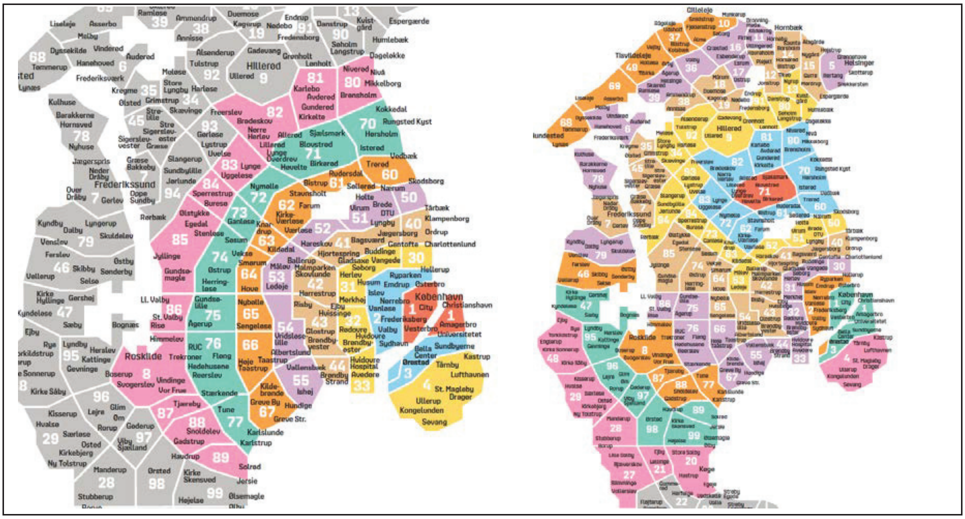

Figure 2-7: Fare structure used for fare calculation seen from inner Copenhagen (red zone left picture) and Allerød (red zone right picture) – from www.moviatrafik.dk 2013-09-10.

Smart card

The fare system for the public transport network of Denmark, including the Greater Copenhagen Area is currently being updated to smart cards. Within a few years the smart card will be the only option for reduced fare payment and the payment methods of season ticket and multiple-ride ticket as described above will partly be incorporated in the new system. Even when the smart card is fully incorporated the most expensive payment method of buying single tickets will still be available (www.trm.dk 2013-09-10).

The smart card comes in three versions:

- The personal card - follows the person but he is allowed to pay for extra passengers, free to acquire.
- The flex card – can be shared among travellers, costs 50 DKK (7 Euro).
- The anonymous card – no registration information is required when buying the card, costs 80 DKK (11 Euro).

The smart card system is used in trains all over Denmark and in buses in most of Denmark.

2.4 Summary

In this chapter the main definitions for this PhD thesis have been presented and described. The terms used in the description of trips have been presented as follows.

A *trip* is the term used for a travel from a point of origin to a point of destination. A trip in a multimodal transport network consists of at least three *trip legs*, where each trip leg describes a use of a *transport mode*. A transport mode is either private or public. In multimodal transport networks the private modes are used as access modes to the first and egress modes from the last public transport stop or station and to transfer between public transport modes.

The thesis deals with the multimodal transport network of the Greater Copenhagen Area. The public transport modes of the network are presented to be:

- Bus.
- Express bus and S-bus (E and S-bus).
- Local, high frequent route (A-bus).
- Suburban train (S-train).
- Regional and Intercity train.
- Metro.
- Local train.
- Ferry/harbour bus.

With various frequencies, stop patterns, areas served, etc. Some examples of alternative routes in the network are presented to give an overview of the high number of alternative routes in the multimodal transport network offered to the traveller.

Finally the fare structure of the area is presented. The public transport network of the Greater Copenhagen Area has a fixed fare structure depending on the number of zones travelled through. Public transport travellers have four options of fare payments;

- Public transport season tickets.
- Public transport tickets.
- Public transport multiple-ride tickets.
- Smart card (which are planned to take over season tickets and multiple-ride tickets within few years).

3 DATA COLLECTION APPROACH

In this chapter the data from and the design of the data collection method developed in this PhD study are described and statistical results of this questionnaire survey to identify route choice in public transport are presented. The survey builds upon the existing Danish Travel Survey, TU, to which was added questions concerning route choice. As a part of the PhD study, the questionnaire was tested via a full scale test among students and staff at the Technical University of Denmark (DTU). The results of the survey are described focusing on distribution of respondents and choice of transport mode, and it is considered which pros and cons such surveys have. The promising results from the PhD test survey resulted in the adding of the questions to the National Travel Survey (TU) and results from this are presented in this chapter. To reproduce and visualise the data a technique using Geographic Information Systems (GIS) was implemented in connection to the PhD study and this method is also described in this chapter.

The test survey had a reasonable number of participants and the results indicated that it is possible to include route choice for public transport passengers in the TU Survey without resulting in a significant drop-out from the survey. This gives the possibility of collecting a large amount of high quality route choice data for public transport users.

The method of collecting public transport route choices proved to be successful since it collected actual route choice data for a large number of travellers and it allows reproducing observed route choices in a GIS network for route choice set assessment purposes.

This chapter builds on the work presented in Anderson (2010b) and Anderson and Rasmussen (2010).

3.1 Introduction and literature review

Knowledge about actual route choices for public transport passengers is important when assessing generated choice sets for route choice modelling and for estimation of route choice models.

Route choices for car traffic have attracted a lot of attention in the research literature, but limited knowledge exists on the route choices of public transport passengers. One of the reasons lies in the difficulty to collect data on actual route choices in public transport networks, since a lot of information has to be provided to describe the routes actually used by travellers. For private transport it is possible to use GPS devices to track routes and then map the data to a physical network (see, e.g., Jan et al., 2000; Schönfelder et al., 2002; Wolf, 2004; and Zabic, 2011). For public transport the same method is of little help for many reasons. Relevant information about the public transport lines used is not retrievable with these devices since often several bus or train lines use the same roads or rails. Because of the limited signal strength and missing visibility of satellites, the signals may fall out in tunnels and, since in the Greater

Copenhagen Area the metro and sections of the urban rail system (S-trains) are in tunnels, a great part of the data cannot be collected. Also information on the trip purpose, which is another fundamental piece of information for uncovering route choice determinants, is not retrievable with GPS devices. Information about the purpose can be collected afterwards through a combination of GPS devices and questionnaires but such a setup will make the survey very extensive and the respondent burden will be high.

3.1.1 GPS and smart phone data collection

The accuracy of GPS data collected for use in route choice analyses for private transport has improved in recent years (see Holm, 2009 and Zabic, 2011). The development has improved from a large share of missing points as reported by Nielsen (2004b) (missing information for 90% of the trips in Copenhagen) to more accuracy in both the device technology and the number of satellites. But still the use of this data source for route choice knowledge is questioned (see Bierlaire and Frejinger, 2009).

On-person GPS devices have been used to log the movements of the person carrying the GPS device. Bricka and Wolf (2008) reported the use of GPS devices in Chicago and pointed out several problems in representing transport modes by this method. If the rail alignment is along a highway and the trains and cars travel with the same speed the algorithms used to deduct the transport mode used might not be detailed enough to assign the correctly used transport mode. Also GPS signals fell out in the underground rail network, in the urban canyons of central Chicago and other times when the signal was blocked. Rasmussen et al. (2013) documented the data collection and data processing from the on-person GPS devices in the ACTUM project. They showed a 92% correct match for identifying transport mode and above 70% correctly identified bus trip legs and exact bus line.

Chung and Shalaby (2005) and Tsui and Shalaby (2006) used GIS software to map-match the observed data from on-persons GPS devices in Jakarta to road and public transport networks to identify the route and transport modes chosen by the traveller. Chung and Shalaby (2005) used a rule based approach using the average speed and rules about the locations of switching transport modes (near bus stop or not) to identify the transport mode and identified correctly 79% of the links and 92% of the transport modes. Schuessler and Axhausen (2009) processed GPS data from on-person GPS devices collected in three Swiss cities and developed algorithms to identify trips and transport mode for each trip leg. Tsui and Shalaby (2006) and Schuessler and Axhausen (2009) used fuzzy logic approaches using characteristic of the transport modes such as speed and acceleration to identify the transport mode. Schuessler and Axhausen (2009) did not have actual behaviour to use in the evaluation of their results, but reported finding the same tendencies as in the Swiss Microcensus on Travel Behavior. Tsui and Shalaby (2006) found a correct detection of walking and car mode of 91% and 97% but the bus detection rate was as low as 76%.

Stopher and Greaves (2007) described the use of GPS devices together with a traditional travel diary and predicted that the future national travel surveys will rely more on GPS data. Chen et al. (2010) and Gong et al. (2012) reported the findings from a travel survey collecting data in a multimodal transport network by the use of GPS devices, and the development of GIS algorithms to determine the transport modes used and the trip purposes for the trips. Up to 90% of the transport modes were correctly identified but the identification was as low as 29% for rail and 53% for buses showing that these methods are still not sufficient for route choice data collection in public transport networks.

The process of collecting travel patterns using an on-person GPS device has a high respondent burden for the participant; the device has to be recharged to prevent data loss, often the participants forget the device, etc. The respondent burden can be reduced by using instead the mobile phone to collect travel observations. The travellers most often have the mobile phone with them. The use of mobile phone applications for the travel data collection was documented in Japan by Itsubo and Hato (2006), for Munich, Deutschland by Mandir et al. (2010) and for San Francisco, California by Hood et al. (2011). The surveys using these applications faced the same difficulties as GPS devices and the issues of matching the observations to transport modes and identifying travel activities caused this data to be far as precise as required for the route choice analysis in this PhD study. Ohmori et al. (2006) described a mobile phone system both collecting GPS observations and providing a travel diary application. Because of the continuous GPS logging and data transmission the battery life was reduced to 6 hours. The mobile phone data collection methods are promising and even more because of the improvements in batteries and other technologies. But since the ownership of the mobile phone is low in some age groups, it is difficult if not impossible to obtain a large and representative sample of the population at this point in time.

3.1.2 Automated fare data collection

Several studies used public transport data from automated fare collection (AFC) sources to describe public transport passenger trips. Pelletier et al. (2011) provided an extensive review of smart card automated fare collection implementations in public transport.

The systems are often smart cards which are swiped at a card reader at the boarding of each new public transport vehicle (New York City MetroCard, see Barry et al., 2002, Slavin et al., 2009 and the Chicago Card/Chicago Card Plus, see Zhao, 2004, Wilson et al., 2009) or at both boarding and alighting (London Oyster Card, alighting at train stations only, see Seaborn et al., 2009 and the Beijing Card, see Sun and Xu, 2012).

A large number of papers addressed the issues of reproducing the passengers' routes in public transport networks and presented solutions for algorithms to identify the boarding stations if not registered by the system, see Barry et al. (2002), Zhao (2004), Utsunomiya et al. (2006), Trépanier et al. (2007), Slavin et al. (2009), Barry et al. (2009) and Wilson et al. (2009). The destination stations were not always possible to identify and in all papers simplifications of trip patterns were made in order to reconstruct the routes. Automated fare collection only collects data on the actual use of public transport modes and not on the transport modes used for access and egress legs to and from the public transport network, and are therefore not in its present form suitable for analysis of multimodal route choice. Section 3.1.4 addresses the literature on the data collection and processing of data from smart cards from the map-matching perspective.

3.1.3 Questionnaire data collection

Since GPS devices, mobile phone, smart cards, etc., could not be used to collect the desired data for this PhD study (no transport mode or trip purpose information, signals fall out, etc. as discussed above) questionnaire methods were considered.

Ramming (2002) collected route choices for car drivers by asking for the origin and destination zones of their routes and the greater road segments used. This procedure returned a high number of incomplete route descriptions, some of which were fixed by using the shortest path between two known points or by using the routes of other respondents travelling between the same points. The method requires a great amount of manual work to map the data afterwards and this is not favourable if the data amount is huge and the collection is ongoing. Also in a public transport network the choice of transport mode between two points are not always as straightforward as in a car road network since not only distance is taken into consideration.

Prato (2005) collected data on route choice in a web-based survey where respondents indicate their chosen and other routes considered by selecting in a numbered order the junctions passed through on an interactive map of the city centre of Turin, Italy. The observations are for car drivers and the method is not applicable to the public transport system since the lines used in public transport cannot be identified by choosing junctions in a transport network. Vrtic et al. (2006) asked respondents to provide information on the origin and destination cities of their trip and up to three cities or locations they passed through on the way. This result in a great amount of missing information and the exact actual route cannot be reproduced from these pieces of information. Even though the methods described are not applicable to this study the method of asking for specific points of the travel is applicable to public transport if the questionnaire is created to obtain all relevant information so that the exact route can be reproduced.

Surveys conducted via mail, telephone, web-based etc. are all conventional ways of collecting data on route choice. Often the data collected are concerning the attributes of the traveller and the trip since the actual route is rather difficult to obtain in this way. Mahmassani et al. (1993) and Abdel-Aty et al. (1995a) described different approaches to the data collection for road users by means of questionnaires. Mahmassani et al. (1993) collected data on respondent characteristics and commuting patterns by using a short paper questionnaire sent to 13,000 households (less than 3,000 of the answers returned were accepted). Afterwards the respondents willing to follow up answered a more detailed questionnaire about their commuting trips describing the chosen routes link-by-link. Abdel-Aty et al. (1995b) combined a computer-aided telephone interview (CATI) with GIS means to register the exact road segments the car driver had used.

For public transport route choice, very few studies have collected questionnaire data to describe the route choice of the public transport passengers. Hoogendoorn-Lanser (2005) collected data on considered route choice sets and actual route choices via face-to-face interviews. The survey was carried out for train users in a specific train corridor in the Netherlands and defined as a Hub-and-spoke network. The collected route choice data was not far as detailed as required for this survey with regard to feeder modes, exact bus lines, etc., and also the sample was not representative because only train trips were investigated.

In 1997-1998 NYMTC (2000) conducted the Regional Travel – Household Interview Survey which included a random sample of approximately 11,000 households in the New York-New Jersey-Connecticut metropolitan area. The travel diary data for a 24-hour weekday was collected via the telephone and included the location of origin and destination for each trip during the day, the departure and arrival time, the transport modes used in the correct order, the activity at origin and destination. When the respondent travelled by a public transport mode, the line name or number was reported. If the respondent travelled by train or metro also the station name was reported. This data was however merely used for statistical reports on the data and not for reconstruction of the routes. The location of origin, destination and train/metro stations were map-matched but the public transport lines were not identified. The geo-coded locations were used for calculation of trip distances in a shortest path calculation for both car and public transport users.

Clifton and Muhs (2012) reviewed the various approaches to collect data on multimodal trips in a travel survey and they stated that historically travel surveys had collected data on the main mode of transportation. Many travel surveys do not collect (or do not store) information on access and egress trip legs neither for public or private transport. Clifton and Muhs (2012) suggested eight recommendations for including the whole multimodal trip in travel surveys, among these: better instructions of respondents of when to define walking as a trip leg, define a minimum walking distance threshold, and implement the use of GPS devices and similar techniques to improve the processing procedures to use the data.

3.1.4 Matching trip observations to a GIS network

In order to use the collected route choice data for comparison with generated route choice sets (see Chapter 5), the route observations should be reproduced in a transport network equal to the network in which the route choice sets are generated. This sets some requirements to the collected data and to the applicability to the transport network in which the route choice sets are generated.

The literature shows scarce effort in matching observed route choice data from a questionnaire to a GIS network and relevant research literature for methods of matching collected public transport route choice data to a GIS network has been difficult to find.

If the data are collected by use of the GIS network, for example by respondents pointing to a map (e.g., Prato, 2005), the data are more or less straightforward to use, but because of computer power the method is only applicable to smaller networks. Pointing to a map might be easier for the internet respondents but will probably not help the interviewer in a telephone interview.

When collecting route choice data described in words, some standard procedures to match the data to the network have to be developed. Ramming (2002) collected route choice for car drivers and use information about origin, destination and greater road segments used to reproduce the route, but an automated matching procedure could not be created because of the many incomplete route choices.

3.1.4.1 Automated Fare Collection Data

Several studies on matching public transport data from automated fare collection (AFC) sources to a GIS network have been carried out. These have various purposes of estimating station-to-station origin-destination trip tables (Barry et al., 2002), route choice estimation (Zhao, 2004 and Wilson et al., 2009) and statistical analyses (Trépanier et al., 2007 and Slavin et al., 2009). Some of the challenges of matching the AFC data to a network are the same challenges met when matching the questionnaire data to a network and some of the assumptions made are presented in this section.

Barry et al. (2002), Slavin et al. (2009), Barry et al. (2009) developed methods to match the automated fare collection (AFC) from the New York City MetroCard data to GIS networks. The New York City Metro card is an entrance-only system which registers information on each boarding of a public transport vehicle (bus, metro, train). For each boarding the public transport line ID and the time are registered and for rail modes also the boarding station. The data is truncated to six-minute intervals to save data storage.

Barry et al. (2002) processed information about stations used by metro travellers and matched these to a database of the metro stations. The systems collected entrance data only making the first stop/station and transfer (boarding) stop/station easy to identify. The destination stations were identified using a set of simplifications; the destination station on the first trip was assumed to be the equal to the first boarding station on the second trip and the destination station on the last trip of the day was assumed to be equal to the first boarding station of the first trip of the day. Barry et al. (2002) looked at trips with one or two trip legs only.

Zhao (2004) and Wilson et al. (2009) used data from the entry-only automated fare collection system from Chicago operated by the Chicago Transit Authorities and included buses in the studies, looking at train-train and train-bus transfers only. The boarding bus stop for each bus trip was identified by comparing AFC data to automated vehicle location (AVL) data for buses and GIS network attributes. Information was given about the exact time of boarding the bus and the GIS network was searched at a given network distance from the boarding stop to identify possible alighting stations. The studies successfully identified destination stops for 65.5% (in Zhao (2004) improved to 71.2% in Wilson et al., 2009) of the trips and with the knowledge of the public transport line used the trips were matched to a GIS network. Finally the matched routes and one alternative route generated by TransCAD were used to estimate route choice parameters.

Slavin et al. (2009) enhanced the work of Barry et al. (2002) and included bus trip legs from the New York City MetroCard data in the study examining the full set of alternatives (also bus-bus and bus-train). For bus-bus transfers an intersection table was created identifying the nearest stops on bus routes intersecting. Similarly, for rail and bus the transfer location of the lines was identified when the lines intersect at a single location only. If the first trip started with a bus trip leg the boarding bus stop was located at the bus stop on the bus line closest to the home location (if known). Finally major simplifications were made if the stops were not yet identified; if the bus boarding stop was not identified a stop was randomly assigned to a stop and if the trip destination was not identified a station was uniformly sampled from all trips starting at the same origin. The authors reported to have identified origin and destination stops for most AFC transactions. Barry et al. (2009) worked with the same data and same methods as Slavin et al. (2009) but instead of sampling destination stops for those not assigned a destination stop using the algorithms Barry et al. (2009) discarded the not-assigned data which were 10% of the trips.

Trépanier et al. (2007) processed smart card data from the city of Outaouais, Canada. The city has buses only (regular, express and special buses) and the smart cards give the travellers access to use either a selection of or all buses (regular cards are not accepted at express routes etc.). Using GPS data from devices on board the buses the boarding stop is identified and stored when the traveller boards the vehicle. The authors assumed that a traveller alight at the bus stop closest to the boarding stop of the next trip. Trépanier et al. (2007) made use of the continuous data and included the possibility that the last alighting stop of the day could be identical to the first stop of the following day. If the destination stop of a trip could not match the origin stop of the next trip (if the destination was not reachable by the train or bus line boarded) data from previous days were searched to find a similar boarding stop.

Utsunomiya et al. (2006) matched information collected via the automated fare collection system Chicago Card to a TransCAD GIS network to calculate the access distance as a network distance. For the first trip of the day (after 3 p.m.) the access distance was calculated as the distance between the billing address (known for 91% of the cards) to the first boarding locations at the first line used. The Chicago Card system saves information on public transport lines and rail stations used but not boarding bus stop and the authors assumed that the traveller would use the bus stop at the selected line closest to the billing address. This resulted in very high access distances for some travellers (exceeded 1.6 km for 34.4% and exceeded 3.2 km for 25.5% of the bus trips) because the billing address perhaps was not the same as the home address, incorrect AFC data, the traveller started from another place than home, etc. Utsunomiya et al. (2006) sorted the data only to use trips with a minimum access distance of 1.6km (the assumed maximum walking distance).

Even though the above methods of matching of automated fare collection separate themselves from the questionnaire route choice data collected in this study the methods can be used as inspiration for the methodology developed. The issues of defining the boarding and alighting bus stops are especially interesting for this PhD study. The AFC data collection has advantages since it is easy to collect and involves large samples of travellers. It has however disadvantages especially in the missing information about the origin and destination locations and the private transport mode trip legs making a reconstruction of full multimodal paths impossible using only this data. In the literature presented many assumptions are made to identify the routes and only trips following these trips patterns are correctly matched to the network. In a detailed public transport network offering many route alternatives the travellers do not always follow a symmetrical pattern when choosing routes for the trips during the day.

The goal of this study is to examine the route choices of passengers in public transport in the Greater Copenhagen Area. To be able to compare the routes with generated choice sets, the information of the routes should be detailed enough to enable the analyst to reproduce the actual chosen route. In order to be able to describe the route choices in the detailed network of the Greater Copenhagen Area the data collection should involve a large number of observations. Face-to-face interviews as in Hoogendoorn-Lanser (2005) would be very costly. The following manipulation of the data in order to reproduce routes should also be as small as possible to minimize costs and errors.

The following lists the important points to take into consideration when developing the collection method:

- Route choices in public transport for all trips during a day.
- Information on all (private and public) transport modes used.
- High level of detail.
- Possibility to reproduce the route in a GIS network.
- Large number of respondents.
- Cover the area of Greater Copenhagen Area and the public transport modes within this area.
- Limited budget constraints.
- Method for continuous data collection.

In this PhD study a questionnaire form was chosen in order to fulfil the objectives of the survey outcome. The questionnaire can be filled out by the respondent via the internet or by the interviewer via a telephone interview. The questions concerning the route choice of passengers were added to the already existing and ongoing TU survey. In the survey, travel diaries for around 26,000 people are collected each year. When adding public transport route choice questions to the existing survey, it has to be clarified how the new questions affect the existing part and to be assured that the time consumption and the difficulty level of filling out the questionnaire does not increase drastically causing a lower completion rate than before.

This chapter presents the TU survey and the new route choice questions and the data set along with the network used in section 3.2 and section 3.3, the design and the results of the full scale test carried out at DTU is presented in section 3.4, and section 0 investigates the implementation in the national TU survey. Section 3.6 explains the methods for matching the observed data to a GIS network, section 3.7 describes analyses of access and egress leg travel speed, section 3.8 shows examples of the final map-matching of route observations, and finally does section 3.9 conclude with the findings of this chapter.

3.2 Public route choice data collection

In the following section, the methods used to create the survey are described. The beforehand requirements are listed and explained and the data collection method chosen in the PhD study is described.

3.2.1 Requirements of the collected data

Prior to the creation of the questionnaire, we set up requirements for the collected data. The data should be collected from a large number of people with a rather simple method to keep down costs and a not too time-consuming method to minimize the dropout rate to get the most representative sample. Even though simple, the collected data should represent the actual chosen route and be reproducible in a GIS network, preferably with a minimum investment of analyst time.

For the current study, it was desired to develop a method to collect the route choice data for a large amount of people. For a small number of collected data, manual work to some level can be accepted, but when collecting thousands of observations the manual work would be very costly and should preferably be kept to a minimum. This means that the respondent should enter as much of the desired data as possible in a form that can be used either without manipulation or with programmed manipulation. If the respondent can enter the data into a GIS network similar to the network of the analyst, the matching part of the quality check can be skipped. For this PhD study this option was not an actual option since data are collected from a large area and the network that the respondents had to work with would be very huge to work. In order to obtain sufficient data we need information of all network attributes (stations, stop, transfers, addresses, etc.) and public transport attributes (lines, stop pattern, schedules, runs, etc.) and at the desired detail level the respondent would have to work with a database similar to the database described in 3.3.4. However, the complexity of this database search would add very much to the respondent burden and might cause respondents to drop out from the survey. Also the access to the database would set high demands for the respondent’s computer power and internet access.

The collection method has to be rather simple to keep costs down and obtain a high completion rate. When interviews are done face-to-face, the questions are allowed to be more complicated if the interviewer is well prepared and can help with the understanding of the questions and how to fill in the questionnaire. When interviewing via telephone, the interviewer can also assist in explaining the questions, but detailed descriptions as figures and graphs cannot be used directly. When the respondents fill in the questions in a printed version or on the internet, the understanding of the questions is completely up to the respondent himself. The questions therefore have to be easily understandable to obtain accurate answers and to avoid losing respondents before completing the survey. In this way it is assured that not a specific part of the population drop out and cause a non-representative sample. On the internet it is possible to create verification procedures for the entered data to point out possible errors to the respondents. This is partly similar to the function the interviewer can take on noticing obvious errors in the answers of the respondents.

A questionnaire survey can be created in a form providing the possibility of collecting route choice for passengers in public transport. The questionnaire form is rather open and offers the possibility of asking for the exact information desired. The formulation of the questions has to be very clear in order to collect the data in a way making it possible to reproduce the exact actual route of the traveller. This problem is somewhat different between car and public transport and even though travelling through a public transport system seems more complex than through a road system, some restraints on the public network make it easier to collect information in a way that the exact route can be restored. For a car driver it is possible to select any given road between two points, but when a public transport user indicates use of a specific bus line between two given points, the route can easily be reproduced because of the knowledge of the route of the bus line. For public transport, for each trip the passenger can travel with many different transport modes also including private modes for access/egress given the locations of origin and destinations, mode availability, public transport service level, and many other factors.

3.3 Data

In the following, the TU survey and the GIS network to which the route choices are matched are described.

3.3.1 The TU Survey

The Danish national TU Survey is an on-going collection of travel diaries alongside respondents' and households' socioeconomic data, and consists of a questionnaire which is either filled out on the internet (20%) or via telephone (80%) (see Christensen, 2013). In the TU survey, respondents are a representative sample of the Danish population between 10 and 84 years who are asked to describe all their trips with both private and public transport modes on the day before the interview.

Respondents provide information on all their trips during the day (e.g., selected modes, time duration, distance travelled, and trip purpose) and all their socio-demographic characteristics (e.g., gender, age, income, location of residence and workspace). The data are a great source of information and makes it possible to reveal many interesting details about travellers' choices in transportation networks. For more information on TU refer to Jensen (2009) and Christiansen (2009).

The TU survey is unique in Denmark since it links information on actual travel behaviour to a list of background variables. From an international perspective it is also unique because of the level of detailed information about the trips (especially the routes for public transport) including the coordinates of the destinations for all trips, the amount and quality of data and the fact that the collection of data is continuous. These facts make the existing TU survey an obvious choice for the addition of the public transport route choice survey questions.

3.3.2 Public transport route choice questions

We formulated the route choice questions to add to the TU survey short and precise in order to keep the questionnaire simple, obtain high completion rate, and collect good and useful observations. It was important for the PhD study that the information was detailed enough to enable the reproduction of the route, but also simple enough for the respondent to provide it correctly. By answering questions about specific points on the trip the route can be reproduced by the analyst with knowledge of the public transport network. The development of the data collection method is described in Anderson (2010b) and the route choice question part of the TU survey is presented and explained in the following sub-sections.

3.3.2.1 Introduction

Before answering the questions concerning public transport route choice, an explanatory text to the following questions is provided to the respondent:

On the next pages you are asked questions about your transportation on [DATE]. Every time you travel on a street / road to get to a new activity / purpose, or a new place, we call it a trip. A trip can also be an end in itself, such as jogging or walking the dog.

Remember all errands en route. All errands, activities, and accommodations during the day must be included, also the short ones. It is important that you also enter if you have visited a kiosk, picked up someone, walked the dog, or performed other activities that led you to move from place to place.

Change of transport mode along the way is part of the same trip. If you use public transport, the same trip can contain many transport modes. Typically there are at least three: walk, bus or train, and walk again. Walk to / from the bus stop / train station is part of the overall trip and must be included along with other vehicles. We therefore ask where you went, and not where you boarded the bus. Similarly, if you parked your car and walked the remaining distance to reach the destination, then the trip has two modes: car and walking.

Trip out and to home are (at least) two trips. A trip can never have the same location as the start and end point. The trip must be divided at all purposes under way. If the trip is a goal in itself, it must be split up, so that the farthest point along the way is the destination of the trip outbound. Remember to finish with the return trip, which probably has the destination in your home.

3.3.2.2 Trip description

The respondent enters the start and the end points of the trip as addresses which are linked to coordinates, the purposes at the trip start and end points, and the departure time. The information has to be filled for each trip during the day, with the convention that the start point for the following trip is defined as the end point of the previous trip.

1. **Where did your day start?**
   - The location the respondent listed as home in an earlier question.
   - Another place in Denmark (type and select address from list).
   - Another place abroad.
     - The last two options return additional questions about the purpose of the trip start location.
2. **When did you leave [previous address entered]?**

3. **After you left […] what was the first place you went to?**
   - Same options as in 1.
4. **What was the purpose of your stay here?**
   - 24 choices among others work, school, shopping.

All transport modes used on the trip are asked for. The respondent chooses from a drop down list (21 transport modes among others car, bicycle, walking, bus, train). When choosing a mode, additional boxes that are required to be filled in appear according to the transport mode entered.

- Walking, Bicycle, Car, Airplane, etc.
  - Enter length and travel time used
- Bus
  - Enter waiting time, bus line, length and travel time used
- S-train
  - Enter waiting time, from-station, S-train line, to-station, length and travel time used
- IC-train, Regional train, Metro
  - Enter waiting time, from-station, to-station, length and travel time used

When the full list of transport modes used for the trip is entered, the respondent returns to the questions from above, starting with no. 2 and continuing until all the trips for the specific day are entered.

The list of transport modes can become relatively long, especially when using public transport, but the boxes to fill in are relatively easy to understand and along the way many checks of the entered values are offered to the respondent (orange boxes in Figure 3-1). The respondent selects the *from-* and *to-station* from a drop down list containing stations that are within a certain distance of the address entered and the distance travelled so far. Along the way, travelling and waiting times are added to the start time for the respondent to check the arrival time. When the *from-* and *to-stations* are entered, the correct length via rails between the two train stations are automatically calculated and suggested to the respondent. This is possible for most pairs of train stations because most train stations are only served by one train type and for the train stations served by more than one the name of the station refers to the type as well (e.g., Nørreport (metro)/Nørreport (Reg.) etc.).

Figure 3-1 shows an example of a public transport route description.

Transport modes in correct order:

	Transport mode		Line	Length		Time		
1.	Walking			0.5	km	8	min	
				Waiting time:		3	min	
2.	Bus		701	2	km	5	min	
3.	Walking			0.2	km	2	min	
	From-station:	Nykøbing F		Waiting time:		7	min	We calculated the distance to 146 km from the stated stations
4.	Other train			146	km	100	min	
	Transfer at:	København H		Waiting time:		8	min	
5.	S-train		E	13.9	km	20	min	We calculated the distance to 13.9 km from the stated stations
	To-station:	Kgs. Lyngby		Waiting time:		5	min	
6.	Bus		300s	5	km	10	min	
7.	Walking			0.5	km	5	min	
			Sum:	168.1	km	173	min	

Figure 3-1: Example of a route description for a public transport trip.

The respondent has travelled with several different public transport modes and has walked to the station from the origin and to the destination. The light blue boxes are to be filled in by the respondent. This is an example of a complicated trip with many transfers, but still the questionnaire is relatively easy to fill in.

In the figure the traveller stated to have seven legs on his trip. Three legs are walking and four are public transport. In theory a walking leg exits between all modes but often the traveller doesn't perceive a transfer using the same platform or different train/bus stops at a station as a transfer involving a walking leg. The walking time is therefore included in the stated waiting time.

When using the list of public transport lines, train stations, etc. the route can be reproduced. Information on the start point of the trip can be used to find the bus stop by searching within a certain buffer for bus stops where the mentioned bus line stops. When the bus alights at a train station this point is fixed and the route to the next train stations is fixed as well. When selecting certain modes (e.g., *other train*) between two train stations, the route is almost definite since not many alternative routes exist between two train stations with the selected train mode. At the end of the trip, the respondent uses bus again and the alighting stop of the mentioned bus line can be found by searching from the destination point.

The procedure of searching for the bus stop used could be avoided if the respondent was asked to provide the name of the bus stop. This option has not been implemented since it would be very difficult for the respondent to fill in correctly. Many people do not know the exact name of the bus stops (often complicated and different from what they are called in daily speaking) and therefore free hand writing would involve a great deal of manual analyst work afterwards. When answering the survey via the Internet a list of the bus stops on a specific line could be presented to the respondent but this list could be very long, the creation of the list demands entering of the exact correct bus line name and still the respondent is perhaps not able to recognize the name of the bus stop used. Another possibility (for the internet version) could be the respondent pointing to a map but this would very much add to the complexity of the questionnaire.

The procedure of matching the data to a route in a GIS network is described in section 3.6 and Anderson and Rasmussen (2010) explain it in details.

3.3.3 Data set

The information collected in the TU survey is in six head subjects: household, respondent, car, journey, trip, and leg characteristics (for documentation of the TU survey see Christiansen and Haunstrup, 2011). For each of these subjects a table is represented in the survey database. Below the tables relevant are described.

- **Respondent.** The respondent table includes information on the socioeconomic data of the respondent, the residence, the household and the family. The respondent lists his/her gender, age, occupation, education level, home address, workplace (address, working hours, public or private employment), ownership of a bicycle, public transport season ticket, driver's license, car availability, car ownership, handicaps, income (also for spouse, family, household). The exact locations of the home and the workplace with coordinates are registered.
- **Trip.** The trip table contains information for every trip during the day namely departure time, trip purpose, origin, destination, primary mode, travel companions, the fare payment amount and method of public transport.
- **Leg.** Alongside the trip table, the leg table includes details about the main components of the trip. Each trip is divided into legs for each transport mode. Information is about transport mode, being a driver or a passenger, respondent's conception of distance, time, and waiting time. For trip legs using public transport modes, information on bus lines, as well as on access to and egress from the public transport stops and stations, is also listed, thus enabling the reconstruction of the chosen route.
- **Journey.** The journey table comprises aggregated data from the trip table. A journey is also defined as a trip chain and both starts and ends at the home or the location of origin for the day. The primary purpose of the journey is the purpose of the destination with the longest stay.

From February 2009 to May 2010, approximately 6,300 interviews were collected in the Greater Copenhagen Area with more than 22,500 trips of which 2,200 use public transport for a part of the trip (minimum one trip leg). The number of trip legs in this data set with bus, S-train, metro, and other train (regional and IC-train) in the Greater Copenhagen Area is shown in Table 3-1. The table also illustrates how many respondents have entered information about the line use (99% of bus passengers, all S-train users) and which train station was travelled via (8% of bus users, minimum 97% of the train users). The low share of station information for bus users is due to the fact that the respondents are not asked about the boarding and alighting stations, but when they travel to and from a train station this information is added to the data set.

Table 3-1: Total number of trip legs using public transport modes, number of trip legs with stated line name/number and from- and to-station names for travellers in the Greater Copenhagen Area.

Mode	No. obs. trip legs	Line Name /Number	From station name	To station name
Bus	1,584	1,581 (99%)	214 (8%)	208 (8%)
S-train	1,039	1,039 (100%)	1,039 (100%)	1,022 (98%)
Metro	459	-	459 (97%)	444 (98%)
Regional + IC-train	259	-	258 (100%)	254 (97%)

3.3.4 Networks

3.3.4.1 Physical network

The public transport network of the Greater Copenhagen Area used for analyses in this thesis. Around 2 million people live in this area, which is the most densely populated area in Denmark. The public transport services include metro, buses, and trains (regional, IC-, S-train and local train). The public transport network was presented and maps were shown in Chapter 2 (see Figure 2-2, 2-4 etc.).

3.3.4.2 GIS network

The public transport network is built using information from the Danish Rejseplan.dk (www.journeyplanner.dk 2013-09-10) which is a data source containing information on lines, stops, schedules, etc. for the Danish public transport network.

Rasmussen (2010) and Anderson and Rasmussen (2010) matched the collected route choice observations to a network representing the Greater Copenhagen Area in a schedule-based public transport network containing addresses, train stations, bus stops, transfers, train lines, bus lines and a road network (see Anderson and Rasmussen, 2010 and section 3.6). The network is also used by the schedule-based stochastic transit assignment model based on MSA used for the generation of choice sets (Nielsen, 2000) carried out in Chapter 5, and is analogous to the one used in Orestaden Transport Model (OTM) (e.g., Jovicic and Hansen, 2003).

The transport network covers the Greater Copenhagen Area and consists of a road and path network used for walking, bicycles, cars, buses, and a rail network for trains and metro trains. Network elements include:

- **Zones:** in this survey exact start and end points are used to investigate the exact route.
- **Connectors:** connect the exact start and end points to the road /public transport network.
- **Road/path network:** links and nodes, used for walking, bicycles, cars, buses, etc.
- **Rail network:** rail for regional trains IC-trains, S-trains, local trains and metro trains.
- **Stops:** bus stops and train stations, where passengers board and alight buses and trains. The stops are defined in stop groups with one or more stops, so that two stops on the opposite side of a road are in the same stop group.

- **Changes:** transfer links connecting bus stops and train stations.
- **Lines:** definition of bus and train lines.
- **Line Variants:** different types of each bus/train line.
- **Line Variant Elements:** each *line variant* is divided in a number of elements, *SQIdx*, for each segment between two stops, with *SQIdx* as a rising number in the driving direction. In each direction the *line variant* has a new *line variant element*.
- **Runs:** different variants of the *line variant*'s stop pattern (the stops served by the line and the order of the stops).
- **Schedule:** links *runs* to *line variants.*
- **Schedule Elements:** information on stops served by the run, whether runs allow for passengers to board and/or alight at specific stops, and arrival and departure time.

Figure 3-2 shows the structure of the public network with the content of the tables and the connections between the tables.

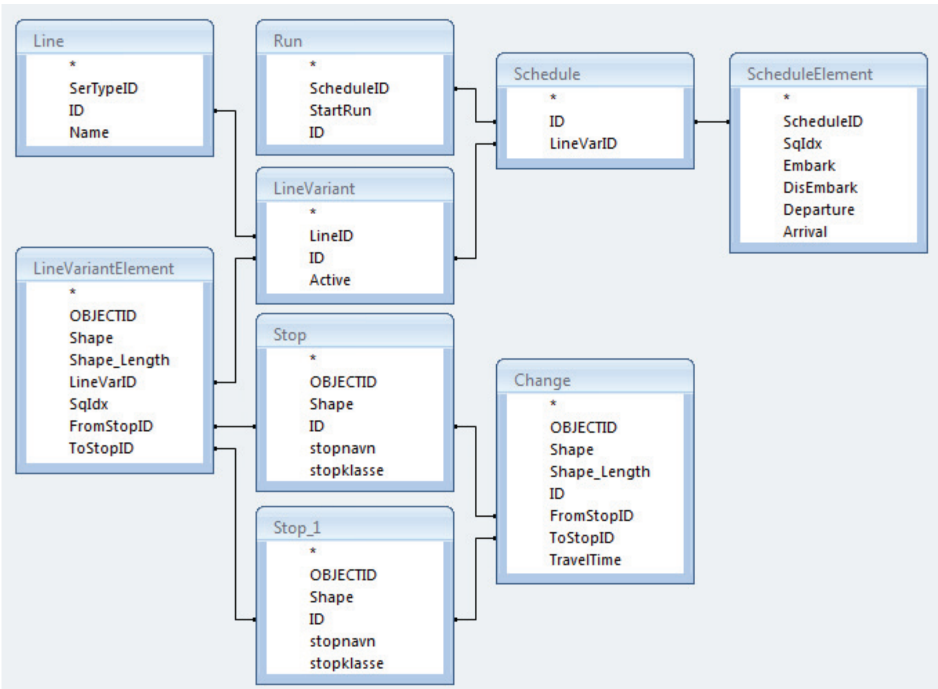

Figure 3-2: Database diagram of the public network structure.

3.4 Full Scale Test at DTU

3.4.1 Adjustment of the TU Survey

In order to describe the route choice of the respondent it is important to be able to reproduce the actual chosen route. In the TU Survey version existing when the PhD study started data concerning transport mode, time and length were found but this was not sufficient to reproduce the route. Therefore the description of trip legs was extended with several parts to include information about public transport route choice.

In the TU Survey every trip leg is described by transport mode, time (minutes) and length (kilometres). If the respondent used public transport the waiting time before boarding the vehicle is entered. In order to avoid greater changes of the questionnaire coding a method to describe the routes in public transport in the present structure of the questionnaire was sought.

To the questions regarding trip legs in public transport were added extra fields in connection with trip legs using public transport. A respondent who travelled by bus had to fill in a field with the line number of the bus line. This field appeared when the bus was chosen as a transport mode. For trips with train legs from- and to-station were asked for (where the train trip started and where it ended). For each alighting or boarding it was to be informed at which station it had taken place. The choice of station was made from an automatic generated dropdown-list of possible train stations.

Routines to pick out possible stations were created so the list of train stations would not be immense to the respondent. In this way only the relevant stations were shown in the list. The respondent had provided information on residence, workplace, etc. in advance of filling out details about the trip and this information was matched to a database with coordinates of most addresses in Denmark. The start of every trip should be entered and this piece of information was used to form the dropdown-list with train stations. If the respondent stated to have started the trip from the residence and to have run by bicycle a certain number of kilometres to a train station the possible train stations could be calculated from the coordinates of the residence and the length of the bicycle trip (with a certain error margin). This was done automatically by an underlying database with information about the coordinates of the train stations. A list of possible to-stations was created by the use of data to identify:

- Which stations were served by the same train lines as served the from-station.
- How far the stations were from the given from-station compared to the number of kilometres and minutes the respondent stated to have used for the trip leg (also with an error margin).

From this information it is possible to reproduce the public transport route of the respondent.

3.4.1.1 The design of the survey

Some of the work of this PhD was to test the route choice questions as a part of the TU Survey in a large scale test before using the survey national-wide. The relative huge number of students and staff at the Technical University of Denmark (DTU) was used for the purpose. It was chosen to carry out the full scale test as an internet based questionnaire only which minimised the costs significantly.

3.4.1.2 The DTU Survey

In order to test whether it was possible to collect detailed and correct information of the route choice for multimodal public transport, the questionnaire was tested in a test survey at the Technical University of Denmark during a week in May 2008. The test survey was carried out amongst employees and students as an Internet based questionnaire. DTU had 4,100 employees and 6,200 students in the fall of 2007 (DTU, 2007), giving a good potentially large data sample for the survey. The quality of the data collected (actual route) and the impact of the new questions on the existing survey (for example change in number of drop-outs, time-use for the survey) were assessed to make a recommendations for the use of the new public route choice questionnaire. See Larsen (2008) for more details.

When receiving an invitation for the survey either of two outcomes could be expected: a certain amount of goodwill to help with the survey was expected especially from other scientists at the university, or a possibility that the respondents would feel that the time consumption was too high and therefore would choose not to answer. Finally there is always a large percentage that does not react to requests regarding such studies (see for example Frick et al., 2001).

The employees were contacted via email and approximately 65% of the staff received an email. No mailing list or the like of student email addresses could be obtained and therefore the students could only enter the survey via a link on the Intranet (where also employees not contacted were informed about the survey). The message on the Intranet was visible to all students and employees, but such a link is easier to ignore or overlook than an email sent directly to the respondent. Furthermore, it was decided to offer a prize to the participants, since it was considered that in particular the students would find greater incentive to participate at the prospect of winning a competition (e.g., Porter and Whitcomb, 2003).

3.4.1.3 Questionnaire – Danish version on the Internet

As mentioned, the original TU Survey is designed to allow the respondent to choose between being interviewed by telephone or via the Internet. In this test study there were no resources to conduct telephone interviews, and therefore the questionnaire was only available via the Internet. From 2006 to 2009, 20% chose to complete the TU survey via the Internet. If these rates are directly transferable, 80% could be expected not to answer the questionnaire.

However, the fact that telephone interviews were not an option is expected to give a significantly smaller drop-out rate among DTU staff and students than in the general population. The vast majority of the DTU staff and students use the Internet in their everyday life. In addition, in this study there were no other ways to participate in the survey. Personal contact is known to increase the response rate but it is assumed that among the group of students and faculty the absent of personal contact is not as problematic as it might have been amongst other groups of respondents. It is not known what proportion of the participants in the TU Survey who would use the Internet version if it was the only option.

In the national survey the respondent is contacted if he has not answered the questionnaire within a certain amount of days. This is not the case for the test survey and this can lead to a lower response rate. The web response rate of 20% in the TU survey can indicate that 20% choose to answer the questionnaire via internet and 80% choose the telephone interview. But among the 80% might be persons missing or postponing the response and do not answer before they are contacted by telephone. If the person has not answered within the two day deadline and therefore gets called by telephone, the telephone is the only option for answering in the national TU. See Christiansen and Haunstrup (2011) for more details on the TU survey.

The questionnaire available to the DTU respondents was only in Danish, since the existing TU survey was only in Danish. It is estimated that the non-Danish speaking share of the employees (about 25% in 2007) and students (about 10% in 2007) of DTU are sufficiently small to make it possible to get good response rates in spite of the lack of an English version of the questionnaire. For the route choice, which the study will identify, it is also useful to exclude persons with a very poor knowledge of the public transport system in the Greater Copenhagen Area to a certain extent. This will often apply to foreign students with shorter stays in the country and foreign guest lecturers or other persons who are temporarily staying at the University. Moreover, foreigners have other travel patterns, preferences or similar. The inclusion of this is also to some extent prevented from the study, by not having the questionnaire in English.

3.4.2 Results from the full scale test

The survey was conducted at DTU from Monday 5th to Friday 9th of May 2008. This week was chosen because there were no holidays, and it was before the students started exams, which should allow for the largest possible number of answers. It appeared, however, that the given week was "Bicycle to work" campaign, which could cause less people travelling by public transport than usual as they took the bicycle to work instead. In the invitations the respondents were asked to choose a day, where they travelled by public transport if possible, but answer the questionnaire although they had only used private transport modes.

The results of the DTU test survey are presented and discussed below after a short presentation of the location of the DTU Campus.

3.4.2.1 DTU Campus

The DTU Campus is placed north of Copenhagen (marked with the black ring in Figure 3-3). The area stretches over 2 km from south to north and 1 km from east to west. The campus is located in the city of Kgs. Lyngby 2.5-4.5 km from the nearest S-train station (Lyngby st.). The DTU Campus is served primarily by buses with the S-buses from Copenhagen and the north of Zealand (150S) and from Lyngby st. (300S) as important public transport modes. Also, smaller bus lines serve DTU from the train station. The Fuglevad st. is a local train station 1.5-2.5 km from campus and no direct bus line serves both this local train station and DTU.

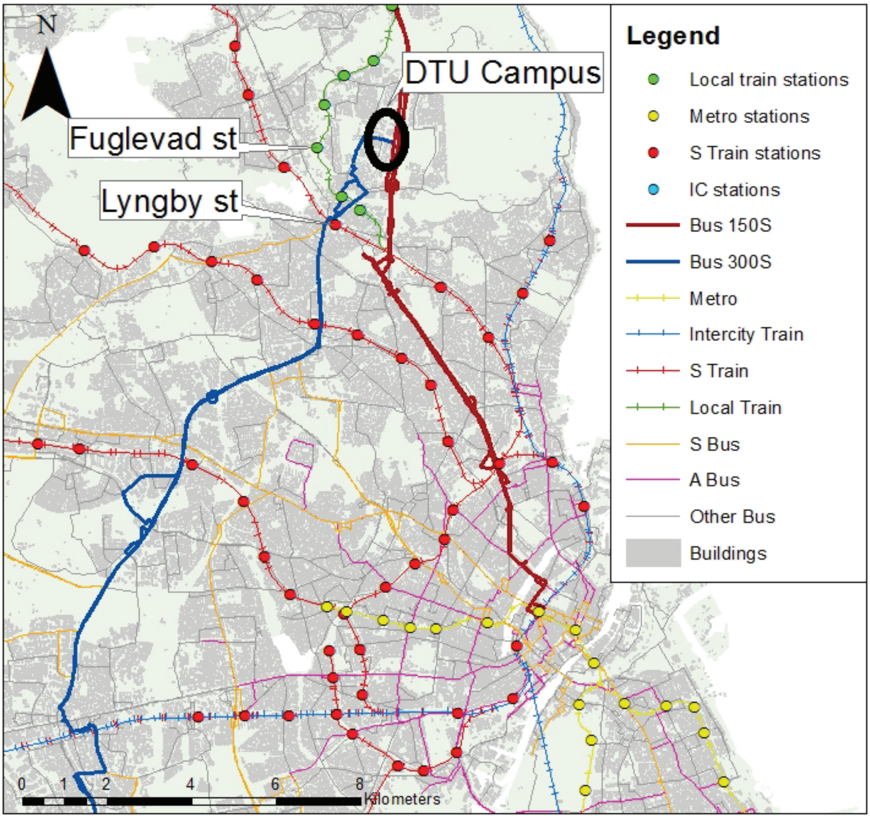

Figure 3-3: Location of DTU Campus (black circle).

3.4.2.2 Number of responses

The test survey at DTU collected 545 full responses which equals a response rate of 5%. Among students only 3% completed the survey, while among employees 8.5% finished.

In total, responses from 748 respondents were collected, who to some degree completed the questionnaire. However, not all answered all the questions. 573 persons entered one or more trip legs. Participants were mainly students and employees as shown in Table 3-2. 1,096 people started the questionnaire, but as mentioned, not all completed this.

Table 3-2: Respondents in each job category with at least one trip.

Position	No. observations	No. trip legs
Unemployed	1	4
Early retirement pensioners	1	7
Students	186	988
Draftees	1	8
Apprentices	3	21
Employed	381	1,797

186 students and 381 employees answered part of the questionnaire. In total 2,825 trip legs were registered in the collected data, meaning that most respondents had more than one trip leg (they also had more than one trip). Especially people using public transport modes had several trip legs because each trip can consist of numerous bus lines or a combination of walking, bus, train etc. (see Chapter 2 for a description of multimodal transport trips).

3.4.2.3 Grouping on main transport mode

The following figures show only the employment categories students and employees, since these are the categories with far most responses. Also, only transport modes used for more than ten trips are shown. Figure 3-4 illustrates which main transport modes are chosen distributed on the position of the respondents.

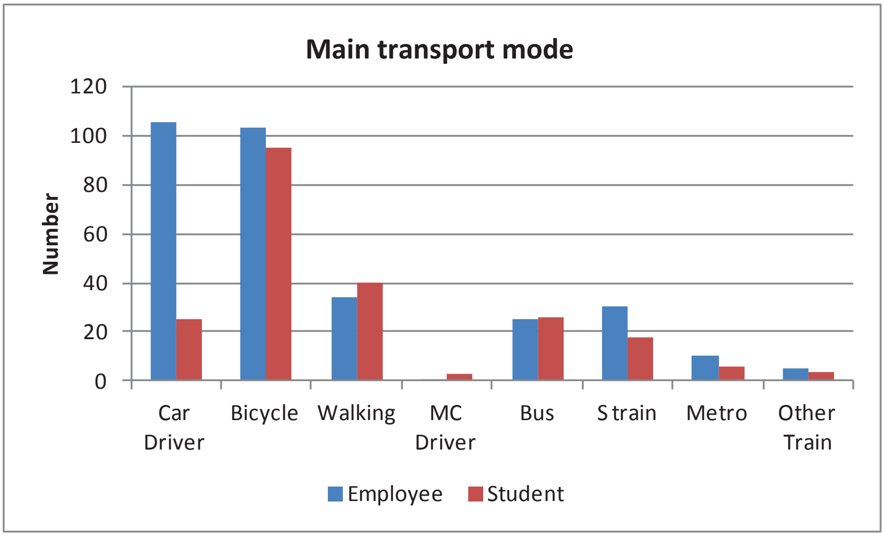

Figure 3-4: Choice of main transport mode divided according to the job of the respondent.

The main transport mode is the mode used for the longest trip leg measured in distance. Most employees are in the categories car driver, bicycling, walking, bus, S-train and other train (regional and IC-train). Considering the fact that twice as many employees compared to students have participated in the survey, a relatively high number of students choose cycling, walking, bus and other public transport modes. Half as many students as employees choose trains, and more employees choose the car.

Figure 3-5 shows the share each job category holds on the total number choosing a certain transport mode. Employees and students represent approx. 60% and 40% of those choosing bicycle, walk, bus and S-train as the main part of the trip. Students represent more than half the MC drivers and approx. 20% of the car drivers and taxi passengers.

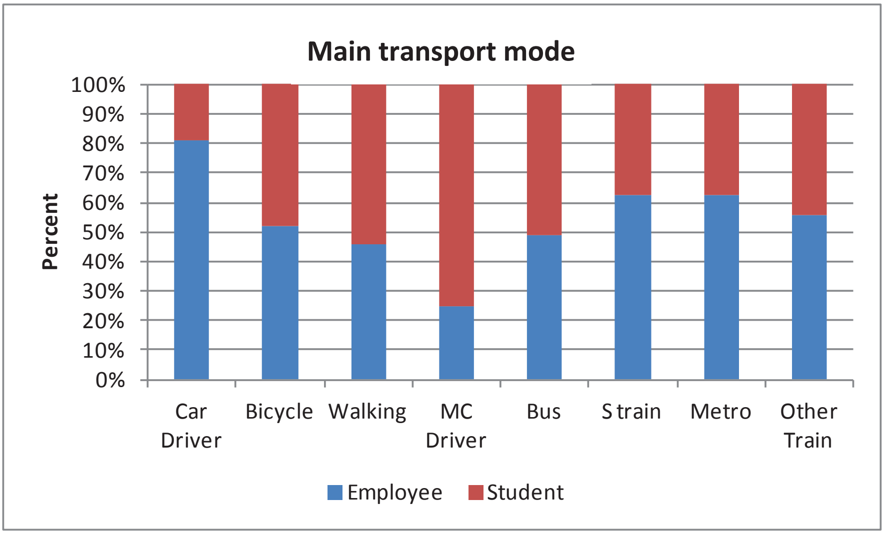

Figure 3-5: Share each job category represents of the total number who chooses the transport mode.

The majority of the respondents are students and employees and Figure 3-6 shows these with use of main transport mode in percentage.

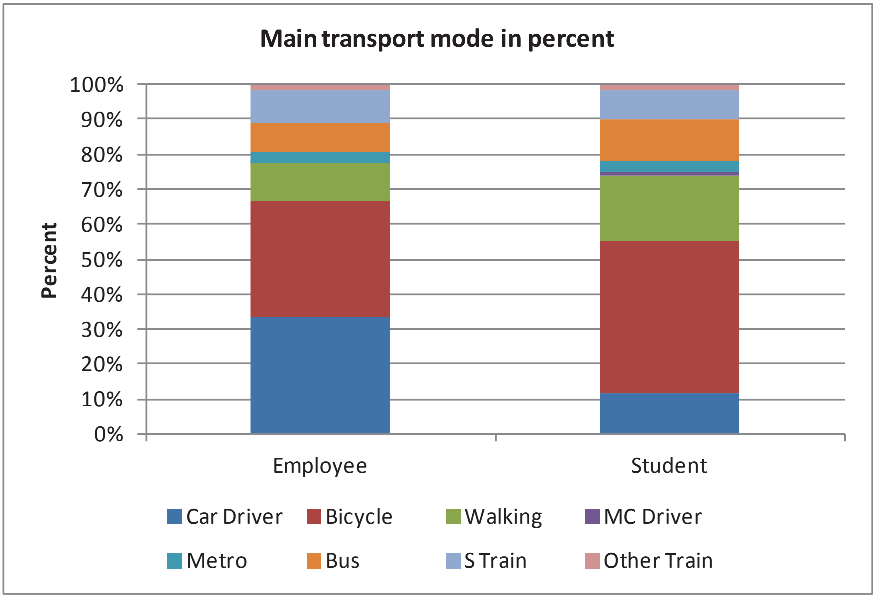

Figure 3-6: Share of students and employees who chooses a certain transport mode.

An approximately equal number of employees choose car, walking and cycling. Among students, a much smaller percentage chooses car, but a higher share is walking and bicycling as part of their trip.

Only a few percent in both categories choose train and metro. More students choose metro and more employees choose train which is describing where the respondents are living. A far larger share of the students chooses bus and slightly more chooses S-train. Approximately the same share chooses either train or S-train in the two job categories. Only among students are enough travellers choosing the motor bicycle to be shown here.

Few respondents in the test data set use train and metro because of the sampling method of the respondents. As explained no public transport on rails is reaching DTU and the nearest train station is 3 km away so students and employees heading for DTU do not use public transport on rails as often as an average traveller in the Greater Copenhagen Area.

3.4.2.4 Choice of public transport

The number of travellers using public transport on a part of the trip is examined. Table 3-3 shows that 170 respondents used public transport on at least one trip leg during the interview day. Of these are 101 employees and 69 students. This equals 37% of all students and 27% of all employees choosing public transport.

Table 3-3: Number and share of each job category choosing public transport on at least one of the trip legs.

Job	Respondents choosing public transport	Total number of respondents	Percentage
Students	69	186	37
Employees	101	381	27
Total	170	567	

An extraction of data from the original Travel Survey data has been created for comparison reason. The data extraction is for main transportation mode and only students and employees are included as in the test survey data set. This extraction is made in the Greater Copenhagen Area as respondents in this area should correspond roughly to the respondents in the test survey. Table 3-4 shows that more than half of the employees in TU are car drivers. 8% of the employees use a public transport mode (bus and S-train) as the main transportation mode. This is far from the 27% that was found in the test survey, due to the statement that in particular data for public travel was interesting for the test.

Table 3-4: Employees and students' choice of transport mode in number and percentage in the TU Survey.

Position	Transport mode	Number	Percentage
Employees	Walk	34,280	12
	Bicycle	40,647	15
	Car driver	151,803	55
	Van driver	3,211	1
	Car passenger	23,492	8
	Bus	10,067	4
	S-train	8,940	3
	Train	4,025	1
Students	Walk	5,986	18
	Bicycle	10,704	32
	Car driver	6,741	20
	MC driver	128	0
	Car passenger	3,075	9
	Taxi passenger	267	1
	Bus	3,088	9
	S-train	2,206	7
	Train	1,092	3

Among the students 32% use bicycle which is a good match with the test survey data. DTU is not placed in the CBD of Copenhagen and there might be that fewer people able to reach the campus by bicycle than are in average able to get between home and study location on a bicycle. But several collegiums are placed close to and the students living here either chose to ride bicycle or walk.

For employees significantly more are choosing bicycle in the test survey data than in TU. This is largely due to the fact that the week chosen for the survey was the first week of the "Bicycle to work" campaign, and that the weather was very good, which encourages more people to ride the bicycle.

Slightly fewer students choose the car in the test survey compared with the TU (13 to 20%). There are 19% public transport trips in the TU survey and as mentioned 37% in the test survey data, and that difference is caused by the design of the survey.

3.4.2.5 Investigation of trips to DTU Campus

Respondents in the test survey travelling to DTU started their trip in the points shown in Figure 3-7 (450 trips). In the figure the trips are split in respondents using public transport and respondents using one of the private transport modes bicycle and car on the longest part of the trip (measured in distance).

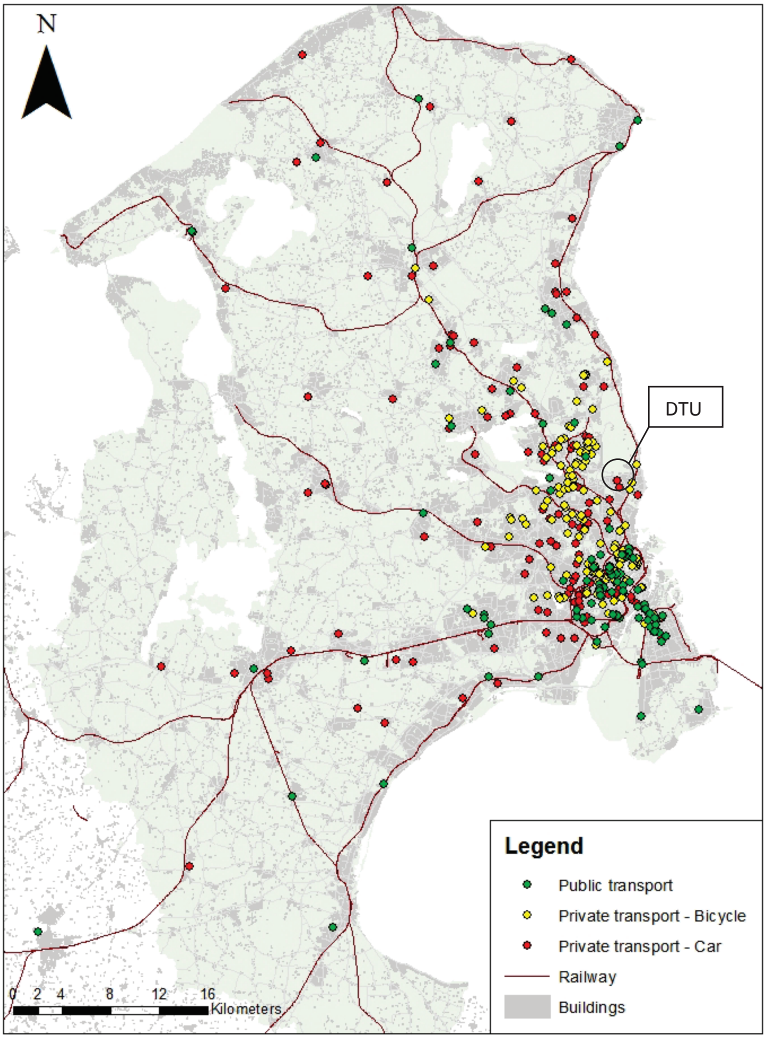

Figure 3-7: Starting points for trips to DTU divided in main transport modes private or public – the Greater Copenhagen Area.

112 persons have used a public transport mode on the longest part of the trip, and 348 have used a private (149 car drivers and passengers, 199 bicyclists). Persons choosing walking as the main transport mode are not shown on the map.

There is a 50/50 distribution of the choices between public and private transport modes in the outer areas of the map. Closer to DTU campus, there is a predominance of people who have used the bicycle due to the short distance. Also many travellers walk from this area.

From the CBD of Copenhagen a large number of travellers choose public transport. Few choose public transport in the area between the CBD of Copenhagen and DTU. This may be due to poor availability of public transport and poor accessibility but is mainly due to the fact that many travellers choose the bicycle. Most bicycle trips start within a radius of 10 km to DTU Campus. The majority of private transport travellers with more than 10 km to DTU choose the car.

The points of origin are distributed along the rail network and in particular the public transport trips are starting close to a railway. As presented in Chapter 2 urban development is attempted along the railways (radial lines leading from the centre of Copenhagen to the suburbs), and therefore most houses are located along the railway lines. In general, more people choose public transport if they live close to the railway and thereby have easier access to the train. Travellers living far from a railway station often used private transport in this survey data.

For analysis reasons the points are merged in geographical areas. The areas are created on the basis of how the points clutch together in areas where travellers can be thought to have the same travel patterns.

Figure 3-8 shows the areas from which trips by public transport are performed in the data set. These have been split into big areas around railways and smaller zones around the centre of Copenhagen. In particular, these trips include travels from Copenhagen and the area north of the DTU. From the grey zones no trips going to DTU using public transport are reported.

The figure shows the percentage choosing public transport in each of the zones. From the area in the south-west many travellers start their trip with public transport (>40%). Many of these trips start from points near the railway. Also many travellers from the island to the east, Amager (>50%) and the CBD of Copenhagen (>30%) choose public transport. The area nearest to DTU has the lowest share of public transport trips which is due to the fact that many travellers use the bicycle when living close to their destination. From the area near Elsinore, near Frederikssund and near Hillerød (the three large light green areas in the middle) only 10-20% choose public transport. This could be because of poor public transport ring connections which causes a long travel time when the traveller has to go through the CBD of Copenhagen. These areas are rather large and include large rural areas where the inhabitants may have a long distance to the public transport system. Many travellers living far from the railways choose private transportation.

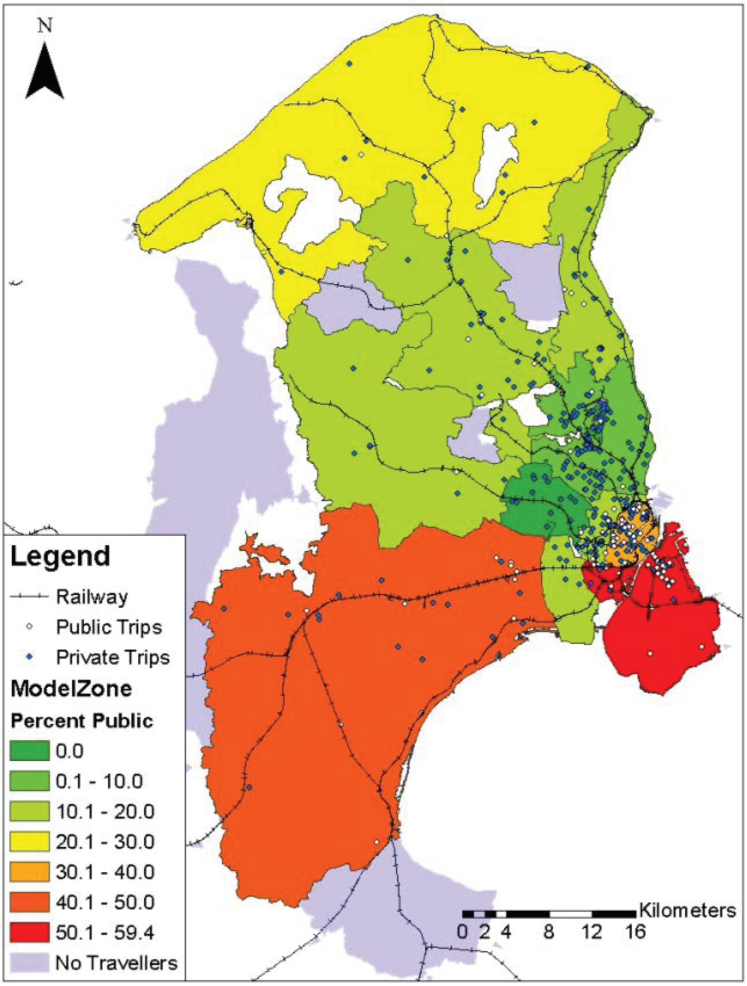

Figure 3-8: Percentages of all travellers travelling to DTU with public transport.

In Figure 3-9 six selected zones and the percentage of travellers from each zone who choose public transport is shown. From these zones at least 20% travels by public transport when going to DTU. Three of the zones near the Copenhagen city centre have more than 50% travelling by public transport.

From some of the zones, most respondents have chosen the same public transport route through the network. This is the case for areas close to bus 150S, where the majority is using this bus. 150S is an S-bus[1], running with 5-10 minutes intervals and only stopping at major stops. The bus runs between Nørreport station and DTU (and to Kokkedal / Nærum in the north) using the

[1] Fast bus connecting the S-train stations and other public transport hubs (see Chapter 2 for more explanations)

motorway most of the way and is used by many students and employees from Copenhagen, since there is no public transport by rail to the DTU.

From various places on Amager, however, very different routes are registered. These respondents are often passing through Nørreport (via metro), and from here they choose between the S-train to Lyngby St. and bus to DTU or bus 150S directly to the DTU.

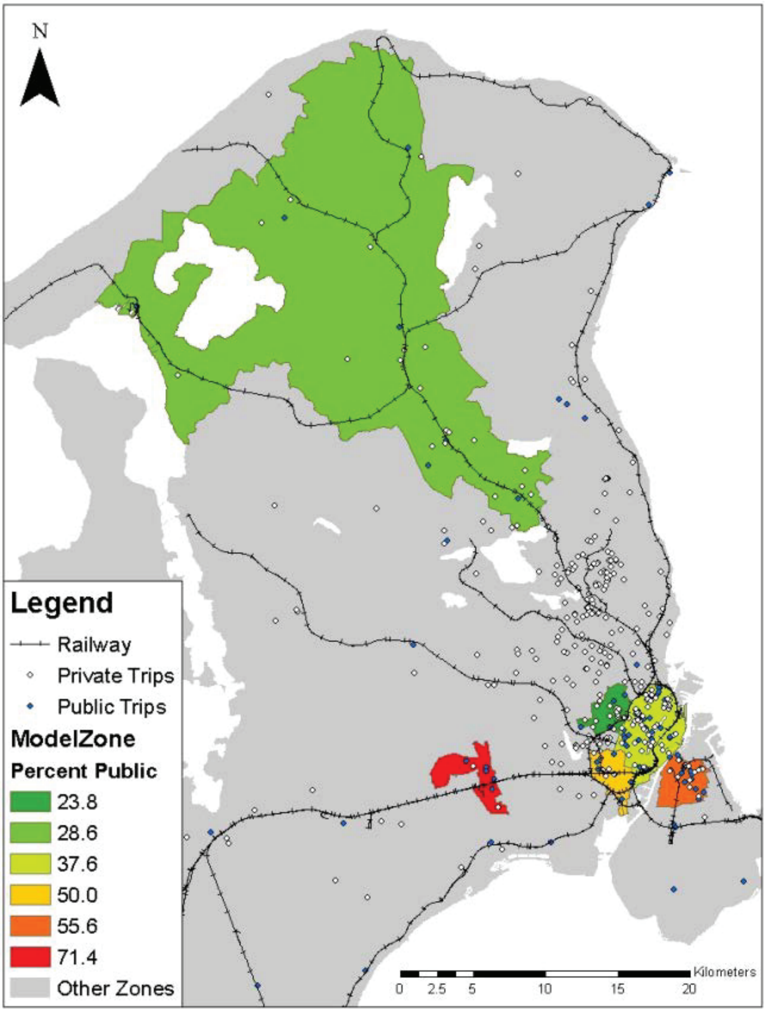

Figure 3-9: Percentages travelling to DTU with public transport – from six selected areas.

3.4.3 Discussion on full scale test

The motives for doing the test survey were met and the survey gave satisfactory results since the amount of data collected was acceptable. Up to 1,000 responses were expected and approx. 600 responses were collected. The final number is thus significantly lower and equals to an answer rate of about 8% of the Danish speaking students and staff at DTU (5.5% of all). In particular, the numbers of responses from students were lower than expected, since less than 3% of all students have filled out the form. Among the employees 8.5% of all employees have responded. Presuming that only the Danish speaking staff contacted via e-mail could answer (approx. 3,000 persons), the response rate among the employees is 13%.

The explanation for the low participation rate among students most likely lies in the fact that this group was not informed properly of the survey. A direct mail to the students had probably given more attention in the survey, and thus a greater response. This supports the theory that a greater response rate is obtained when contacting respondents directly (via e-mail) compared to an invitation via the intranet which is very easy to miss.

The expectation was that a larger proportion of students than employees would answer the questionnaire because students in general have more spare time or value their time lower than employees. This expectation was not fulfilled, but since only a few students were contacted via email, the low response rate is obvious.

Some persons started the questionnaire without completing it, and there is not enough information about why these dropped out. The vast majority of the dropouts (95%) spent less than the estimated 20 minutes to complete (which they were informed in advance), and some of these probably had no intention of completing it from the start. Others might get tired or bored after a shorter period, did not receive from the survey what they expected, etc.

1,096 people opened the questionnaire and 545 completed it. 164 people only opened the questionnaire and did not answer a single question, so of the actual respondents 64% completed the entire survey. It was registered the time when the respondent accessed the questionnaire via the link and every time the respondent clicked “Next” the answers were saved and it could therefore be seen how far the respondent got. It is not known the reason why the potential respondents showed interest in the survey by pressing the link and then gave up before they actually started. A respondent who completed the questionnaire gave the following comment: “A lot of text is presented at the first page”. This might explain why some possible respondents do not begin the actual part of the questionnaire. On top of the persons not proceeding from the introductory page 127 persons gave up after less than 2 minutes. It was reported that the survey was going to take between 10-20 minutes to complete. This was estimated from various tests of the questionnaire and considered an acceptable time use for a questionnaire (Umbach, 2004).

On average 49 minutes were spent on the questionnaire, but this number was significantly affected by the respondents who opened the questionnaire and chose to complete it at a later time and therefore kept it open for several hours. The median ignores the most extreme observations and is therefore providing a more accurate representation of the time compared to the average. The median is calculated on the entire data set avoiding trying to define a border for outliers. The median is 14.8 minutes, which means that the completion time has been estimated quite accurately.

Data were collected in the ordinary TU survey (without route choice questions) in the same period as the pilot survey. Here, Internet respondents on average spent 79 minutes with median equal to 21.1 minutes, thus a slightly smaller time use is observed in the pilot survey than in the ordinary survey. More time spent for the pilot survey than in the ordinary survey was expected because of the introduction of the additional route choice questions and fields to be completed. The lower time use may be caused by the fact that the survey was conducted among students and staff at DTU, which are likely to be more Internet competent than the average population.

The greatest dropout from the survey is as mentioned from the introductory page or within the first two minutes. The extra questions about public transport route choice are asked in step 5 of 6 and it is very difficult to answer the questions in the first four steps in less than two minutes. It is assessed that none of these persons (164+127=291 persons) were presented the questions about public transport route choice and consequently these dropouts are not due to the extra questions.

It could be considered to change the long intro text to ease the reading for the respondents and to ensure that more respondents actually read the text. An idea is to split up the information given by the respondents. The respondents could start to describe the locations they have travelled to during the day and add information about the travel purposes to this. Afterwards the more detailed questions about the trip legs could be asked. With this method the intro text could be given in smaller pieces with only the information relevant at the specific moment and this might lead to more respondents reading the text. The questionnaire is however kept as described to ensure a good flow throughout the survey.

The test survey results showed that it was possible to introduce the additional public transport route choice questions without extending the duration of the investigation significantly. There had previously been reluctance to include route choice questions for fear of how much it would extend the duration of the survey, but the findings in this test survey showed that this fear was not justified or the problem, at least, was not at the assumed large extent.

3.4.4 Conclusion on the full scale test

The expected number of participants was higher than the actual number. There was a great sample potential at DTU, but only a relatively small percentage of the invited employees and especially students chose to participate. The results show that a higher percentage chooses public transport than in the ordinary Travel Survey. This reflects the fact that the invitation indicated, that especially those who have used public transport, should participate.

For trips to DTU approx. 25% have used public transport, particularly bus 150S (which is a fast bus skipping stops and running directly to the DTU campus) or the S-train to Kgs. Lyngby and bus the rest of the way. Respondents are spread evenly across the Greater Copenhagen Area, and only 9 of the 450 respondents come from outside this area. Those choosing public transport are mainly travelling from the Municipality of Copenhagen and from locations along the railway network in general.

The survey has shown a significant number of drop-out of respondents who simply open the questionnaire but never answer any questions. This dropout is not observed to the same extent in the ordinary TU Survey. Approximately 64% of those who actually start to answer the questionnaire in the test survey complete it. This number is 83% in the regular TU. So there is a somewhat higher dropout in the test survey. The median time used to complete the questionnaire is 14.8 minutes in the test survey and 21.1 minutes in the ordinary TU. This suggests that the difficulty of completing the questionnaire has not increased significantly, after the inclusion of questions about route choices in public transport. It is therefore concluded that it is possible to examine the route choices in public transport by using a questionnaire on the Internet, and the next step is to get the survey out to a larger group of respondents.

3.5 Implementation in the TU survey

As a result of the promising results shown by the DTU test survey carried out in this thesis, the public transport route choice questions have been an implemented part of the TU survey since February 2009; see Christiansen (2009) for more details. From February 2009 to May 2010, more than 25,000 interviews were collected with approximately 78,000 trips, of which 4,400 used public transport for at least a part of the trip.

Each use of a transport or access/egress mode during a trip is defined as a trip leg in the TU data. The number of trip legs with bus, S-train, metro and other train is presented in Table 3-5. The table also illustrates how many respondents have entered information on line use (99% of bus passengers, all S-train users) and which train station they travelled via (8% of bus users, minimum 97% of the train users). The low share of train station information for bus users is due to the fact that respondents are not asked this piece of information, but when they travel to and from a train station this piece of information is added to the data set.

Table 3-5: Number of trip legs using the four public transport modes, with respect to number of stated lines and from- and to-stations.

		Number of stated		
Mode	**Trip legs**	**Line**	**From-station**	**To-station**
Bus	3,611	3,576	291	293
S-train	1,122	1,122	1,122	1,103
Metro	885	-	859	866
Other train	491	-	490	476

The number of complete public transport trip observations is high and shows that it is possible for the respondents to fill in the route choice questions. The highest share of non-completed observations is for *from-station* for *Metro* and *to-station* for *Other train* (97%). For other train the missing stations are often in foreign countries (Sweden or Germany) and these stations do not appear on the drop down list. Another explanation for the missing information could be that the automatic calculation of stations to select from the drop down list does not present the actual used station because the respondent has entered wrong information in terms of travel distance, etc.

In average the number of trips legs observed in the TU survey has not been negatively affected after adding the public route choice questions to the national survey as illustrated in Figure 3-10 for the months from February 2008 to January 2010. The figure shows a tendency towards an increase in the average number of total trip legs (black bars) and a constant number of public transport trip legs after the enrolment of the questionnaire with the route choice questions in February 2009. This indicates that the adding of the route choice questions also improved the quality of the travel survey data in general since a higher number of the access and egress trip legs to and from the public transport system and the (walking) transfer legs inside the system are observed after the new questions are added. According to a chi square test the averages are not significantly different from each other.

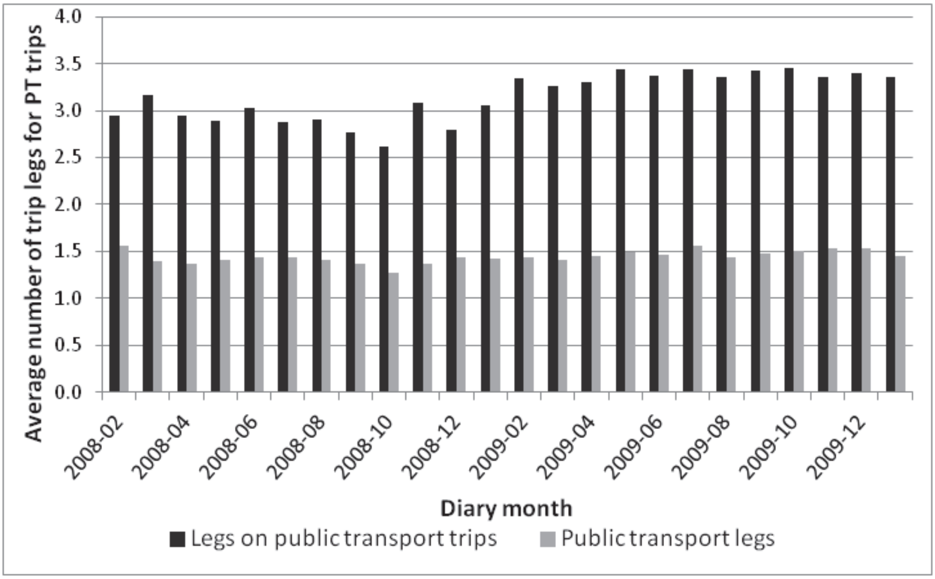

Figure 3-10: Comparison of number of trips legs and number of PT trip legs for public transport trips in the Greater Copenhagen Area before and after adding the route choice questions to the TU survey (primo February 2009).

Both public transport travellers and other travellers have had an increase in the amount of time used to fill in the questionnaire via the Internet after the route choice questions were implemented. Table 3-6 shows the time use, for the data set from May 2006-May 2010, split into before and after the route choice questions were added in February 2009. The time used is shown as median value. Some people log on twice to fill the Internet-based questionnaire and the time use is registered from the first login to the completion of the survey. In these cases the time use could be several days.

Table 3-6: Median [min] of the time used to fill in the TU survey for the 2006-2010 data set.

Interview period	All	Public	Other	Internet	Telephone
All (May 06-May 10)	8.5	11.2	8.2	21.6	7.5
Before Feb. 2009	8.3	10.5	8.0	21.8	7.3
After Feb. 2009	8.8	12.2	8.4	21.4	7.8

The increase in time use for all respondents is in average 0.5 min. For public transport users, the increase is 1.7 minutes. The time use for the public transport users was also the highest before the new questions were added, but the increase for public transport users is higher in percentage than for non public transport users. The median time used by Internet respondents is almost twice the time used for telephone interviews. The median time used by Internet respondents has dropped after February 2009, and this can be caused by many other changes than the route choice questions (java scripts were added to create faster searches and to correct illogical entries). Time use via the telephone use has increased, which could be an effect of the route choice questions.

Table 3-7 shows the difference in time use between the questionnaire version just before and after the implementation of the route choice questions. The version used until February 2009 has a lower time use for all the respondent groups tested. The increases in time use between the before and the after route choice version are from 9 to 24%, the highest for public transport travellers and the lowest for internet respondents. In the comparison between the new and the whole data set, time use for entering public transport trips increases 10%, while time for non public transport users only increases 1%. Internet and telephone interview time use both increase 3%.

Table 3-7: Median [min] of the time used to fill in the TU survey for the 2006-2010 data set.

Interview Period	All	Public	Other	Internet	Telephone
All (May 06-May 10)	8.5	11.2	8.2	21.6	7.5
Oct. 08- Feb. 09	7.8	10.0	7.5	20.4	6.8
Feb. 09-May 09	8.7	12.4	8.3	22.2	7.7

Table 3-6 and Table 3-7 show that there has been an increase in time use for public transport respondents after the implementation of the public transport route choice questions. Comparing the data set just before and after the implementation, a greater increase is registered for public transport users than for private transport users, but the difference rather quickly minimises. The increase can be caused by the fact that the telephone interviewers had to learn how to work with the new questions and, since the highest amount of interviews are completed via telephone; this had a great impact on the results. The time use has not increased too dramatically to be accepted.

3.5.1 Summary and conclusions on full integration in the TU Survey

The question of how to create a survey to collect route choices of public transport passengers was investigated and answered in this chapter until now. This section has investigated the impacts of adding the public transport route choice questions to the national TU Survey. The impacts are measured in additional time use when filling in the questionnaire and the quality of the data is checked.

The requirements for the collection method were:

- Ability to collect for many people.
- Simplicity of the choices.
- Easiness to answer correctly.
- Ability to present the actual chosen route by making it reproducible in a GIS network.

Of the different collection method solutions considered, a questionnaire form was chosen. The original questionnaire was accessible in an internet based form or by an interviewer via a telephone interview. The route choice questions were created with the GIS network in mind and added to the existing TU Survey. The internet based questionnaire was tested in a test survey at DTU in May 2008 proving that it was possible to add the route choice questions to the existing survey and get satisfactory answers. From February 2009 and onwards, public transport route choice data have been collected in the TU survey providing a data set of more than 5,500 route choice descriptions within the Greater Copenhagen Area.

The test survey and the results of the implementation of the route choice questions show that the route choice for public transport passengers can be collected via a questionnaire. The questions are rather easy to answer and do not extend significantly the time used for the survey, and they fulfil the requirements of collecting data enabling reproduction of the route choice in a GIS network (see section 3.6 for details of the matching in a GIS network).

When implemented to the TU survey, great data amounts are being collected from all over Denmark and the number of observations is increasing instantly. This provides opportunity for many different uses of the data.

This section has provided insight on the basic effects of the implementation of route choice questions in the TU Survey. The data describe the actual route choices that passengers make in this public network and will provide the basis for improved estimation of route choice models for public transport. Additional to the research in this thesis several other studies at DTU Transport have benefitted from the route choice questions. A study has investigated the feeder modes to train stations in the Greater Copenhagen Area (Halldórsdóttir, 2010) and this data source provide previously inaccessible amount of information. Rasmussen (2010) used the data for the assessment of generated public transport route choice sets. These studies prove that the public transport route choice survey is useful and has been successful.

3.6 Matching observed public transport route choice data to a GIS network

In this section the methods developed by Rasmussen (2010) and improved to this PhD project to match the public transport route choice data to a GIS network are described. By matching the routes to the network it is possible to compare the observed routes with the routes generated by means of the schedule-based method proposed by Nielsen (2000). The GIS network is a schedule-based public transport network containing addresses, train stations, bus stops, transfer links, train lines, bus lines and a road network. The method developed is able to:

- Identify the train stations used.
- Identify the bus stops used.
- Map observations onto the relevant links with knowledge of line used and points travelled through (origin, bus stops, train stations, and destination).

The matching of the routes is important for the reconstruction of the routes. When matching to a GIS network the exact distance travelled is measurable and comparable for different individuals. When answering the questionnaire, people describe how long each part of the trip is in kilometres and minutes. This measure is very uncertain since people often have an incorrect perception of especially their travel distance (Ankomah et al., 1995, Walmsley and Jenkins, 1992) and to represent the observed routes more exact the matching procedure is developed. The matching of collected data has also proved to be very important for use in the assessment of generated route choice sets and the method is developed and described in this section.

3.6.1 Method

6,547 observations are matched to a GIS network covering the Greater Copenhagen Area. This is a schedule-based public transport network containing addresses, train stations, bus stops, transfers, train lines, bus lines and a road network. The network looks very different at different points in time since over time bus lines are rerouted, new train lines are built, and timetables are changed. It is therefore important to have the correct network to match the observed data to in order to be able to find the actual used lines, departure and arrival stations, etc.

To obtain the correct network for the years the data were collected, the GIS network is built on the basis of the data behind “Rejseplanen.dk” (the Danish Public Transport Route Planner, see Rejseplanen, 2011). These data provide information about lines, stops, stations, time schedules, etc. at a given point in time, and accordingly allow the construction of the public transport GIS network. The important public network data are correct train stations, bus stops, train lines, bus lines and schedules. The network data for a weekday in June 2010 are selected as a good representation of the observed data since bus stops, train stations, public transport lines, and schedules are very similar to the observations.

The following sections explain the steps in the matching procedure and along the way the network elements are referred to. Please refer to Figure 3-2 for the database structure providing an overview of the elements of the different layers in the GIS model.

3.6.1.1 Identification of stations for train trips

The identification of train stations used is straightforward because the respondent enters the train station names in the questionnaire. The respondents choose from a drop down list minimising the chance of wrong entries. The train stations serve S-train, Metro and/or other train (regional and IC), and the station type (if served by several train types) is identified using the train type stated.

The train station names do not have the exact same name in the TU data and in the network, because the respondent might get confused if presented to a too long drop down lists. If train stations were named according to the infrastructure serving them, the respondent had to enter a walking distance between the S-train and Metro even though they stop very close to each other.

When joining the list of train station coordinates to the TU observations, the use of mode on the leg is used in order to identify the correct train station and stop type.

This first step identifies a train station (ID in the Stop table) for 5,704 FromStation StopIDs and 5,703 ToStation StopIDs. For each stop type 58 entries were manually adjusted afterwards.

3.6.1.2 Identification of bus stops for transfer to/from trains at train stations

When transferring between bus and train, a bus stop close to the train station will be used. In many cases several bus stops are located close to a train station, and the following explains the method to identify the actual bus stop used for each leg. By investigating the transfer links from the train stations, all potential bus stops can be found. These stops are compared to the bus stops served by the bus line used and, if several bus stops are found, the one with the shortest transfer distance is used.

The method is divided in several smaller parts.

- The trip legs examined either arrive at or depart from a train station and consequently either the *to-stop* or the *from-stop* should be identified.
- The mode used at the next/previous leg is important to keep track of, because of the different train *types* and *transfer links* to train type stations.
- Most of the *transfer links* in the network are defined as going from a train station to a bus stop, but this is not consistent and hence both directions have to be examined.

This step identifies a bus stop (ID in the Stop table) for 913 FromBus StopIDs and 894 ToBus StopIDs. Afterwards 29 and 31 entries were manually adjusted.

3.6.1.3 Identification of from-stop for bus trips starting near origin – walking or bicycle as feeder modes

It is assumed that the traveller always uses the nearest bus stop served by the desired bus line. This is not always correct, since people are often willing to walk longer towards a bus stop in the travel direction rather than against the travel direction of the bus, even though the latter may be the closest to the starting point.

The identification is carried out for trips where the first part is by bus, or the second part is by bus and the first is by walking, bicycle or car with a threshold of 8 kilometres. The traveller is assumed to use the bus stop on the specific line closest to the origin which is an approximation which is not necessarily correct for the longest access/egress legs.

The identification is carried out by using a network approach in ArcGIS for one trip at a time.

1. All bus stops on the bus line used are selected.
2. The origin coordinate set is selected as starting point.
3. The network distances between origin and bus stops are identified.
4. The bus stop closest to the origin is identified and defined as the *from-stop* for the bus leg.

This step identifies a bus stop (ID in the Stop table) for 3,338 FromBus StopIDs. 5 entries were corrected manually.

3.6.1.4 Identification of to-stop for bus trips ending near destination – walking or bicycle as feeder modes

For identification of bus stops used near destination the same approach as above is used with a few changes.

1. The trips examined use bus mode on the last leg or the second last leg and walking, bicycle or car (-> 3 km) on the last leg.
2. Destination coordinate set is selected as starting point.
3. The network distances between destination and bus stops are identified.
4. *To-stop* is identified and updated.

This step identifies a bus stop (ID in the Stop table) for 3,335 ToBus StopIDs. 8 entries were corrected manually.

3.6.1.5 Identification of transfer stop when transferring between two bus lines

In the public transport TU data set 7.2% of the trips have transfers between two bus lines and, since the respondent does not state the exact bus stop used for transfer, a method to identify the stop(s) is used. Two types of trips are examined, namely trips with two successive bus legs (423 observations) and trips with two bus legs with a walking leg in between (48 observations).

All possible transfers between two bus lines are identified to be stop groups served by both bus lines and bus stops connected with a *change*. If only one possible transfer is identified, this is the transfer used. When more than one possible transfer is identified, additional steps have to be run through.

Not all the transfers identified above are realistic because some can cause great detours for the traveller. Rules are applied to identify the transfer stop:

If the two lines run parallel for some distance, often several bus stops are served by both lines and could be possible transfer locations. The actual chosen transfer stop depends on several factors as service level (travel time, comfort, frequency etc.) of the two lines, service level of the bus stops, transfer time, etc. This is not accounted for and the problem is simplified as follows:

- If the arrival stop is placed so that bus no. 2 drives in the opposite direction of bus no. 1, the transfer is carried out at the first stop possible (left case in Figure 3-11)
- If the arrival stop is placed so that bus no. 2 drives in the same direction of bus no. 1, the latest stop possible is chosen (right case in Figure 3-11)

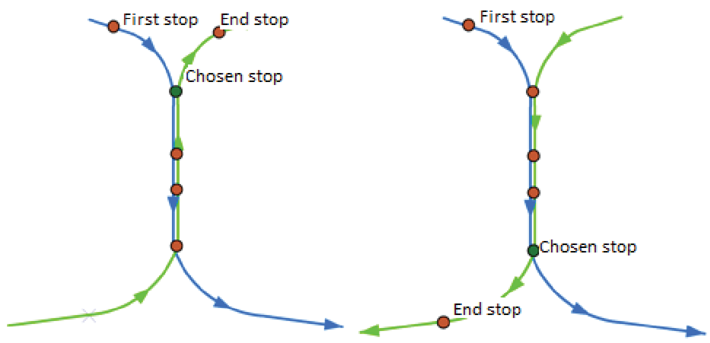

Figure 3-11: Transfer between two bus lines.

Often the routes of the bus lines in the Greater Copenhagen Area are really long and for these even more possible transfer locations exist. Since we do not know the direction of the bus lines used several possible cases could occur, as illustrated in Figure 3-12 .

The first case shows two bus lines in the same direction with a transfer on the last stop before the lines split the first time.

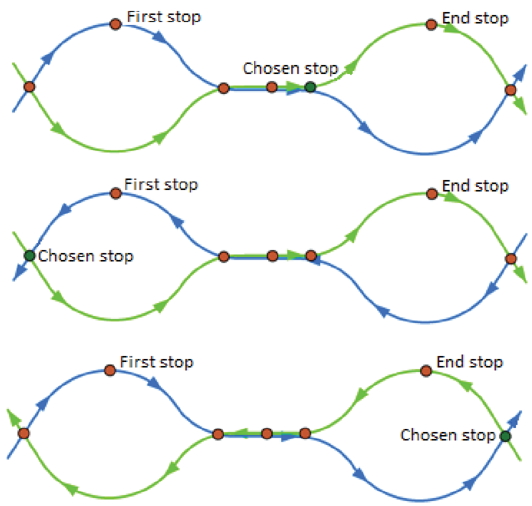

Figure 3-12: Long bus routes meeting and separating several times.

In the next case the first line takes the traveller away from the end stop, followed by a transfer to the second line leading the traveller to the end stop.

The last figure shows the opposite where the traveller travels longer than the end stop and changes for a bus going back to the stop. When the first case shown is possible then this is selected. The above method identifies the arrival stop for the first bus line. When the buses stop at the same bus stop group this stop is also identified as the departure stop for the second bus line. If the traveller has walked between the stops the stops with the shortest *transfer link* in between is chosen.

This step identifies a bus stop (ID in the Stop table) for 562 Bus StopIDs or pair of StopIDs. 402 entries were corrected manually (see more in discussion of the map-matching methods).

3.6.2 Identification of link pieces

In the TU survey the departures of the travellers are truncated to five minute intervals. This simplification of the data imposes some uncertainty in the identification of the exact schedules used for the trip. The following describes the identification of the schedules used by the traveller.

The travel time from the origin to the first public transport stop/station is identified for each traveller. If the traveller used one public transport mode only for the whole trip the departure closest to the departure time from home plus the travel time to the first stop is used to identify the public transport line departure used for the first public transport stop. The departure closest to the point in time calculated by departure time + access time is chosen.

When the trip includes a transfer between two public transport lines the first public transport line is identified following the same method as above. In the transfer, the travel time on the transfer link is added to the arrival time of the first public transport mode and the second public transport mode is identified as the first schedule departure after this point in time.

When the stops used on the route are identified, the link pieces used in between are identified in order to map the complete route. The link pieces on the Line Variant Elements table are selected for each line variant and in the direction of the travel. The element *SQIdx* defines the direction of the line since the attribute is rising in number from the start to the end of the line.

This step identifies the link pieces for all trips (ID in the Line Variant Element table).

The results from the matching are shown in the next section.

3.6.3 Results

With the mapping method described we are able to map 91% of the trips onto the public transport network in ArcGIS. For the remaining trips, data for the exact route are missing or incorrect. In some observations the name of the train station used is missing, and in some the line number of the bus is missing or is incorrect. In some cases, if the train station or bus line used is obvious, the observations can be corrected manually.

The trips which failed to be map-matched by the algorithms can be divided according to five characteristics as presented in Table 3-8.

Table 3-8: Characteristics of trips which are not map-matched, measured in percentages.

Characteristics	Number	Percentage of trips not matched	Percentage of all trips
Bus line (missing/incorrect)	457	83.1	7.1
Train line/station (missing/incorrect)	22	4.0	0.3
Transfers	27	4.9	0.4
Network and other errors	44	8.0	0.7
Sum	550	100.0	8.5

When assessing the table it is important to remember the high number of trips manually corrected during each step in the map-matching algorithms as described above. Anderson and Rasmussen (2010) made similar analyses on a smaller data set and found the difficulties of matching transfers to be the explanation for the high number of not map-matched trips. For this data set the corrections are made manually for 402 records. The share of unsolved transfers (0.4% of the full data set) is very low compared to Anderson and Rasmussen (2010) (6.1%) due to the high number of manually corrected transfers in the present study.

The table shows that the highest number of non-matched trips is found within the missing or incorrect bus line information. As described the most obvious errors are corrected manually but in some cases it is not possible to identify the correct bus line used by the traveller if missing or incorrect information. Because we require the exact route used by the traveller we have to sort out the trip observations with no or wrong information. One explanation for the high number of incorrect bus lines is the fact that observations from a period of three years (Feb. 2009 - April 2012) are map-matched to a network represented by a specific schedule. Over time some buses change line number, line route, stops served etc. and this can lead to a classification of the observation as incorrect. In cases of a bus line changing line number and still serving the same route the observation is changed to represent the actual line in the network but many of these observations are discarded.

3.6.3.1 Example of results visualisation

Figure 3-13 shows the map-matching of the observed routes of four travellers between Høje Taastrup and Frederiksberg/Copenhagen N. All four routes are mapped according to the observations and the method has proven to be able to reconstruct the routes correctly. The figure visualises why mapping the observations is useful. If the information was presented in a table, the routes would be difficult to compare and to assess, especially for people with less knowledge of the network. When the data are matched to the GIS network, the routes are visually comparable and easier to assess also for people with less network knowledge.

None of the four routes in Figure 3-13 are completely identical to another route. Three of the four travellers used the S-train lines B and F for the greatest parts of the trip. One used a bus at the beginning of the trip (line 154E) in order to get to the train station. One of the S-train users chose to disembark the B-train at an earlier station, take a bus (line 13) and depart the F train at a station closer to the end station of the route. The traveller has used twice the time of the other travellers in order to take this bus detour. It is not possible to know why the traveller has chosen this detour, but the explanation can be problems with the S-train departures on the given day or the convenience of travelling along with fellow travellers on the bus line (e.g., bringing a child to school). The fourth traveller used the regional or IC-train from Høje Taastrup to Nørreport and continued by bus line 150S. The destination of this trip is somewhat different from the others, but still this route would have been relevant to the other travellers if they were willing to walk for 1-1.5 kilometres to their destination.

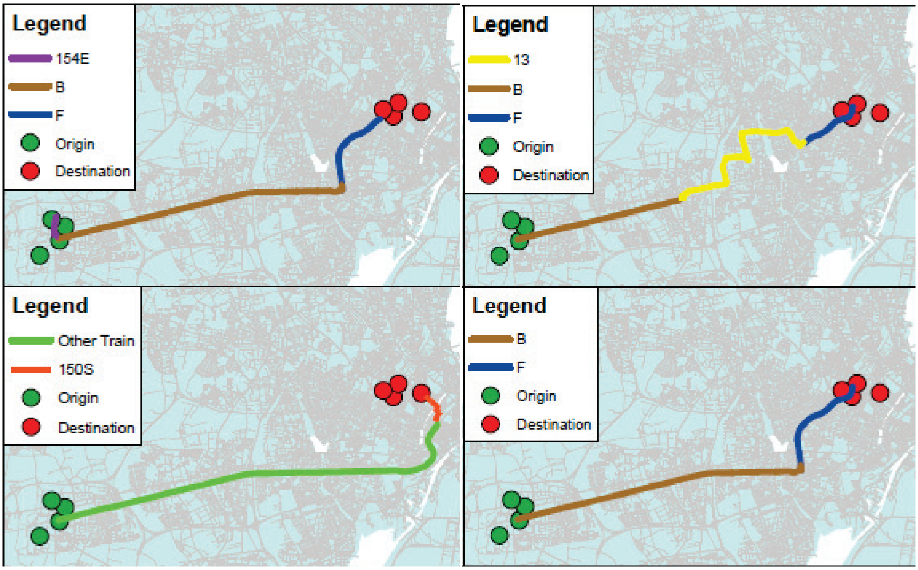

Figure 3-13: Maps of routes for four trips from Høje Taastrup to Frederiksberg/Copenhagen N.

The alternative routes in the figure visualise the differences in the routes that the model should be able to generate. All routes are actually chosen by travellers and are therefore assessed as relevant and must be in the generated route choice set for the given OD pairs.

3.6.4 Discussion of map-matching

For the bus stops some assumptions in the matching method could be questioned. When transferring between bus and train, departing near the origin, and arriving near the destination, it is assumed that the closest bus stop is used. However, the traveller will often benefit more from using a bus stop which is not the nearest. When arriving by bus to a train station the traveller will sometimes alight at the earliest stop if more than one stop is close to the station. This might enable the traveller to run in order to catch an earlier connection. The opposite is the case when departing from a bus stop near a train station. From origin or at destination the same can happen, for example by using a stop later on the bus route than the bus stop closest by, etc.

The assumptions concerning choice of transfer stop when transferring between two buses can be questioned. The last possible stop is chosen with this method, but this can also depend on the service level of the bus (comfort, travel time, etc.) or the situation (can the traveller spot the next bus when sitting in the first).

The issues of identification of bus stops could be solved using the network attributes. When the analyst has information about the lines used it is possible to pick out only the relevant bus lines and carry out a route choice assignment between the origin and destination (or train stations) of the traveller. In this way the assignment model can assist in the identification of the most relevant transfer location between the bus lines. This of course requests the network attributes to be as precise a description of the real network as possible.

The method developed for identifying transfer bus stops is very much dependent on the underlying network attributes. If no transfer link is defined between the two bus lines the algorithm is not able to identify any possible bus stops. As presented the highest number of non-matched trips is found within bus-bus transfer trips and the network attributes could be responsible for this. Ideally the observations from respondents of actual transfer in the network could be used to define extra transfer links to improve the network.

Incorrect or missing information about the bus line used is the most significant reason for a failure in the map-matching (83.1% of the not map-matched trips, 7.1% of the full data set). As explained this high number is mainly due to the fact that actual observations of the route choices are collected over a time period of three years and the schedules in the network represent one day only. Ideally the observations should be map-matched to a network with public transport lines and schedules representing the actual day of travel. To do this not only the correct schedules but also planned changes to the schedules should be taken into consideration in the network creation. A great effort has to be put into the construction and validation of one network only and it is assessed not to be essential to create a separate public transport network for each new schedule.

Overall the differences between the actual network and the network model are acceptable having the high number of correctly map-matched trips in mind. The number of discarded trips with missing or incorrect bus line information is high and a method to ensure a higher match with the network data could be looked into in further research. In this thesis we map-match to the mentioned network and schedules only.

3.6.5 Conclusion on the map-matching procedure

In the above methods to map-match the collected data in a few steps are described. The results from the study shows that it is possible to map-match public transport route choice data collected via a questionnaire in a travel diary form to a GIS network.

The identification of the train stations is easy when the names of the stations used are given in the observations. The identification of bus stops is more cumbersome since these are not mentioned in the observed route choice data. Several assumptions have to be made to identify the bus stops used.

At the start and/or end of each trip including a bus:

- The traveller is assumed to board at the bus stop closest to his origin point served by the stated bus line.
- The traveller is assumed to alight at the bus stop closest to his destination point served by the stated bus line.

When transferring between bus and train:

- The traveller is assumed to board/alight at the bus stop closest to the train station.

When transferring between two bus lines:

- If the two bus lines serve the same bus stop the traveller is assumed to transfer at this stop.
- If the two bus lines serve bus stops connected by a transfer link in the GIS network the traveller is assumed to transfer here.
- If multiple transfer locations possible the traveller is assumed to:
  - Stay as long as possible in bus one if the buses travel in the same direction.
  - Transfer as early as possible if the buses travel in opposite directions.

The “first/last bus stop” and “transfer at train station” methods only consider distance and in several cases this will be different from the actual choice. The bus-bus transfer method assumes that the network contains all transfer links used by the travellers, which is not always the case.

Identification of transfer stops using an assignment model only including the relevant bus lines and origin/destination locations would most likely provide a higher map-matching percentage or a more precise description of the actual route choices of the travellers. This will be an issue for future studies.

The map-matching algorithm provided a successful map-matching of 91% of the observed trips. This number is acceptable for the future use of the data, since a high number of the observed trips are made useable for research purposes. The list of characteristics of the non-matched trips suggests that especially the transfer between two lines should be looked into in future research. Also methods could be improved by offering the respondent a list of relevant bus lines to choose between in order to minimise the number of missing or incorrect bus lines. These issues can be investigated in future studies. The study emphasizes the importance of the high level of detail in the route choice observations and shows that with this level of detail it is possible to develop simple methods which reproduce more than 90% of the public transport route choice observations.

The matching of the actual routes to the GIS network is very important for the future use of public transport route choice observations and the results have been used in several projects at DTU Transport. Halldórsdóttir (2010) used the data to assess and model choice of feeder mode to train stations. Rasmussen (2010) and Larsen et al. (2010) used the matched routes to assess generated route choice sets and finally this PhD study is using the map-matched routes together with the generated route choice sets to estimate parameters for route choice preferences.

3.7 Access and egress leg travel speed

After collecting the route choice data and map-matching the public transport trips to the network also the access/egress parts of the trips have to be considered. The travellers do not provide any information about the route choice of this part of the trip and as mentioned they are assumed to use the shortest route through the road and path network to get from origin to the first stop/station and from the last stop/station to the trip destination. This leaves the analyst with the distance travelled but as shown in this section the time travelled is often even more important than the distance and also has to be estimated.

The respondents are offered the opportunity of entering the length travelled and time used for each of the legs in the trips. Since the departure time from the origin is truncated to five minute intervals, there is a possibility that travellers change the time travelled to match the time they arrived to the first stop/station. Rasmussen (2010) showed that the TU travellers' perception of the access and egress distance travelled is underestimated compared to the real network distance (calculated as the shortest path in the GIS path and road network). The figure is for access and egress legs leading to and from train stations.

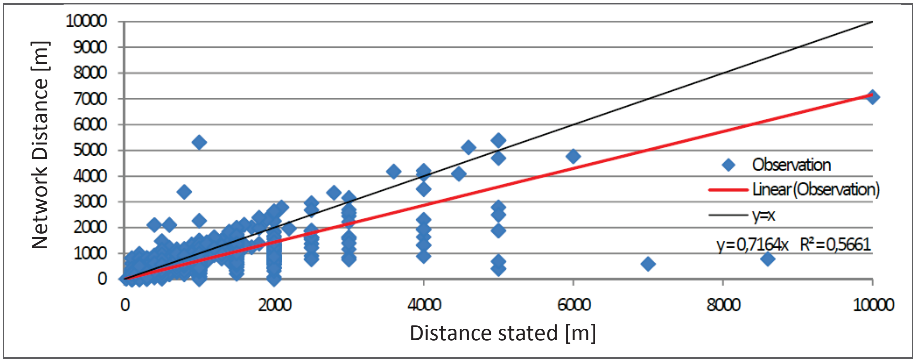

Figure 3-14: Network distance as a function of stated distance for access/egress legs – from Rasmussen (2010).

The regression line shows that there is a tendency towards the TU travellers overestimating the distance travelled. This might also be the case for the time they have travelled on the access and egress legs. In the following we use the proportional relationship between the time and distance travelled mentioned by the traveller to calculate a measure of the speed relationship for access and egress legs in multimodal network trips.

In the literature it is common practice to use the average speed as a constant speed for the access and egress legs. Krygsman et al. (2004) used mean access and egress speeds of 4 km/h for walking and 12 km/h for bicycle. From Dutch data, they found that most travellers were willing to walk up to 550 m and bicycle up to 1.8km for access legs and 600 m and 2.4km for egress legs.

Based on a Swiss study Meister et al. (2010) used average speeds of 2.8 km/h for walking and 12 km/h for bicycling to public transport.

With the great amount of data available for this thesis a study on the speeds used for access and egress to and from public transport is carried out.

In this thesis the access and egress legs are defined by using the private transport modes; car, bicycle and walking which are very different modes which have different speed profiles. It is however very difficult to differentiate the speed of the private vehicles for short distances. The car is faster but the traveller use more time relatively on parking compared to a bicycle which is slower moving but faster to park and finally to walking which has the slowest speed but no parking involved. Below a procedure presented which describes the speed of the access and egress modes and it is assumed that the speeds of the modes are comparable and the speed increase with increased distance to resemble the change in transport mode with distance.

Since access and egress legs are served by the three private modes car, bicycle and walking we search for an average measure of a speed function for the three modes. The TU data shows the travellers having increased travel speed increasing with the access/egress length travelled. The speeds are distributed within the intervals

- Walk: 4 km/h for the shortest trip legs, 8 km/h for the longest trip legs.
- Bicycle: 6 km/h for the shortest trip legs, 20 km/h for the longest trip legs.
- Car: 15 km/h for the shortest trip legs, 40 km/h for the longest trip legs.

The distribution of the used transport modes changes with the access/egress leg length and whether the leg is for access or egress. The end points are shown in Figure 3-15 for the home-end and Figure 3-16 for the activity-end.

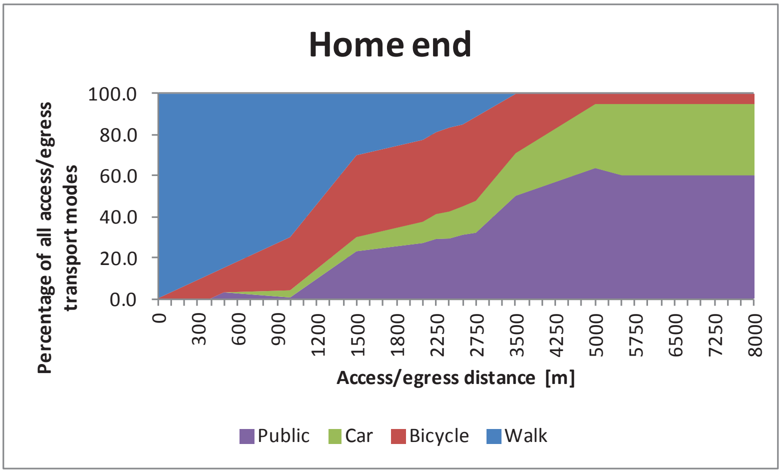

Figure 3-15: Distribution of private transport modes on access and egress home-end legs from the TU Survey.

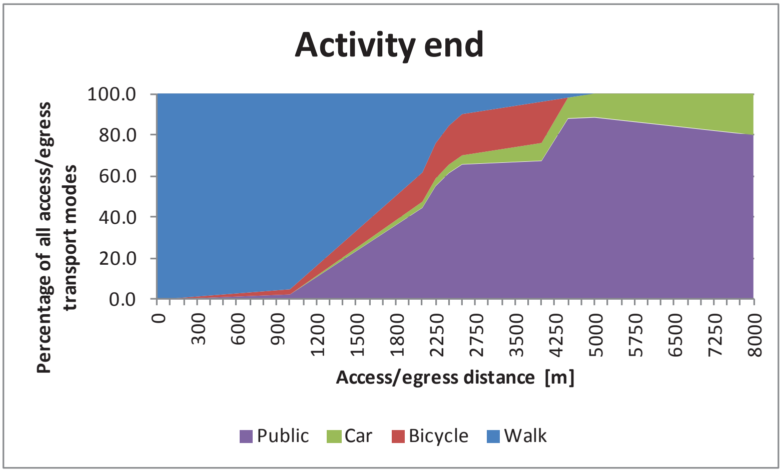

Figure 3-16: Distribution of private transport modes on access and egress activity-end legs from the TU Survey.

By using the information above we define a logistic curve describing the access/egress travel speed as a function of the access/egress travel distance:

$$TravelSpeed_{Access/Egress} = \frac{40}{\left(1+7.75\cdot EXP\left(\frac{-LN(7.75)}{4500}\cdot TravelDistance_{Access/egress}\right)\right)} - 1.2 \qquad \textbf{(3-1)}$$

$$= \frac{40}{\left(1+7.75^{\left(1-\frac{TravelDistance_{Access/egress}}{4500}\right)}\right)} - 1.2$$

Where the *TravelDistance*$_{Access/Egress}$ is the access/egress leg network length measured in m, and the *TravelSpeed*$_{Access/Egress}$ is in km/h. The "40" is the maximum speed defined for the transport modes above.

It follows by Formula (3-1) that the travel time in minutes is calculated as:

$$TravelTime_{Access/Egress} = \frac{TravelDistance_{Access/Egress}/1000}{TravelSpeed_{Access/Egress}/60} \qquad \textbf{(3-2)}$$

where the *TravelTime*$_{Access/Egress}$ is calculated in minutes.

The travel time for formula (3-1) and the travel speed as a function of the access/egress length are shown in Figure 3-17 and Figure 3-18 together with the proposed calculations of access/egress speeds.

The formula (3-1) is based on the shorter travel distances and applies very well to distances up to approximately 2 km. When exceeding 2 km the speed which increases with the length has increased to a degree where the travel time becomes constant. This means that all access and egress legs with a travel distance exceeding approximately 2 km have the same travel time even with increasing distance and the formula is therefore not a good description of the speeds at increasing distances.

The figures also present the travel speed and travel distances for a constant speed of 10 km/h. This speed seems very high at very short trips since 10 km/h assumes that most people use the bicycle or are running instead of walking. At the longer distances however, a constant speed of 10 km/h is equivalent to a travel time of 48 minutes for 8 km access/egress legs. Some travellers in the TU do have long access/egress legs but they state travel times higher than 10 km/h. Also the share of transport modes used indicates that travellers with long access/egress legs use a bicycle or car and do not walk.

An alternative to the travel speed description is a combination of the logistic curve and a fixed speed. The figures show curves for a speed fixed at 10 km/h. Formula (3-1) approaches 10 km/h for 2.4 km long access/egress legs and for longer trips the speeds do not increase further. The 10 km/h is still a low number as mentioned above and according to the TU travellers a travel speed of 20 km/h would be more appropriate.

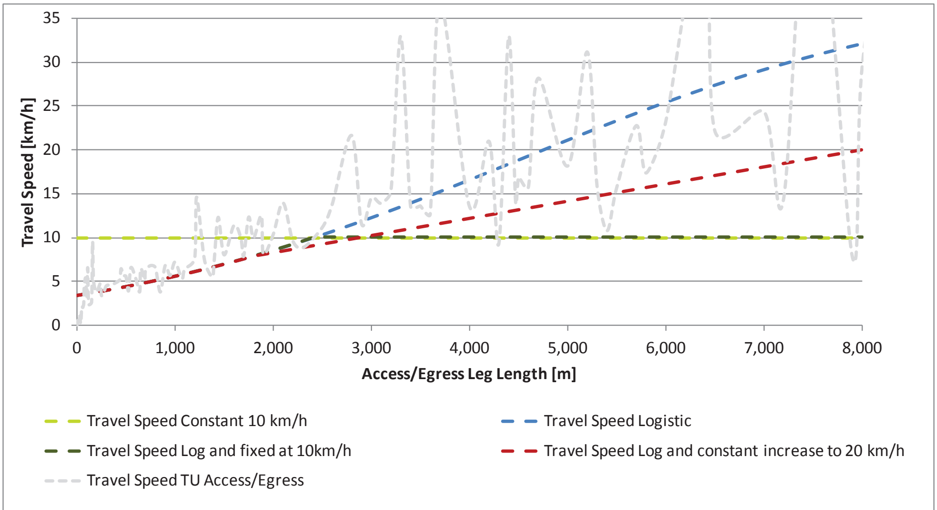

Figure 3-17: Access/egress travel speed in km/h as a function of the access/egress leg length.

The final curves (the red) assume an increase in the travel time from 0 to 1,800 meters following the logistic curve. From 1,800 (with a travel speed of 7.9 km/h) to 8,000 meters the travel speed increases constantly according to the formula:

$$TravelSpeed_{Access/Egress,1800-8000m} = 7.9km/h + 12.1km/h \cdot \frac{TravelDistance - 1800m}{6200m} \quad \textbf{(3-3)}$$

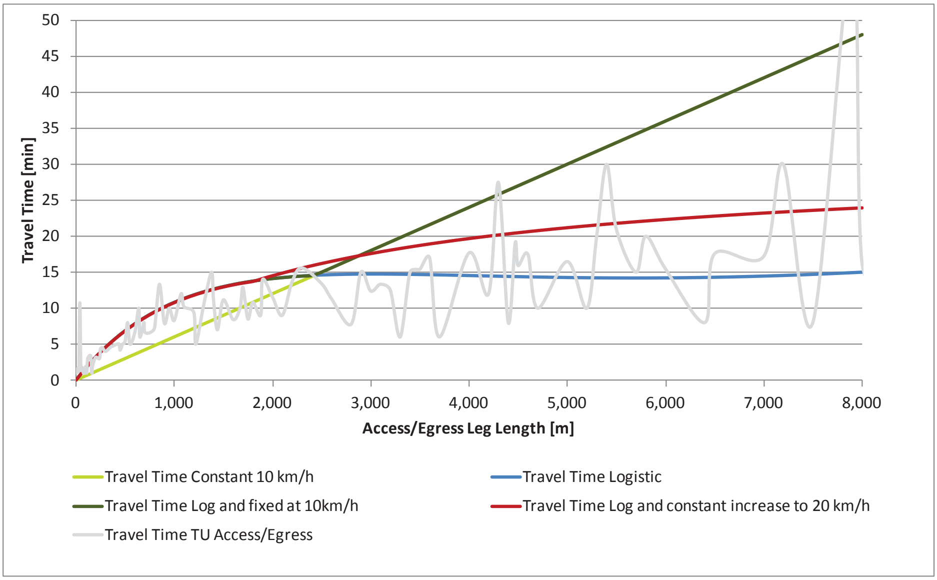

Figure 3-18: Access/egress travel time in minutes as function of the access/egress leg length.

We see from Figure 3-18 that with this constant increase in the travel speed also the travel time will increase continuously up to 8,000 meters. The increase in travel time per increase in meters is greatest at small distances and the increase is smaller at higher distances as described by the data collected in the TU Survey.

To compare the models goodness of fit values for the different speed formulations are calculated. We see in Table 3-9 that for the modelled speeds with constant speeds of 5 and 20 km/h the hypothesis is rejected for all distance intervals. For the 100-3000 meter interval we cannot reject the hypothesis that a constant speed of 5 km/h describes the observed speeds.

For the modelled speeds with the new logistic formula the hypothesis cannot be rejected for any of the shown access/egress intervals. For the high distances the formula of logistic + constant speed does not seem to be as good as the "all obs log speed" but this formula is chosen to avoid a constant travel time for access/egress distances above 2500 m as described.

Table 3-9: p-values for the chi-square of testing the hypothesis that the observed speeds are equal to the modelled speed

	p-values for chi-square test				
Access/egress	Log formula		Constant speed		
distance [m]	For all obs.	+Constant speed	5 km/h	10 km/h	20 km/h
0-8000	0.9966	0.4062	2.2E-48	7.3E-18	1.4E-57
100-8000	0.9999	0.3930	7.3E-40	0.0080	3.0E-20
0-3000	0.9754	0.9485	1.1E-08	3.0E-06	5.0E-43
100-3000	0.9988	0.9943	3.4E-08	0.3245	6.3E-21

A solution to the travel speed and travel time problem for access legs is therefore defined to be a combination of the logistic curve and the constant speed improving the fit with the actual travel speed compared to the simple assumptions of constant speed for all distances. The procedure described is added to the model used for choice set generation in Chapter 5. This model did not distinguish between the transport modes used for access/egress and therefore the access/egress speed formulation is an improvement of this model.

3.8 Examples of trips in the TU survey

In this section examples of map-matched trips from the national survey are visualised. Figure 3-19 shows the route bundle for the whole data set of public route observations.

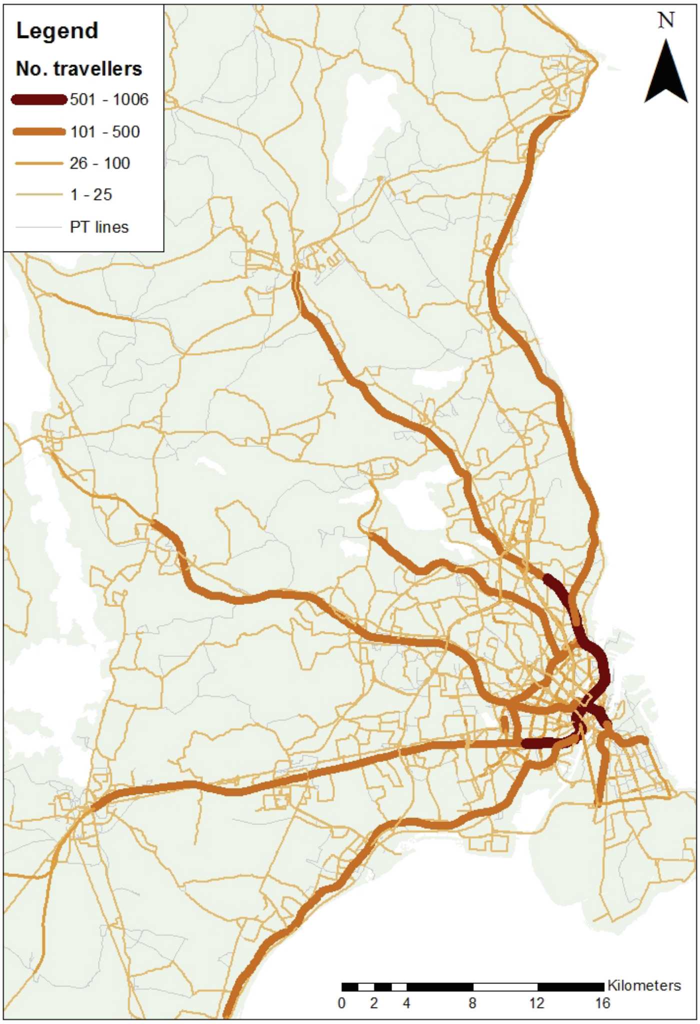

Figure 3-19: Route bundle for all observations (6,451 observations map-matched).

The figure shows that the highest share of travellers is using a public transport mode in or near the centre of Copenhagen (15% of the travellers used the same rails in the CBD of Copenhagen). The thickest lines leading to/from Copenhagen are the train lines and these are used by the highest number of people. Many of the bus lines in the periphery of Copenhagen are used for less than a half percent of the observations. Some routes are not used by any people in the sample, but most of these are small bus lines serving a local area and it is acceptable for the further analyses that no one have stated to use these. When the sample gets larger, users of all the small routes will be included at some point.

Figure 3-20 shows a trip from Roskilde to DTU campus. The traveller uses three public transport modes for his trip: IC-train from Roskilde to Copenhagen, S-train line E from Copenhagen to Lyngby St. and bus 300S from Lyngby St. (bus) to DTU Campus. As can be seen a ring rail line in the outskirts of Copenhagen would be ideal for this trip but since such does not exist the traveller has to travel through Copenhagen, despite the detour given by this.

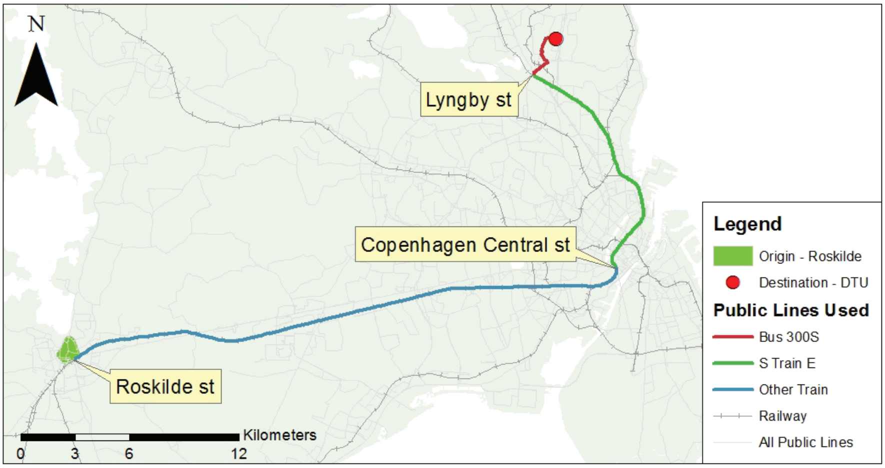

Figure 3-20: Use of public transport lines from Roskilde (aggregated area) to DTU Campus – 1 trip.

Figure 3-21 shows the public transport modes used by three travellers from Albertslund/Glostrup to DTU Campus. The two travellers from the east part of the zone walk for 400-700 meters to the bus stop and use the rapid bus line 300S from Glostrup Hospital. The traveller from the west of the area walks 300 meters to bus line 143 which he uses to go to Vallensbæk station (in the south) and transfers to bus 300S which he uses to DTU Campus. To use 300S from the start he had to walk/bicycle 2.1 km and he prefers the detour (driving south even though his target is to the north) over the long access trip to bus 300S.

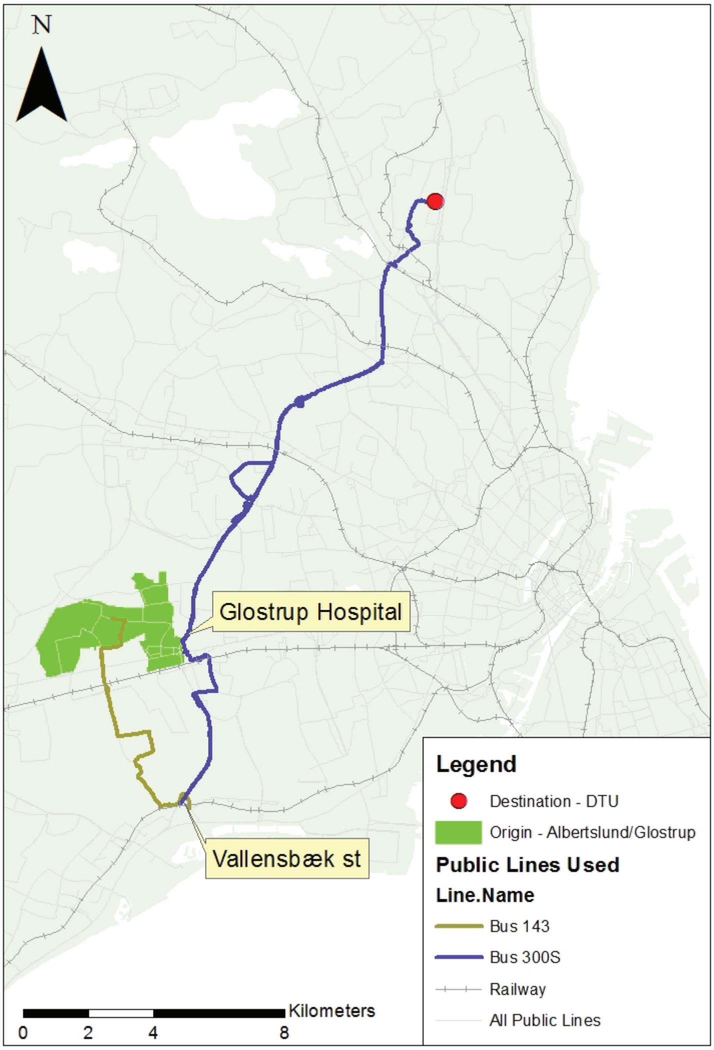

Figure 3-21: Use of public transport lines from Albertslund/Glostrup (aggregated area) to DTU Campus – 3 trips.

In the data survey seven respondents travel from the north of Amager to DTU campus. The travellers all travel via Nørreport station (one person bicycles to Østerport st.) either by bus or train. From Nørreport st. either the S-trains line E and B (B has more stops than E but the E train does not overtake the B train so the travellers most often use the first train arriving) or the buses 150S and 173E are used. The buses use the same links (primarily the motorway) and from the stop it is only a short walk (300 m) to the destination. The S-trains are faster but stop at Lyngby st. (3 km from DTU) and most travellers use the bicycle or the bus as egress mode from the train. The seven trips are as shown in Table 3-10 (primary mode in bold):

Table 3-10: Mode chains used for trips from Amager to DTU Campus, with access and egress mode in numerical order.

	Access Mode no.		Primary Mode	Egress Mode no.	
No	1st	2nd		1st	2nd
1	Bicycle to Østerport		**S-train to Lyngby**	Bicycle to DTU	
2	Walk to Christianshavn	Metro to Nørreport st.	**S-train to Lyngby**	Bus to DTU	Walk
3	Walk to Christianshavn	Metro to Nørreport st.	**Bus 150S to DTU**	Walk	
4	Walk to Christianshavn	Metro to Nørreport st.	**Bus 150S to DTU**	Walk	
5	Walk to bus stop	Bus to Nørreport st.	**Bus 150S to DTU**	Walk	
6	Bicycle to Nørreport		**Bus 173E to DTU**	Walk	
7	Walk to Islands Brygge st.	Metro to Nørreport st.	**Bus 150S to DTU**	Walk	

Figure 3-22 shows the public transport line variants used by the seven travellers. As can be seen the distances for routes using bus (dark purple) and train (light purple) are very similar. This makes the choice of route for these travellers very interesting since depending not only on time and distance but also on other characteristics. The figure implies that the characteristics of the transfer locations might be important for the travellers and this will be investigated when estimating the route choice models in Chapter 6.

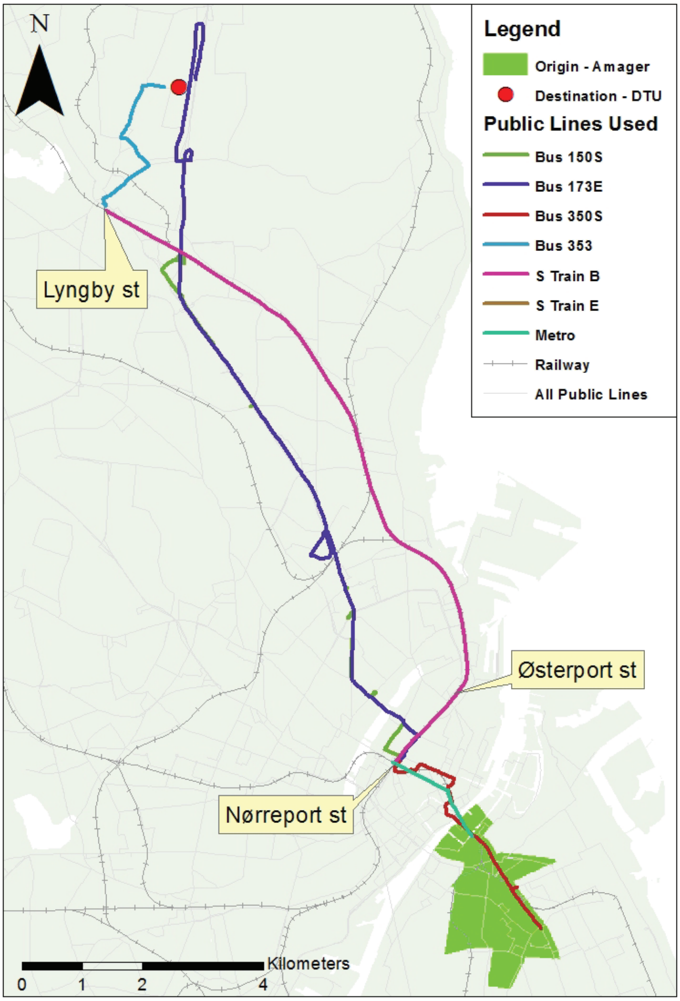

Figure 3-22: Use of public transport lines from Amager (aggregated area) to DTU Campus – 7 trips.

Nørreport st. is an important point in the Greater Copenhagen Area public transport network since almost all trains and many buses pass through here. All the seven travellers between Amager and DTU Campus travel via Nørreport st. but none of the seven travellers use the exact same route to go from approximately the same point of origin to the same destination. This is a good example of the variety of the route alternative in the multimodal transport network and emphasizes why the route observations in this public network are a good basis for route choice model estimation purposes.

3.9 Summary and conclusions

3.9.1 Data collection

In this chapter the method of collecting public transport route choice data has been presented and described. The data were collected using the existing TU Survey and adding questions concerning public transport route choice for this.

The route choice questions added to the TU survey were shortly and precisely formulated in order to keep the questionnaire simple, obtain high completion rate, and collect good and useful observations. The requirements for the method was to collect information detailed enough to enable the reproduction of the route, but also simple enough for the respondent to provide it correctly. By answering questions about specific points on the trip the route can be reproduced with the knowledge of the public transport network.

The questionnaire was tested in a full scale survey at the Technical University of Denmark in May 2008. Potentially 10,000 staff and students could have participated in the survey. Employees were invited to the survey via an e-mail and students were invited via a link on the intranet. 600 respondents completed the questionnaire providing a great source of data to assess the quality of the survey.

On February 2009, the public transport route choice questions were added to the national travel survey and data has been collected continuously since. The adding of the questions has not caused a significant increase in time use for the questionnaire and the method is therefore accepted as a permanent part of the national survey. The route choice data are collected in terms of public transport lines used and stops/stations travelled via. The observations are complete for more than 97% of the total number of public transport trips and the method is therefore assessed as being able to solve the challenge of collecting route choice data in a simple and effective way.

Some ideas for improving the collection method are introduced:

- Possibility of clicking an interactive map. If the interviewer had a map of the buses he would be able to point to issues with wrongly stated information, for example bus lines. This demands a network equal the exact network of the day of travel.
- If respondents stated bus stops used (just writing the name in free text) the manual fixing of the observations would be easier since only bus stops served by the line and with a name equal to the stated should be considered.
- At each major change in time schedule the network should be created and saved for further use.

The method for collecting data shows to be very useful for collection of route choice data for passengers in multimodal public transport networks and adds to existing literature by introducing a method to collect this detailed information in a simple way, still allowing for the data to be reproduced and visualised in a GIS network.

3.9.2 Map-matching

Methods to map-match the collected data in a few steps were described. The results from the study showed that it is possible to map-match public transport route choice data collected via a questionnaire in a travel diary form to a GIS network.

The identification of the train stations is easy when the names of the stations used are given in the observations. The identification of bus stops is more cumbersome since these are not mentioned in the observed route choice data. Several assumptions have to be made to identify the bus stops used.

At the start and/or end of each trip including a bus:

- The traveller is assumed to board at the bus stop closest to his origin point served by the stated bus line.
- The traveller is assumed to alight at the bus stop closest to his destination point served by the stated bus line.

When transferring between bus and train:

- The traveller is assumed to board/alight at the bus stop closest to the train station.

When transferring between two bus lines:

- If the two bus lines serve the same bus stop the traveller is assumed to transfer at this stop.
- If the two bus lines serve bus stops connected by a transfer link in the GIS network the traveller is assumed to transfer at these stops.
- If multiple transfer locations possible the traveller is assumed to:
    - Stay as long as possible in bus one if the buses travel in the same direction.
    - Transfer as early as possible if the buses travel in opposite directions.

The map-matching algorithm provided a successful map-matching of 91% of the observed trips. This number is acceptable for the future use of the data, since a high number of the observed trips are made useable for research purposes.

The number of not-matched trip is mainly caused by incorrect or missing bus line information. If obvious the bus line is corrected manually but we require the exact information and have to discard the observations with no or wrong information. The route choice observations are collected during a three year period and map-matched to a network representing a specific point in time. The public transport network (lines, line names, schedules) has changed over these three years and this can lead to a classification of the observation as incorrect. Some observations can be corrected if the transport line has a changed line number only but if the route is not the same the observation is discarded.

The study emphasizes the importance of the high level of detail in the route choice observations and shows that with this level of detail it is possible to develop simple methods to reproduce the public transport route choice observations.

The matching of actual routes to the GIS network is very important for the future use of public transport route choice observations and the results have been used in several projects at DTU Transport.

3.9.3 Access and egress leg travel time and travel speed

By use of the observed route choice data from the TU Survey a measure of the travel speed as function of the travel leg distance is defined. A logistic curve according to the following describes the travel speeds on access and egress leg from 0 to 1,800 meters:

$$TravelSpeed_{Access/Egress,0-1800m} = \frac{40}{\left(1+7.75\cdot EXP\left(\frac{-LN(7.75)}{4500\cdot TravelDistance}\right)\right)}\text{-1.2}$$

A constant increase in the travel speed is defined for access and egress legs from 1,800 to 8,000 meters:

$$TravelSpeed_{Access/Egress,1800-8000m} = 7.9km/h + 12.1km/h\cdot\frac{TravelDistance-1800m}{6200m}$$

The logistic and constant increase in the travel speed as functions of the access and egress leg distance describes very well the observed travel speeds from the TU Survey. The travel speed increase is caused by the shift in transport mode choice according to the access/egress leg travel distance.

In the last section of the chapter examples of map-matching of actual routes were given. The matching gives a good visualisation of the observation which eases the accessibility of the data for people with little knowledge of the network. Since the observations are matched in the same network as used for the generation of route choice sets in Chapter 5 the observed routes and the generated routes are possible to be compared and to be used for the estimation of route choice models. The visualisation of the observed routes also shows examples of several travellers travelling from the approximate same point of location to the same location and the observations show the variation in the alternatives since many different route alternatives are chosen by the travellers.

4 PUBLIC TRANSPORT ROUTE CHOICE DATA

The following chapter present the data used in the remainder of the thesis. The descriptions and analysis are both route choice and mode choice related in order to show the variety of the data and to prepare for the route choice part of the thesis. The focus of the analysis is to investigate relations between private and public transport modes and between unimodal and multimodal trips. The analysis presents both analyses similar to general findings to show the viability of the data and analysis of more unique character. The analyses are in general focused on number of trips since the specific trips are used in the PhD study.

4.1 Characteristics of trips and mode chains in the TU Survey Data

In this section the data collected in the TU survey are presented and various analyses are provided in order to understand the data and provide greater insight into the travel habits and mode chains of Danish travellers. The analyses are conducted in two parts:

- The data collected in the Danish Travel Survey, TU, from 2006 to 2009 is used for the *characteristic of public and private trips*.
- The detailed public transport route choice data collected in the TU Survey from 2009 and to present is used for the *characteristics of multimodal mode chains*.

The investigations of these data first enable pointing out which travellers' and trips' characteristics to pay special attention to when examining the choice between private and public transport. When examining the data in point two, it is possible to point out the characteristics in the choice between unimodal trips (e.g., train) and multimodal (e.g., car and train) to pay special attention to.

This is important not only for the future estimation of mode choice and public transport route choice models, but also for the general information about mode chains and knowledge about which issues should be taken into account when trying to improve the conditions for public transport users. The section also shows which initiatives could convince more people to travel by public transport and the factors which affect the choice of modes.

The analyses show that income is very important for the choice between public and private transportation since income also reflects gender, place of residence, etc. and that both trip characteristics and demographic factors are important for the choice. For mode chains the six most important are discussed further. The chapter builds partly on the work presented in Anderson (2010a).

4.1.1 Introduction

This section illustrates and discusses results from the analyses of the TU data. Two main issues in the analysis are described.

The first is the choice between *public and private transport modes* which is an important issue in a planning context. The use of the private car is much more polluting than the use of public transport (e.g., bus, train) depending on occupancy rate. In fact, cars induce congestion that has a great cost for the travellers and the society. The investigated aspects related to the mode choice include characteristics concerning the trip, the traveller, the home, the workplace, etc.

In literature, the analyses of the TU have often focused on a specific aspect of the survey and a more detailed analysis of one issue. Christensen (2000) investigated the impact of the public transport service on the behaviour of travellers in order to understand the possibilities of transfer from car to public transport modes. Christensen (2001) analysed the effects on private transport and the environmental impact of urban size, structure and localisation, and focused on the three main areas of the locations of residential areas, workplaces and centre functions. Christensen (2001) looked at transport mode and amount of generated traffic. Christensen and Jensen (2008) focused on the potential of switching short car trips up to 22 km to walk and bicycle trips.

The second issue investigated is the choice of *mode chains*. Multimodal transport is often thought to be beneficial to society since it is considered more sustainable and environmentally friendly than unimodal car transport. Multimodal trips combine the best characteristics of each transport mode in terms of accessibility, speed, use of space, etc. and a multimodal trip often consists of at least one public transport mode. Multimodal transportation is less often used than unimodal transport and it was interesting to investigate what factors cause people to use more than one transport mode. The main issue in this research was the choice between unimodal and multimodal travelling, and various aspects were investigated to determine important factors affecting this choice.

At a political level, multimodal systems have been the focus of many projects. The European Commission (1995) encouraged the use of multimodal transport by improving opportunities for changing between private and public transport networks, for example, by improving park-and-ride facilities. Several projects under the European Commission have focused on multimodal (also called intermodal) transportation of passengers, e.g., the project LINK (2010), which had partners from 18 European countries and recommended multimodal travel, listing ways of improving the opportunities for multimodal travel in Europe. A number of national authorities urge the use of multimodal transport for environmental reasons, amongst others. The Danish Ministry of Transport (2009) had its focus on the interaction between private and public transport modes and the relevance of improving parking facilities. The Government of India (2006) defined the best public transport system as a system that allows seamless travel in terms of intermodal transfer and ticket use.

The choice of transport mode is affected by several factors. Ortúzar and Willumsen (2011) classified important factors into three groups: characteristics of the trip-maker, of the journey, and of the transport facilities, including car availability, driver’s license, residential density, trip purpose, etc. For multimodal trips, in particular, several characteristics have been investigated in several contexts. Van Nes (2002) analysed Dutch observed travel data and defined a framework for multimodal travelling. He found that the important factors affecting the choice of multimodal travelling were the main transport mode, trip distance, urbanisation level, and trip purpose. Van Nes (2002) also showed that more than 20% of the interurban trips to major cities in the Netherlands were multimodal, and that trains were used in 60% of these trips. Krygsman and Dijst (2001) found that important aspects in multimodal travelling are short access and egress distances, the purpose of the trip (mostly education and work travellers choose multimodal), urban intensity, and car ownership. Barnes (2005) emphasised the importance of destination density for the choice of public transport.

However, the literature includes very little research that makes use of actual observations of data on the attributes of each trip leg and of the transfer locations, since this data is extensive and it is difficult to identify a simple and thorough method to collect the data. Hoogendoorn-Lanser et al. (2006) paid special attention to the transfers in multimodal travelling and found 13 transfer attributes to be important for the multimodal trips in a train hub-n-spoke network.

In the TU, questions to reveal the exact route chosen with public transport were implemented in 2009 (Anderson, 2010b and the previous chapter) and data has been collected since February 2009. Besides the trip and mode information, the demographic characteristics of the respondents are also collected, which reveals many interesting details about traveller choices in the transportation network. Halldórsdóttir (2010) used this data to show that, in trips with the train as the main mode, the socioeconomic variables have the greatest influence on the choice of access/egress mode.

The analysis of the *private and public transport trips* points to important aspects of the mode choice and is important input to the following designing of transport models. The results are presented as frequency and distribution analyses and concern trip distance, trip purpose, gender, age, and many other traveller and trip characteristics. The analyses are shown as figures and tables and the information read from these are discussed. The discussions build on this information, existing knowledge and some hypothesis of the mode choice behaviour are suggested.

This analysis of the mode chains uses the detailed data from the TU Survey to investigate the choices of multimodal trips in further detail. A binary logit model applied to the characteristics of travellers, trips and journey are used to pick out the most important factors in the choice between unimodal and multimodal travel, which provides new insights into to how the share of multimodal trips can be improved.

4.1.2 Data

The TU survey is described in details in Section 3.3.

Two different data sets extracted from the TU survey have been used for the analyses in this chapter. For the analysis in Section 0 the data used are collected between May 2006 and December 2009 including 54,695 interviews with 166,994 observations of trips. The analysis in Section 4.3 is built on data from May 2006 to April 2011 from the Greater Copenhagen Area with

23,427 interviews and 65,527 records of vehicular trips. Only trips in the Greater Copenhagen Area have been used for the second analysis regarding choice of multimodal transport, since only the public transport network in this area offers realistic multimodal trip alternatives to travellers compared to the rest of Denmark.

4.1.3 Definitions used in the TU Survey

The following introduces the classification of some transport modes and trip purposes from the national travel survey into more convenient categories.

Respondents in the TU have the possibility of selecting one of 21 transport modes when describing each trip part. These 21 modes cover the supply of modes in the transportation network but in order to ease the overview of the tables and the graphs presented in this chapter they are divided into the six categories described in Table 4-1. It should be noted that *car driver* and *car passenger* are not listed as separate transport modes in the TU, but the purpose of this analysis advised to introduce this differentiation. Similarly, the *ferry* mode was also separated in car ferry and passenger ferry since these two modes satisfy different requirements for the travellers.

Table 4-1: Classification of the transport modes from the TU.

Name	Description	No. observations	Percentages of full sample
Walking	Walking, skateboard	27,891	16.7
Bicycle	Bicycle	27,414	16.4
Car Driver	Car (driver), van, MC, ferry (car ferry)	77,142	46.2
Car Passenger	Car (passenger)	20,720	12.4
Public	Bus, S-train, other train, metro, telebus, passenger ferry, boat, airplane	9,581	5.7
Other	Moped (30,45 km/h), horse wagon, truck, tractor, taxi, tourist bus, (blank)	4,256	2.5

In the TU survey, respondents select among 25 trip purposes to describe the purpose of the trip. Similarly to transport mode, trip purposes are divided into six categories in order to simplify the description and the discussion of the data, according to the classification presented in Table 4-2.

Table 4-2: Classification of the 25 trip purposes from the TU.

Name	Description	No. observations	Percentages of full sample
Home	Home	67,401	40.4
Work	Work	17,946	10.8
Education	School, education,	5,548	3.3
Errand	Shopping, bank, library, social/health	38,329	23.0
Leisure	School care, youth club, visiting family/friends, sports, entertainment, weekend cottage, holiday, private meetings, evenings school	34,572	20.7
Business	Meeting, conference, customer visit, craftsmen, business trip, business transport of goods or persons, roadwork, police trips	3,105	1.9

4.2 Characteristics of trips and travellers in private and public transport

In this section, the different analyses are illustrated and discussed. The discussion is kept to a minimum only bringing out the most important issues and much information can therefore only be read from the graphs and tables and not from the text.

4.2.1 Analyses of the TU survey data

4.2.1.1 Analysis of primary mode share and distance

Table 4-3 shows the average distances between the origin and destination of the trips for the six categories in which modes have been differentiated. Average trip distances are significantly different for each category with respect to the others.

Table 4-3: Average and standard deviation of the distance between origin and destination by mode category.

Primary Mode	Avg. Distance [km]	St. Dev. distance	No. observations
Walking	0.68	1.00	27,891
Bicycle	2.01	3.12	27,414
Car Driver	12.17	21.85	77,142
Car passenger	15.89	28.81	20,710
Public	17.30	31.39	9,581
Other	14.53	31.25	4,256

Quite interestingly, trips using public transport as primary mode are the longest, and trips as a car passenger are longer than trips as a car driver, suggesting that people most likely drive alone on shorter distances and have passengers in their vehicles on longer distances. Leisure and holiday trips are often performed over long distances, and people often travel with friends or family and therefore more people are in the car.

Figure 4-1 shows the choice of transport mode related to the distance between origin and destination. Mode choices are very different according to the distance, especially for shorter trips.

When considering trips of a few kilometres, the order of the mode shares is obvious and the choice from approx. 8-40 km is relatively stable. Exceeding 40 km the car driver share drops in favour of the car passenger share. Car as a driver is chosen by the majority, but at distances exceeding 40 km the car as a passenger is also rather high, indicating that more people travel together at long distances. Even though the share of car trips as drivers and passengers are alike, it does not mean that all car drivers have a passenger in the car. Often when people have passengers in the car, they have more than one passenger and in most cars it is possible to transport 3-4 people additional to the driver. Longer trips are often leisure and holiday trips having a higher number of people travelling together.

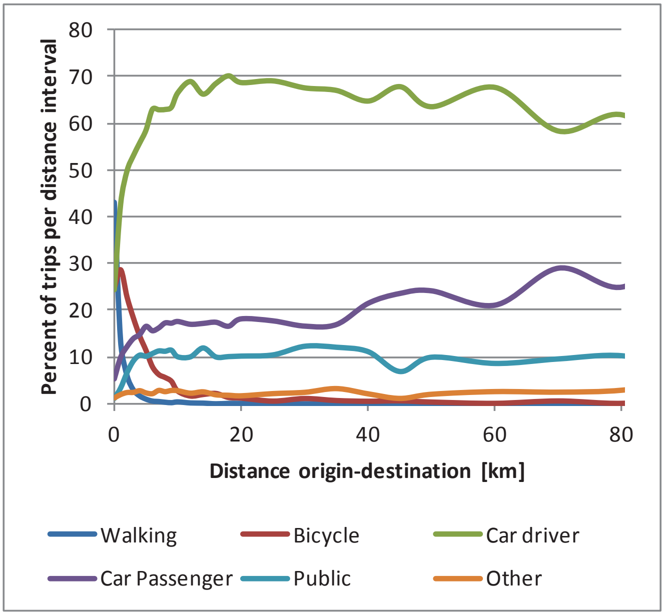

Figure 4-1: Choice of mode related to the distance between the origin and destination of the trips.

For all trips with distances between origin and destination exceeding a few kilometres, public transport is used in approximately 10% of the trips. The public transport classification covers many different transport modes used for different trip lengths. At the shorter OD distances, bus and local trains are used for the majority of the trips. At longer distances, regional and national trains and airplane are used. The choice of public transport can be caused by the convenience of using these modes or the fact that a car is not available. When exceeding 4 and 8 km, respectively, the distance gets too long for the traveller to walk or bike and he/she chooses car or public transport instead. At long distances, the use of the train can be more comfortable than car since it is possible to read, sleep, etc.

Car driving is often convenient for longer distances if a car is available to the traveller. Travelling as car passenger can be difficult at the ends of the trip. At the start point, the traveller perhaps has to travel to a pick up point. The passenger is also dependent on the desired departure time of the car driver. At the end point, the passenger is perhaps dropped off at a point different from the destination point and the arrival time is perhaps not the optimal. These difficulties cancel out at greater distances because the passenger obtains benefits closer to the benefits of the driver at longer distances.

Figure 4-2 shows a zoom of the first 8 km of the graph in Figure 4-1. This is the part of the previous graph where the mode shares really change. From this point on, the shares are rather stable. Walking trips drop from 65 to 8% within the first two kilometres. For trip distances of 500-750 m there are equal shares of people walking, biking and driving a car. At the shortest distances it is easier to walk, bike or drive a car, but for distances of 3 km and higher the public transport share reaches 10%. If the traveller has to walk or bike to either the bus stop or the train station, it might be easier to walk or bike for the whole trip, but at longer distances the benefit of public transport is higher and therefore this mode has a higher share. The share is stable from this point on because not all people consider public transport and for many people living in the countryside using public transport modes is very difficult.

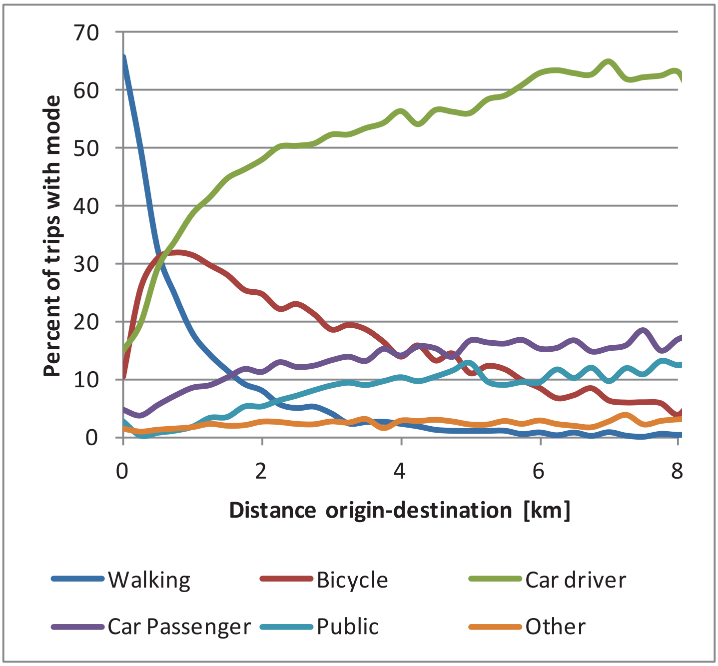

Figure 4-2: Choice of mode related to the distance between origin and destination of the trips (zoom of Figure 4-1).

Even though the trip with public transport modes is short it can contain many transfers. Studies show that the majority of travellers have impedance against transfers between public transport modes and therefore try to minimize the number of transfers (see Daly and Gunn, 2002, Fosgerau et al., 2007, Nielsen et al., 2001). Public transport modes are most often chosen for trips where it is possible to have no or few transfers (use of train and high class bus) and less often for trips that would include many transfers. Car as driver increases in the whole interval, and for longer distances this mode is much more practical than walking or bicycling and has greater benefits in terms of comfort, travel time, etc.

4.2.1.2 Analysis of trip purposes

Table 4-4 shows the distances between the origin and destination and the number of trips for the six destination purposes in the survey data. The average trip distances of all purposes are significantly different from the others except the comparison of education and errands having a t-value of 0.7. Business trips are the longest trips, since people travel longer distances for meetings, seminars etc.

Table 4-4: Average distance in kilometres and standard deviation between origin and destination by destination purpose.

Destination purpose	Avg. Distance [km]	St. Dev. Distance [km]	No. observations
Home	10.12	23.59	67,401
Work	12.31	18.62	17,946
Education	6.07	12.47	5,548
Errand	5.94	13.72	38,329
Leisure	9.71	21.65	34,572
Business	21.51	36.08	3,105

Work trips are the second longest and are not significantly different from the car driver trips in Table 4-3. The highest number of trips is homebound, for example if a person goes to work, then home, then to a secondary activity and then back home, two trips that day have been with home as a destination. Errands and leisure trips are the next most numerous. Only 18,000 trips are to work compared to the 26,000 employees in the sample because of weekends, vacations, people working from home or being sick, self-employed people, etc.

In Figure 4-3 the choice of primary transport mode in relation to trip purpose is observed.

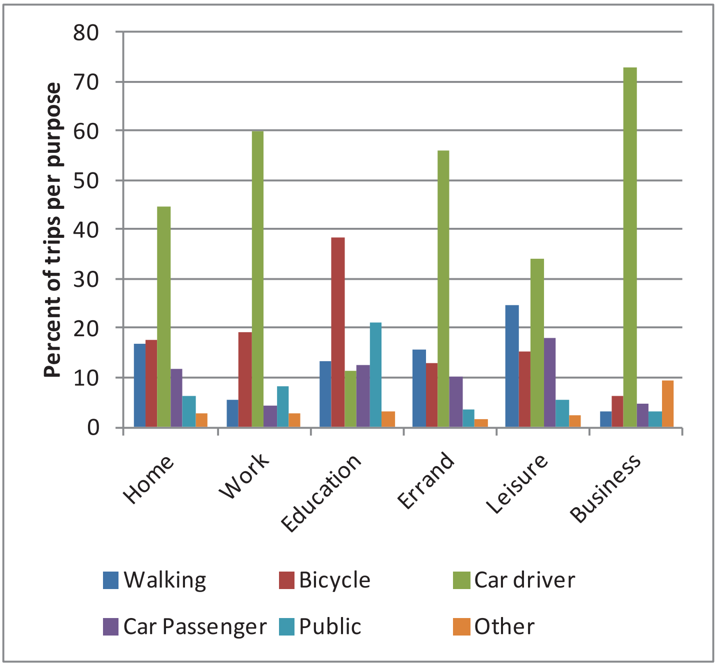

Figure 4-3: Choice of primary transport mode in relation to purpose at destination.

For all trip purposes except education, most people travel by car. For educational purposes, most people use the bicycle. The highest share using public transport (over 20%) is found among the students. Most likely, students have lower car ownership and therefore choose other transport modes. The reason for the high shares of public transport modes might be that there is often a high public transport level of service at schools and universities, making it more convenient to choose public transport modes. However, the most significant factor for this difference is the low income among students and thereby low car ownership.

The highest share of car passengers is found in the leisure purpose group, which also has the second lowest car driver share. This is probably caused by many people driving together for leisure purposes (e.g., visiting family, sport events). Walking is also popular for leisure trips, possibly because of low distances and more time to spend on the trips because the travelling itself can be a significant part of the trip.

Most people drive by car (or vans, trucks) for business purposes, since this is often the most flexible solution when carrying goods, going to several specific addresses in a short time, etc. (European Commission, 2001).

Figure 4-4 shows the cumulative distributions of the trip distances divided by trip purpose. The graph cuts at 50 km since at this point five of the six purpose categories have reached a level of over 95%.

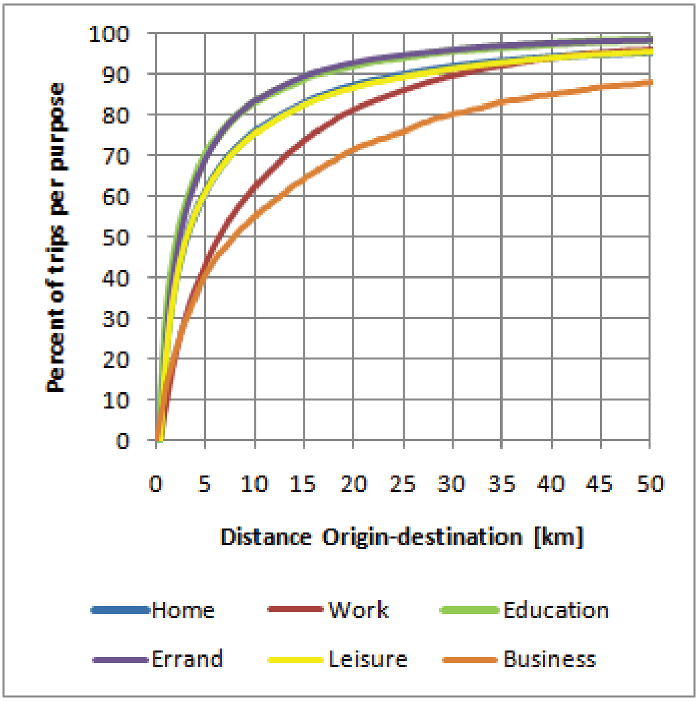

Figure 4-4: Cumulative distributions of the share of trips.

The business trips are below 90% and do not reach 95% before 137 km. Business trips have the flattest curve all along, meaning that their average trip distance is the highest.

The figure shows that 70% of the errand and education trips are shorter than 5 km. This is also the case for 60% of the home and leisure trips. The education trips are rather short since school trips are a part of this category and pupils often go to school close to their home. For higher education students, residential location (e.g., apartments for students, dorms) is placed close to the university and therefore gives some of the students a short distance to their education place. The short education trips are also in line with the high share of biking trip showed in the previous figure. People often select a destination close to the origin of their trip when going for errands, which explains the short errand trips. Shopping malls and the like also attract people, thus explaining why some of the errand trips are relatively long.

80% of the work trips are up to 20 km. The commuting distances are longer than the other trips (except business) because the choice of workplace is not as flexible as other choices (e.g., errands). People are willing to travel longer to reach a better job. The curve for work trips starts as the curve for business trips and approaches the curves for home and leisure at 35 km.

4.2.1.3 Analysis of gender differences

The women in the survey perform a higher number of trips than the men. The TU survey has more female than male participants (i.e., 28,444 versus 26,251) and the total number of trips by gender shows that women perform 3.13 trips per day while men perform 2.96 trips per day.

In Table 4-5 the average distance between origin and destination for men and women are shown. On average, men travel 2.5 km longer than women. The men also have a higher standard deviation for the distance.

Table 4-5: Average distance and standard deviation between origin and destination for men and women.

Gender	Avg. Distance [km]	St. Dev. Distance	No. observations
Men	10.74	23.02	78,099
Women	8.22	18.88	88,895

Figure 4-5 illustrates the mode shares for each gender. Car driver is the most chosen mode by both genders, with more than half of the men and 40% of the women choosing this mode. The data also show that, in households with only one car, men most often use this car and women either get a lift or choose alternatives such as walking, bicycle or public transport.

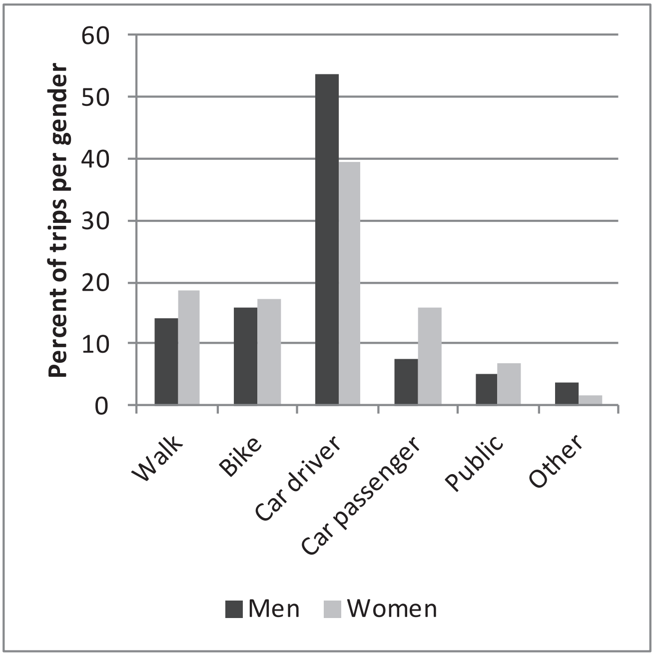

Figure 4-5: Mode shares for men and women.

With the exception of the category *other*, all remaining categories are chosen by a greater part of women rather than men. The higher share of women walking or biking could be explained by the fact that women work closer to home and therefore have a better chance of either walking or bicycling to work (also see Figure 4-6). Other explanations are that women perform tasks such as shopping, bringing and collecting children, close to home and therefore the majority of their trips are short. Also the fact that women on average have lower income than men can explain this phenomenon.

Figure 4-6 shows the percentage of travellers by gender according to the trip distance measured between origin and destination. It should be noted that the graph is for distances between origin and destination up to 15 km, since, for longer distances, shares approximate zeros for both genders.

For shorter distances, women have a higher share of trips than men. The difference is relatively rapidly decreasing and the shares for the two genders are almost equal for trips exceeding 1.5 km. For trips with distances exceeding two kilometres, men have a slightly higher share of trips.

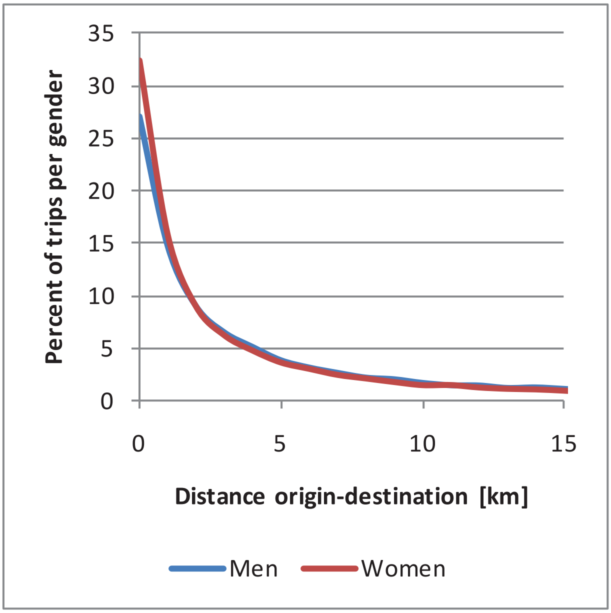

Figure 4-6: Percentage of trips for each gender.

4.2.1.4 Analysis of age differences

Figure 4-7 shows the percentage of people with a specific age choosing the different transport modes. People younger than 18 years old have a very different pattern from the rest, since they have not yet had the chance to obtain a driving license (can be obtained by the age of 18 in Denmark). The young population often travels as car passengers or by bicycle. The bicycle use is increasing up to 12 years (40%) and then decreasing to a rather constant share of 15%. The share of public transport modes is increasing up to 18 years of age, and then dropping after the possible achievement of driving license. This share increases again with age and the elderly population's share of public transport is almost as high as the young ones. The car as driver is the most often chosen mode from the age of 18 to 77 years, with a peak at 40 years of age.

Walk is increasing from the age of 40 and a very high share of the trips conducted by the elderly in the survey is by walking. The people in the middle age group are often employed and therefore have a high demand for transportation. They often can afford one or more cars and, because of the convenience of using the car, this choice is superior to the rest (Lorenc et al. 2008). The older people have more time when travelling and often feel safer when being a passenger (both public and private) than when driving (Rosenbloom, 2004). The curve is clearly turning around the point of retirement (60-67 years of age) where the demand for transportation is changing, since working trips are no longer as important and other trip patterns occur.

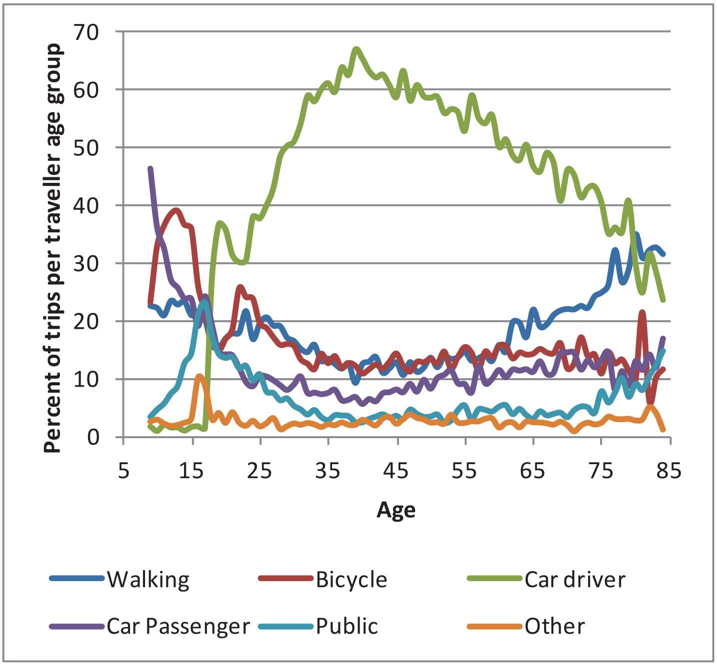

Figure 4-7: Choice of transport modes in relation to the age of the respondent.

Table 4-6 shows the average OD distances for the trips in each age group. The two age groups which are very different from the others are the youngest (9-17) and the oldest (70-85). The two groups have an average of less than 6.0-6.8 km compared to the third lowest of 9.3 km. The reason for the short trips for both groups is the decreased possibility to use the same transport modes as the other age groups. The youngest have not yet had the chance to obtain driver's licenses and among the elderly more people have lost their license, not renewed it, etc. The oldest respondents probably feel safer when walking because they can choose their own pace and have more time to assess the surroundings when walking.

Table 4-6: Average distance and standard deviation between origin and destination for age groups.

Age Group	Avg. Distance [km]	St. Dev. Distance	No. Observations
9-17	5.97	16.85	20,581
18-29	9.34	20.29	21,328
30-39	10.02	21.40	27,010
40-49	10.82	22.47	29,671
50-59	11.10	22.36	24,902
60-69	9.34	21.76	19,507
70-85	6.79	17.79	10,887

The average distance for the travellers in the age groups 18-29, 30-39 and 60-69 are not significantly different from each other at the 0.01 confidence interval, which is also the case for the groups of 40-49 and 50-59 years.

4.2.1.5 Analysis of car ownership

In Figure 4-8 the choice of transport mode is compared to the number of cars owned in the household. Logically, the use of walking, bicycle and public transport modes is greater for the people without car than people with car, since car is not always available.

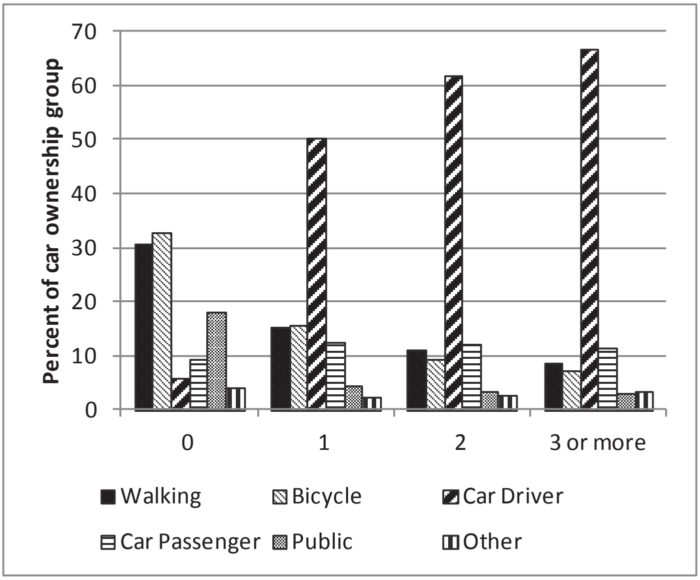

Figure 4-8: Choice of transport mode in relation to car ownership (percent of car ownership group).

The high use of walking and bicycle points to the fact that households without cars are placed close to the traveller's destination points (e.g., in a city) or that non-car owners walk and bike longer distances than car owners. The use of car as driver increases with the number of cars in the household, but car as passenger shows the same share of people with 1, 2, 3 or more cars.

Interestingly, travellers from households without cars travel as car passenger less often than people with car, pointing to the fact that car passengers often drive along with another household member.

Figure 4-9 shows the average number of cars for each household in the 98 municipalities in Denmark. In the municipalities of the Greater Copenhagen Area (east) and around the greater cities: Aarhus, Aalborg, Odense and Helsingør, the car ownership is the lowest with less than 0.9 cars per household. The people living there have good access to public transport and some do not have parking possibilities around their home, and therefore people minimise the number of cars in the household.

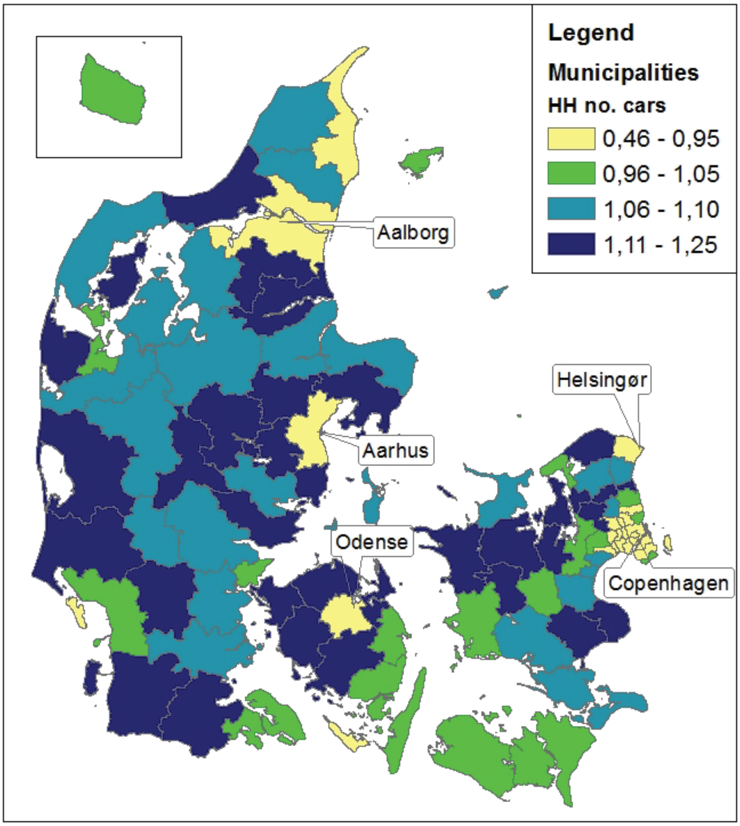

Figure 4-9: Average number of cars in households in each municipality.

In the areas around Copenhagen the households own approximately one car, likely because the members of these households have access to better public transport services than the rest of the country. The other green zones are spread around the country and most of these are in municipalities where the inhabitants earn less than the average and therefore can afford fewer cars. The municipalities with the highest number of cars per household are found with some distance from the larger cities and are most likely people who have to commute for long distances. Geographical trip distances are investigated in section 4.2.1.8.

4.2.1.6 Analysis of income differences

Figure 4-10 shows the choice of transport mode related to the total income in the household of the respondent. For low income groups, the choice of the car is very low and the choice of public transport is the highest. These groups can often not afford to have a car and therefore have to choose other transport modes. In 2009 the average income for Danish inhabitants over 15 years was 278.500 DKK so very few fall in this low income group, and the data show that it is primarily very young people. For households earning more than 100,000 DKK per year, the choice of the car as a driver is the most often selected.

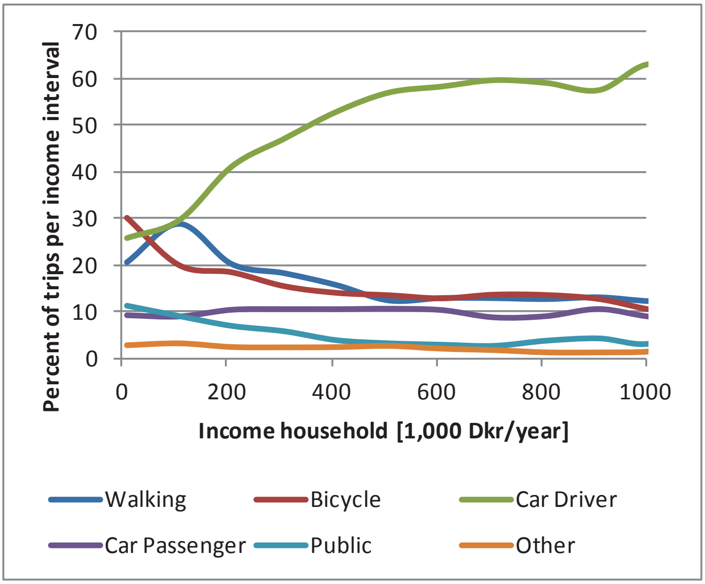

Figure 4-10: Choice of mode related to the household income of the traveller.

The share of travellers using public transport is decreasing with increasing income and is rather stable at 3-4% for respondents from households with income exceeding 400,000 DKK per year. For most of the mode choices, the curves are rather stable from 500,000 DKK per year, meaning that the income has only little effect on the mode choice when exceeding a certain amount of income.

4.2.1.7 Analysis of the distance to the nearest train station

Data on the exact location of home and work were only available for the Greater Copenhagen Area, and hence the analyses in this section are made only for this area. Figure 4-11 and Figure 4-12 show the choice of transport mode related to the distance to the nearest train station from origin purpose home and work, respectively. In these analyses, the public transport modes are split in bus and train since graphs with public transport modes as one category showed clear signs of a difference between bus and train. Also the distance to stations should explain more about the use of trains than about the use of buses. The graphs are cut at 5 km since only few people have greater distances to a train station from home or work in the Greater Copenhagen area.

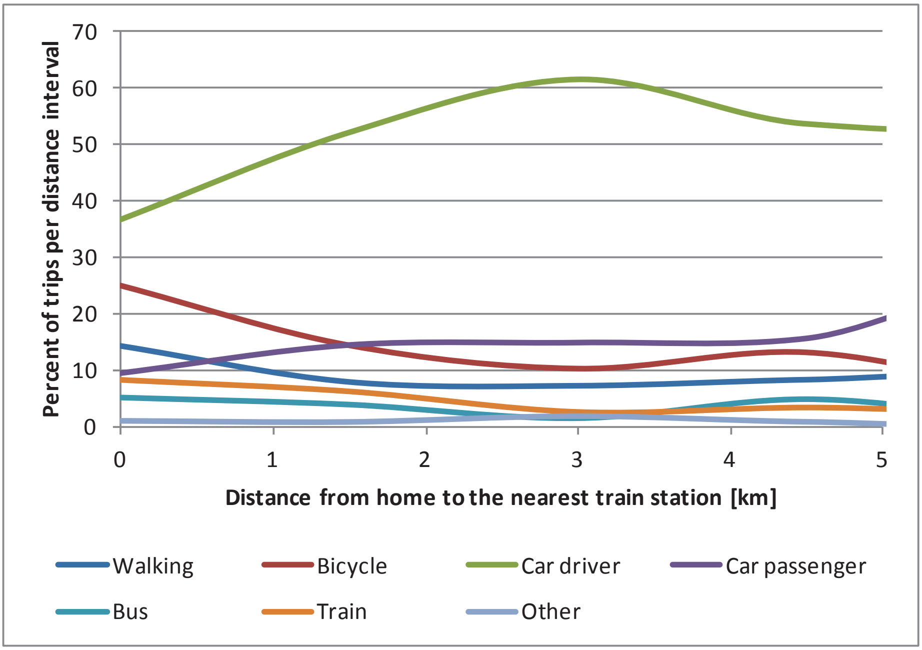

Figure 4-11: Choice of primary mode at different distances to the nearest train station from home.

At all distances to train stations, most people choose to drive by car. The use of public transport is higher at small distances. From home (Figure 4-11), the use of bus is higher at short distances to stations, but drops to half within the first 2-3 kilometres. The same applies also to train.

From work (Figure 4-12), the use of the train is rather high at small distances and the bus use is low compared to trips from home. Use of bus increases with trip distance and peaks at more than 15% with a distance of 2 km from the work location to the nearest train station.

The car as driver is chosen much more often from work than from home (this choice is also dependent on the choice the traveller did in the morning). From home travellers often have short distances to shopping and leisure facilities making walking and bicycle realistic and even attractive choices to the traveller.

Overall, the transport modes used from home is very different from the transport modes used from work. A reason for this is the different accessibility to transport modes from the two origins. Private transport modes such as car and bicycle are accessible from home, but from work only if the traveller "brought" the mode from home in the morning (or if the traveller planned ahead and placed a bicycle at work or the work end of the public transport trip).

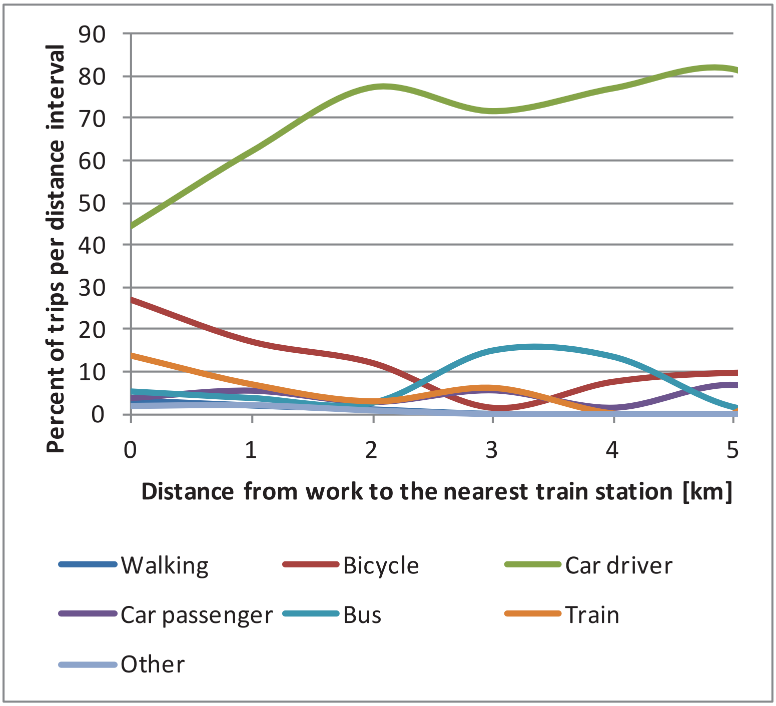

Figure 4-12: Choice of primary mode at different distances to the nearest train station from work.

The choice of transport mode is also dependent on the destination purpose of the trip and the number of trips in a trip journey (number of trips and trip destinations before again reaching the starting point of the first trip).

4.2.1.8 Analysis of geographical location and mileage

Figure 4-13, Figure 4-14 and Figure 4-15 show the number of kilometres an average person in each of the Danish municipalities travels in public transport modes, car and in total per year. Note that the scales for the figures are not the same since the mileage for cars is approximately seven times higher than for public transport.

Figure 4-13 shows that the people in Zealand travel more often with public transport than the people in the rest of the country. Especially inhabitants of the Greater Copenhagen Area, as described in Figure 4-9, also often do not have a car and therefore travel fewer car kilometres. People from northern Zealand and around Roskilde also travel many kilometres with public transport. This can be caused by the fact that they have a relatively long distance commuting to Copenhagen and the train services between the cities and Copenhagen is rather good, so they often choose public transport modes instead of driving cars on the congested main roads of Copenhagen. In almost every zone in Jutland people are using public transport for less than 1,500 km/person/year because of the less good train services and the higher car ownership.

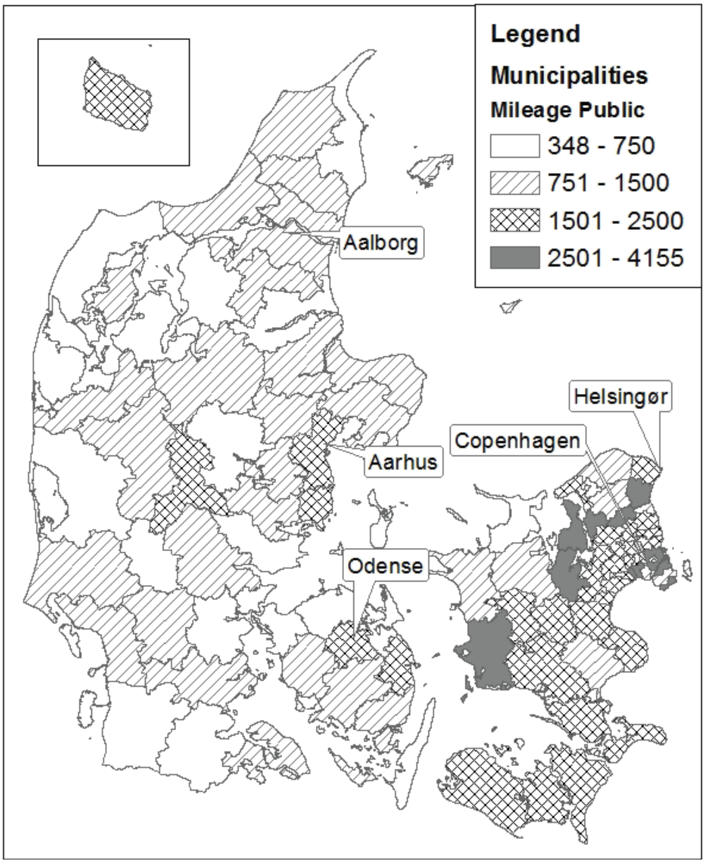

Figure 4-13: Mileage travelled in public transport modes for inhabitants in each Danish municipality [km/person/year].

Figure 4-14 shows the number of kilometres per person with a car. The smallest scale is up to 7,500 km/person/year and only the municipalities closest to Copenhagen, the area around Aalborg and some of the islands fall in this category. These areas either have a good public transport service or a travel pattern with short commuting distances. From the south-west part of Zealand people drive long distances by car, and many people from this region commute to Copenhagen or live far from the nearest larger city.

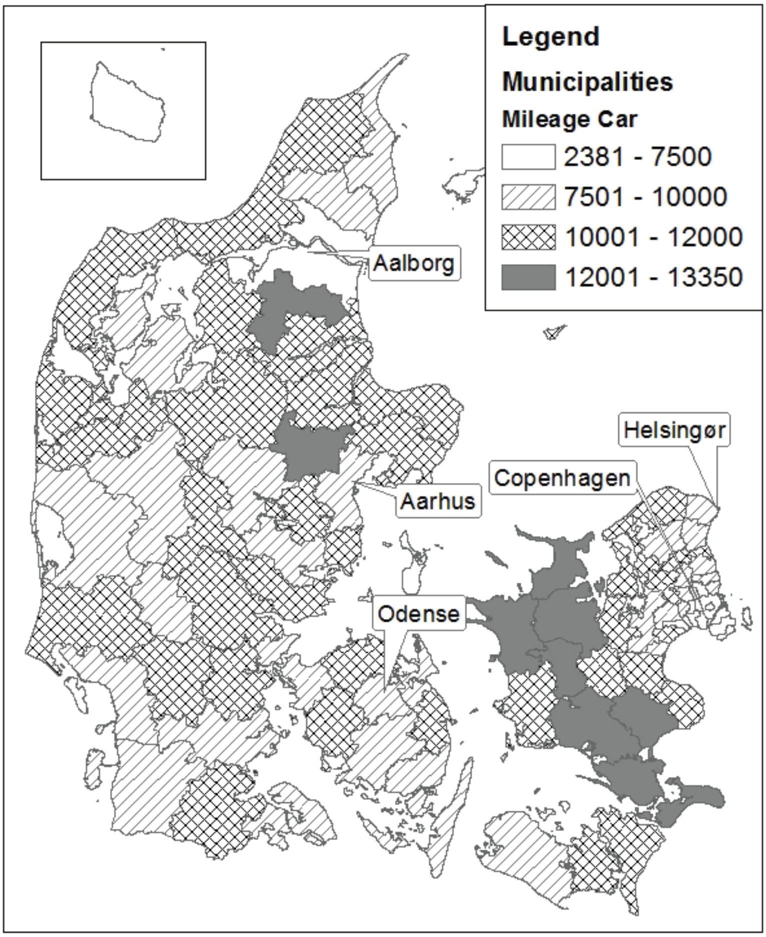

Figure 4-14: Mileage travelled in car for inhabitants in each Danish municipality [km/person/year].

Figure 4-15 shows the total distance travelled for inhabitants in the 98 Danish municipalities. The greatest total distances are found at most of Zealand (the Greater Copenhagen Area excepted). Often people living in this area work in Copenhagen and therefore have a large commuting distance (>13,000 km/year).

The inhabitants living close to the centre of Copenhagen travel the shortest yearly distances.

The maps show that people living on larger islands with no bridges to the mainland have the shortest yearly travel distances (shorter commuting distances). The people living in the outskirts of the main capital often have jobs close to Copenhagen and therefore commute long distances resulting in a large yearly travelled distance.

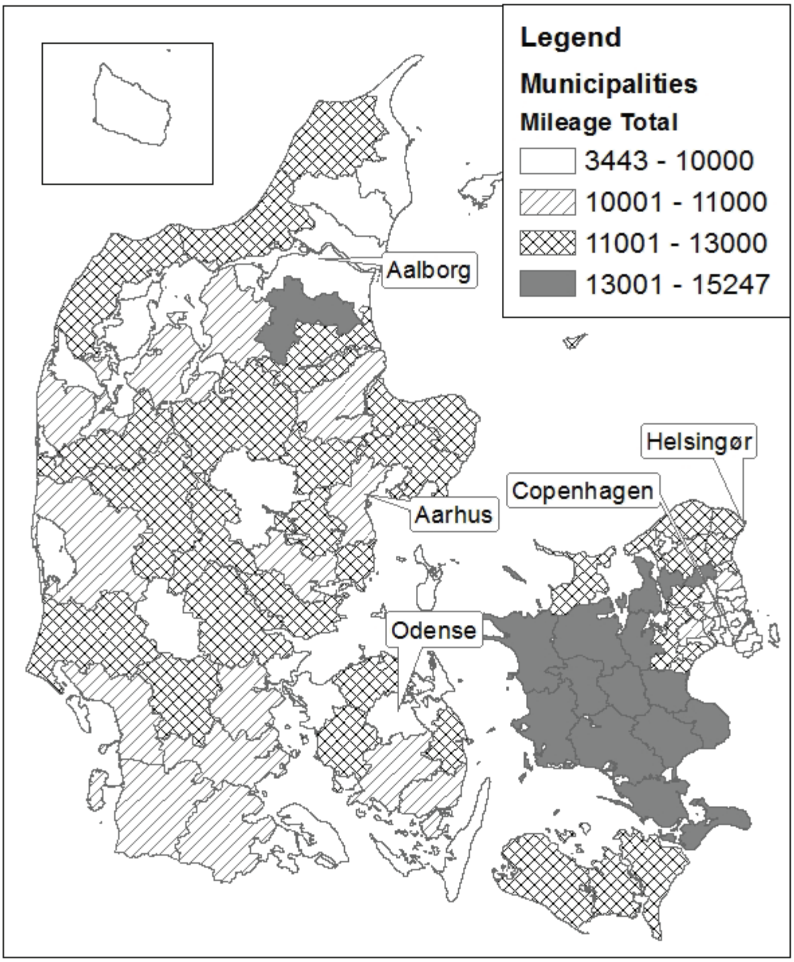

Figure 4-15: Mileage travelled in total for inhabitants in each Danish municipality [km/person/year].

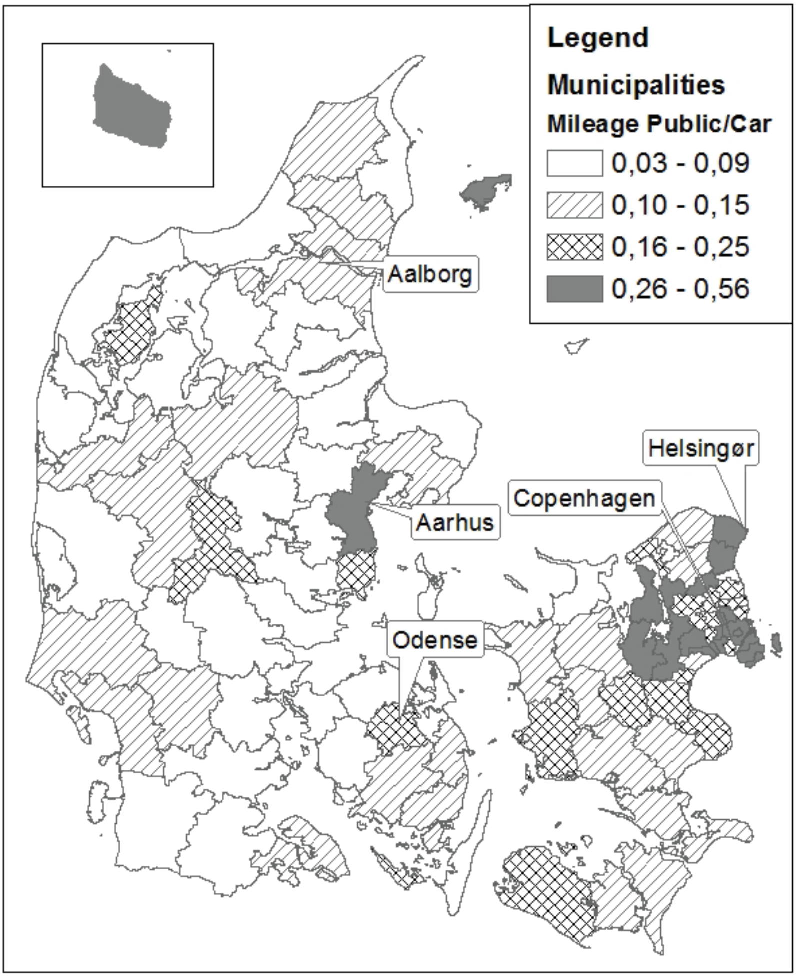

Figure 4-16: Ratio of mileage travelled in public related to private transport modes for inhabitants in each Danish municipality [km/person/year].

Figure 4-16 and Figure 4-17 show how many of the travelled kilometres are in public transport versus private transport and versus total travelled distance. The figures show that a high share of the kilometres travelled by people living in Copenhagen and Aarhus (Mid-East of Jutland) is by public transport. The areas in the centre of Copenhagen which were white (short total distances) in Figure 4-15 are the ones with the highest share of public transport use.

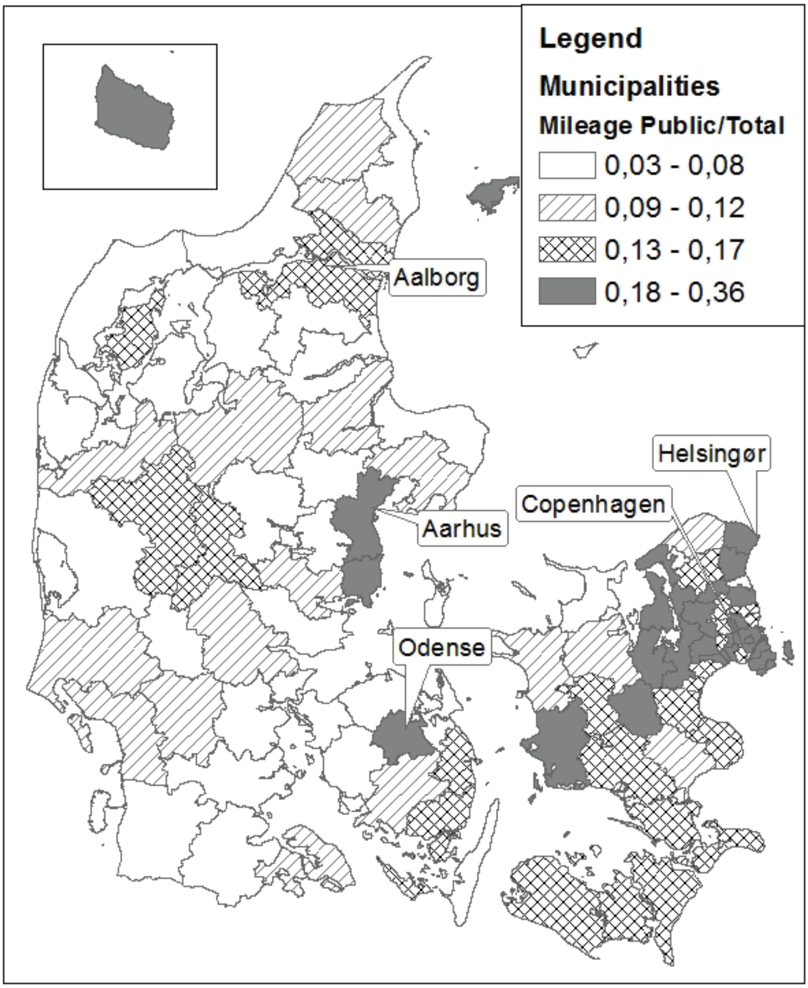

Figure 4-17: Ratio of mileage travelled in public transport modes related to total distance travelled for inhabitants in each Danish municipality [km/person/year].

4.2.2 Summary and conclusions on private and public transport trips

This section has presented and discussed different analyses from the TU survey, and different aspects which can influence the choice of mode were touched upon.

The distance between the origin and the destination influences the choices, especially between short (primarily walking and bicycle) and long (primarily car and public transport) distances. For both car and public transport, the shares increase with the length and are rather stable from five kilometres. The distance does not directly influence whether the travellers choose car or public transport.

The choice of mode varies very much for the different trip purposes. For all purpose categories but education, the most often chosen mode is the car, but for example for leisure one third drive car compared to the two thirds for business trips. Education has the lowest share of car drivers and the highest share of public transport user.

The analysis of gender differences shows that far more men than women drive cars, whereas women are more often passengers in a car than men. Women also have a higher share of public transport users than men.

The mode choice showed to be very different among the different age groups. Car driving is the highest at the age of 40 and lowest for the young and the elderly population. For public transport, the pattern is very different showing the lowest share from 35-65 years and the highest for the young and the elderly population.

The car ownership also affects the choice of transport mode. When no car is owned, the use of a car is of course very low and the use of public transport modes is the highest in this group. When owning at least one car, car use increases with the number of cars and the use of public transport decreases accordingly. The geography of car ownership shows the lowest number of cars per household around Copenhagen and the highest number at some distance from the largest cities (but not in less wealthy rural and remote areas).

The income of the respondent’s household has an effect on the mode choice, and especially the choice is very different between the lowest and highest income groups. The share using public transport is the highest for the lowest income travellers and the share decreases with increasing income, while the car driver share shows an opposite tendency. From 500,000 DKK per year, the choice is stable and additional income only affects the mode choice very little.

The highest use of public transport is for trips conducted by people living within 0-2 km from the nearest train station. With up to one kilometre from the workplace location to the nearest train station 10% of the trips are by train. With 3-4 km from work to the nearest station more than 15% of the travellers use a bus as the primary mode.

The geographical analyses of the kilometrage for car and public transport show that public transport modes often substitute car since many of the zones have a high number for one mode and a low number for the other. This is especially the case for areas around the larger cities. Farther away from these cities, people have a higher demand for transport and especially some areas in Zealand have a higher kilometrage for both car and public transport.

The examined factors are all characteristics of the traveller or the trip, and some of these can be difficult to modify in order to change the mode choice of the traveller. However, all these factors are important to be aware of when planning transportation service and when informing about these services. Even though they are hard to change the findings can be taken into account when planning the transport system. Travellers to some specific trip purposes use public transport very seldom. For some business travellers the public transport modes might not be considered an alternative since carrying goods, visiting customers at remote locations, etc. If the non-selection of public transport is caused by inconvenience of not getting a seat, missing regularity, etc. the transport authorities are in fact able to do something about this and thereby they might raise the market share of public transport users among business travellers.

More women than men use public transport. To increase the share of men choosing public transport the reasons for the men deselecting it has to be revealed and these areas can then be improved in order to attract the men.

The group of travellers from 35-65 years has low shares of public transport use. The youngest in this group often travel with children for the whole trip or delivers / collects the children at day care or school at the start or the end of the trip journey and therefore find it difficult to use public transport. Increasing attention to families with children might improve the conditions by providing higher public transport service in specific geographical areas, improving the frequency of the buses, ensuring easier access with children, etc.

The number of cars in the household is difficult to affect immediately by improving the public transport modes. But at the long run improvements in the public transport network can encourage people not to invest in the first car or not to buy an additional car, or the services can affect people's choice of place to live and work since in places with good public transport access the investment in a car might not be necessary.

The largest potential for moving traffic from car to public transport modes is found in the highest income groups and the conditions offered to these might be improved to attract these groups.

The analysis of distance to nearest train station shows that the location does affect the choice of public transport versus private transport modes. By using this in more depth when developing the urban and rural areas more people can be attracted to the public transport modes. Workplaces with many employees should be placed close to a station because this increases the possibilities of meeting the wishes from the travellers to use the transport modes most convenient to them.

The analyses have shown that many of the investigated factors do have an effect on the choice between private and public transport. The method of analysing the data can be used for further analyses of the TU data. The characteristic of the respondents and the trips can be investigated further, also going more into details about for example the public transport, service level, distance to public transport, etc.

4.3 Characteristics of public transport mode chains in the network of the Greater Copenhagen Area

In the following sections analyses concerning the choice of multimodal versus unimodal trips are conducted. The trips are defined in Chapter 2 to be unimodal when using only private modes, public multimodal when using several public transport modes and full multimodal when using both private and public transport modes. The analyses are for the Greater Copenhagen Area and the selection of data used are described in section 4.1.2.

4.3.1 Characteristics of the multimodal trips in the Greater Copenhagen Area

The majority of the multimodal trips consist of two legs (not including walking). Two-leg trips make up 63% of all the multimodal trips in the dataset. 31% of the trips have three legs and 6% have four or more legs.

Table 4-7 shows the share of all trips, unimodal, multiple-leg public and multimodal public and private trips in the Greater Copenhagen Area for which each mode is the primary mode.

Table 4-7: Modal split with the distinction between unimodal, multi- leg public and multimodal public and private trips (TU data 2006-11, Greater Copenhagen Area) – pure walking trips excluded.

Primary Mode	All trips Number	All trips [%]	Unimodal [%]	Multiple-leg public [%]	Multimodal public+private [%]	Percentage multimodal public+private
Walking	145	0.22	0.20	0.02	0.00	1.38
Bicycle	15,840	24.11	24.09	0.00	0.02	0.08
Car driver	31,794	48.38	48.33	0.00	0.05	0.10
Car passenger	8,919	13.57	13.45	0.00	0.12	0.91
Bus	3,331	5.07	3.42	1.27	0.39	7.60
S-train	2,695	4.10	1.20	1.43	1.47	35.84
Other train	1,528	2.33	0.43	0.67	1.22	52.42
Metro	787	1.20	0.61	0.41	0.18	14.87
Other	673	1.02	0.99	0.00	0.03	3.12
Totals	65,712	100.00	92.72	3.80	3.48	
Number		65,712	60,931	2,494	2,287	

The bicycle is the primary mode of 24% of all trips. 48% use a car as a driver and 14% as a passenger. The public transport modes bus, train and metro have rather low shares (1-5%) of all trips. Unimodal trips with a private mode (bicycle, car as driver or passenger) count for 86% of all trips.

Walking, bicycle, and car as a driver or passenger are rarely used as the primary mode in multimodal mode chains. S-train is the most frequently used primary mode for multimodal mode chains accounting for 1.5% of all trips. 7.6% of the bus passengers and half of the metro passengers use both private and public transport modes for the trip.

Table 4-8 shows the percentages of trips with a public transport mode as the primary mode. The unimodal share of public transport trips is 44.6%, while 29.7% use several public transport modes, and 25.6% use a combination of public and private transport modes. The bus is used for more than half of the unimodal public transport trips and the S-train is used for almost half the multimodal public and private trips.

Table 4-8: Modal split for public transport trips (public primary mode) with distinction between unimodal, multimodal public and multimodal public and private trips (TU data 2006-11, Greater Copenhagen Area).

Primary Mode	All trips Number	All trips [%]	Unimodal [%]	Multiple-leg public [%]	Multimodal public+private [%]	Percentage Multimodal public+private
Bus	3,331	39.94	26.93	9.97	3.03	24.98
S-train	2,695	32.31	9.48	11.25	11.58	34.81
Other train	1,528	18.32	3.40	5.31	9.60	28.99
Metro	787	9.44	4.83	3.20	1.40	33.93
Totals	8,341	100.00	44.65	29.73	25.62	

Figure 4-18 shows the modal share of home- and activity-based legs on multimodal trips. The figure illustrates the difference in mode availability at the home-end and the activity-end. At home various private modes are available and at the activity-end the private mode availability often depends on the mode choice made from home. The private mode used from home is often parked at the entrance to the public transport network. It is possible for travellers to bring the bicycle along in the trains and in some buses. Bicycle and car as a driver are used more than three times as often from home as from the activity. Bus is used for one third of both home and activity legs. The S-train, other train and metro are used as the first leg in multimodal trips twice as often from activity-end as from home-end. Often the public service is better at the activity-end of the trip, which is reflected in the different transport modes used at the home-ends and activity-ends of the trips.

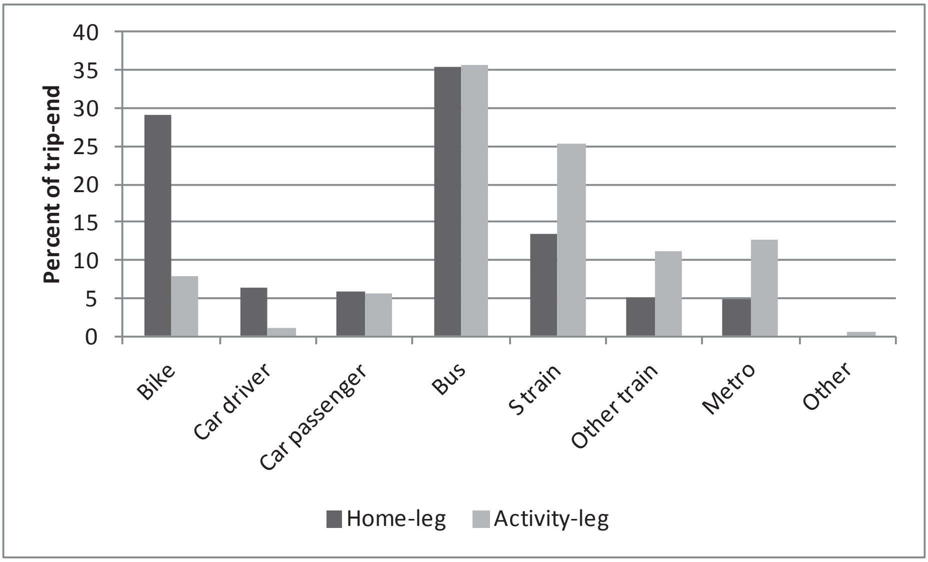

Figure 4-18: First mode used from home- and from activity-end for multimodal public transport trips.

In Appendix 2 is a table presenting the composition of the most frequently used mode chains used in the exact mode sequence by more than 20 travellers. The sample is split into home- and activity-start of trips because of the above-mentioned different availability of private modes at the home- and activity-ends. The most frequently used multiple-leg mode chain consists of two bus legs (0.8% of all trips, 10.6% of all multiple-leg trips). In fact, this combination is more often used than unimodal trips with other train or metro. The nine combinations of transport modes most often used consist of two legs, the most often used three-leg combination is Bicycle - S-train - Bicycle and the second most used three-leg combination is Bus - S-train - Bus. The trips consist of at least one public transport mode leg, but in fact 48% of the trips with at least two legs are a combination of private and public transport modes. The table shows a very clear distinction between home-end and activity-end trips, since the home-end trips more often start with a private mode and the activity-end trips more often end with one.

Table 4-9 shows the aggregated mode chains used by at least 10 travellers. S-trains and other trains are referred to as *Train*, both car drivers and passengers use *Car*, and the same transport mode type is only listed once per mode chain. The sequence of the modes (mode 1-2-3-4) is not significant for the sequence in the actual mode chain. The most frequently used aggregated mode chain consists of *Train* and *Bus* modes and the secondly most often used is *Train* and *Bicycle*. 67% of all public transport travellers with multiple-leg trips use two different transport modes (16% use one, 16% use three, and 1% use four or more).

Table 4-9: Aggregated mode chain combinations, total, from home, and from activity (mode chain used by 10 or more travellers) (Percentage of all full multimodal and public multiple-legs).

Modes used				Number of trips			Percentage		
Mode 1	**Mode 2**	**Mode 3**	**Mode 4**	**All**	**Home**	**Activity**	**All**	**Home**	**Activity**
Train	-	-	-	223	98	125	4.6	2.0	2.6
Train	Metro	-	-	301	124	177	6.3	2.6	3.7
Train	Metro	Bus	-	104	50	54	2.2	1.0	1.1
Train	Metro	Bus	Bicycle	17	10	7	0.4	0.2	0.1
Train	Metro	Car	-	79	33	46	1.6	0.7	1.0
Train	Metro	Bicycle	-	99	46	53	2.1	1.0	1.1
Train	Bus	-	-	1,033	489	544	21.5	10.2	11.3
Train	Bus	Car	-	168	71	97	3.5	1.5	2.0
Train	Bus	Bicycle	-	216	111	105	4.5	2.3	2.2
Train	Car	-	-	478	191	287	9.9	4.0	6.0
Train	Car	Other	-	10	3	7	0.2	0.1	0.1
Train	Car	Bicycle	-	39	11	28	0.8	0.2	0.6
Train	Bicycle	-	-	760	346	414	15.8	7.2	8.6
Metro	-	-	-	12	7	5	0.2	0.1	0.1
Metro	Bus	-	-	280	121	159	5.8	2.5	3.3
Metro	Bus	Bicycle	-	21	9	12	0.4	0.2	0.2
Metro	Car	-	-	36	11	25	0.7	0.2	0.5
Metro	Bicycle	-	-	65	28	37	1.4	0.6	0.8
Bus	-	-	-	534	249	285	11.1	5.2	5.9
Bus	Car	-	-	62	26	36	1.3	0.5	0.7
Bus	Bicycle	-	-	164	80	84	3.4	1.7	1.7
			Totals	4,701	2,114	2,587	98.7	44.4	54.3

4.3.2 Logistic regression model of multimodal choice

Binary logit models were used to model the choices between unimodal and multimodal trips (private and public) and between unimodal and multimodal within the public transport system. The models were made to investigate the data and make a simple model for the choice private-public transport choice which is assumed already decided upon in the following chapters. The dataset with all observed trips were used and the trip attributes and traveller characteristics were used as explanatory variables for the choice between uni- and multimodal. Two models were estimated and the input data are shown in Table 4-10.

Table 4-10: Data in the two models.

Model	Multimodal	Unimodal
All	Multimodal public and private	All other trips
Public	Multimodal public and private	All other public transport trips

The choice of choosing multimodal over unimodal was estimated and the significance of each variable to the model was assessed by comparing the statistics with a Chi-square distribution. The odds ratios were estimated to show the effect of a change in the variable to the model.

The factors important to the binary choice between unimodal (in the following, defined as unimodal and public multiple-leg trips) and multimodal trips were analysed using a logistic regression model estimated using SAS software. The binary logit model models the choice of multimodal trips and indicates which variables are important for this choice. The variables are quantitative variables and classification variables. The explanatory variables are related to the *trip* (trip distance, trip purpose and choice of primary mode), to the *journey* (distance to nearest station from journey start and city size at primary destination), and to the *traveller* (gender, income, education, and availability of private modes), as suggested by Ortúzar and Willumsen (2011).

4.3.2.1 Interpretation of the model results

The Chi-square value represents the difference in log likelihoods between fitting a model with only an intercept term and an intercept and the given variable. The statistics can be compared with a Chi-square distribution, and a resulting p-value (Pr>ChiSq) of less than 0.0001 indicates that the variable is highly significant.

The *odds ratio estimate* represents the change in odds for choosing a unimodal trip (up or down) when changing the value with one *unit* (1 or 10,000). A point estimate for a classification variable of more than one means that the odds of choosing a multimodal trip when the variable is 1 (=yes) is higher than the odds of choosing multimodal when the variable is 0.

The model was estimated both for all trips and for only public transport mode trips, and Table 4-11 shows the number of trips used for estimation of the two models. 3.5% of all trips are multimodal while 25.6% of the public transport trips are multimodal.

Table 4-11: Number of trips used for model estimation All and Public transport mode only.

Mode chain type	All	Public
Multimodal	2,215	2,078
Other	61,845	6,048
Totals	64,060	8,126

Table 4-12: Binomial logit model and odds ratio estimates modelling the choice of multimodal travel (multimodal = 1) (SAS) (0=no, 1=yes for classification variables, *ref*=reference category).

Variables	DF All (Publ)	Wald Chi-Sq All	Wald Chi-Sq Public	Pr>Chi-Sq All	Pr>Chi-Sq Public	Odds Ratio Estimate Unit	Odds Ratio Estimate All	Odds Ratio Estimate Public
Traveller Characteristics								
Number of Cars	1	28.7	42.7	<.0001	<.0001	1	1.246	1.336
City Size Home	1	6.7	9.8	0.0098	0.0018	10,000	0.998	0.998
Gender	1	21.1	20.6	<.0001	<.0001	1	1.305	1.322
Has Bicycle	1	298.3	302.7	<.0001	<.0001	1	0.222	0.191
Has Season Ticket	1	4.6	0.0	0.0324	0.9014	1	0.872	1.008
Main Occupancy								
Student	1	3.9	2.4	0.0476	0.1198	1	1.242	1.200
Working	1	8.1	5.6	0.0045	0.0176	1	1.349	1.311
Non-Working	0	.	.	.	.			
Journey Characteristics								
Distance from Journey Start to Nearest Station [km]	1	37.2	42.7	<.0001	<.0001	1	1.104	1.173
City Size Primary Target	1	2.5	0.0	0.1139	0.9729	10,000	1.001	1.000
Trip Characteristics								
Trip distance [km]	1	184.5	110.6	<.0001	<.0001	1	1.008	1.009
Trip Purpose								
Work	1	1.0	0.4	0.3290	0.5236	1	0.850	0.884
Education	1	3.9	3.9	0.0472	0.0472	1	0.686	0.652
Errand	1	40.8	25.9	<.0001	<.0001	1	0.309	0.342
Leisure	1	10.0	6.5	0.0015	0.0105	1	0.585	0.605
Business	0	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
Primary Mode						**1**		
Walk	1 (-)	0.8	.	0.3816	.	1	1.966	
Bicycle	1 (-)	39.4	.	<.0001	.	1	0.080	
Car Driver	1 (-)	58.2	.	<.0001	.	1	0.080	
Car Passenger	1 (-)	1.8	.	0.1743	.	1	0.660	
Bus	1	55.0	46.2	<.0001	<.0001	1	9.229	0.413
S-train	1	175.9	52.1	<.0001	<.0001	1	49.602	2.323
Other train	1	197.6	42.6	<.0001	<.0001	1	59.085	2.503
Metro	1 (0)	92.7	*ref*	<.0001	ref	1	20.370	*ref*
Other	0 (-)	*ref*		*ref*		*ref*	*ref*	

Table 4-12 shows the variables significant for both models. There is a distinct difference between the two models in the sense that a variable significant for one model is not necessarily significant for the other model. For example, the size of the city where the traveller's home is located (*City Size Home*) is significant for the *public transport trips* model but not for the *all-trips* model. The variables significant for both models are bicycle ownership, public transport season ticket, distance from home to nearest station, trip distance, and primary mode bus.

The ownership of bicycle and public transport season ticket both have high odds ratio estimates. Since “0” means no ownership and “1” means ownership, the travellers with a bicycle or season ticket have higher odds for choosing multimodal trips compared to unimodal. The impacts of both variables are greatest for the all-trips model.

The lowest odds ratios are for primary modes bicycle and car driver meaning that drivers using bicycles or cars for the longest trip distances have the lowest odds for travelling multimodal. These variables are only used in the all-trips model.

The distance from journey start to the nearest station has odds ratios exceeding one for both models meaning that when distance increases, the probability of using multimodal trips increases. The odds ratio is highest for the public transport trips model, which can be explained by the fact that the public travellers who live far from a railway station have to use several modes to travel from origin to station, use the train, and finally travel from station to destination. The high value for the all-trips model could be explained by people living close to stations only using one mode (often the train) and if people living far from stations use the train they have to use other modes to get to the station.

The effect of the city size is estimated by intervals of 10,000 citizens. The odds of choosing multimodal trips increase when the number of citizens in the home city increases. The odds ratio is higher for the public transport trips model than for the all-trips model.

4.3.3 Descriptive analysis of the public transport multimodal trips

In the generalised linear model analysis, six variables particularly proved to be significant in the choice of multimodal trips (highest Wald chi-square), and these were trip distance, trip purpose, primary mode, distance to nearest train station, ownership of the public transport season ticket, and ownership of a bicycle. These important factors are discussed in more detail in this section. The following analyses compare the three groups, unimodal, public multiple-legs, and multimodal public and private.

4.3.3.1 Trip distance

The impact of trip distance on the choice between unimodal and multimodal mode chains was analysed. Figure 4-19 shows the share of trips in each of the trip distance defined below. 65% of all public transport trips are in the interval of 0-20 km. For multimodal trips, 62% of the trips are between 10 and 50 km.

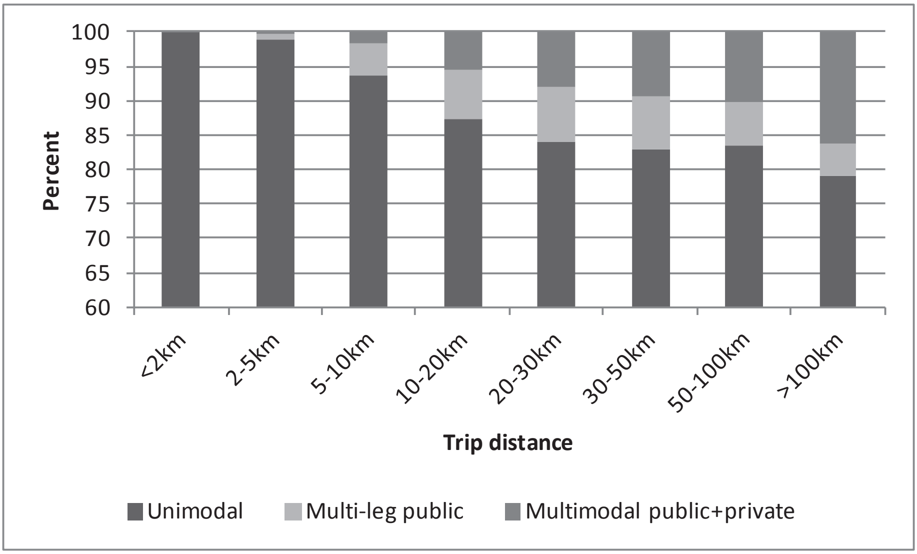

Figure 4-19: Trip distance distribution of multimodal and all public transport trips (sum to 100 for each trip distance interval –second axis 60 to 100%).

Figure 4-20 shows the average and the 10 and 90 percentile for the trip length compared to the number of trip legs (a trip of two or more legs is multimodal). The average length increases with the number of legs and so does the span of the trip lengths. 17.4% of the trips with a length exceeding 20 kilometres have more than one leg, this number increases to 19.0% for trip lengths over 50 km, and to 22.7% for trips over 100 km.

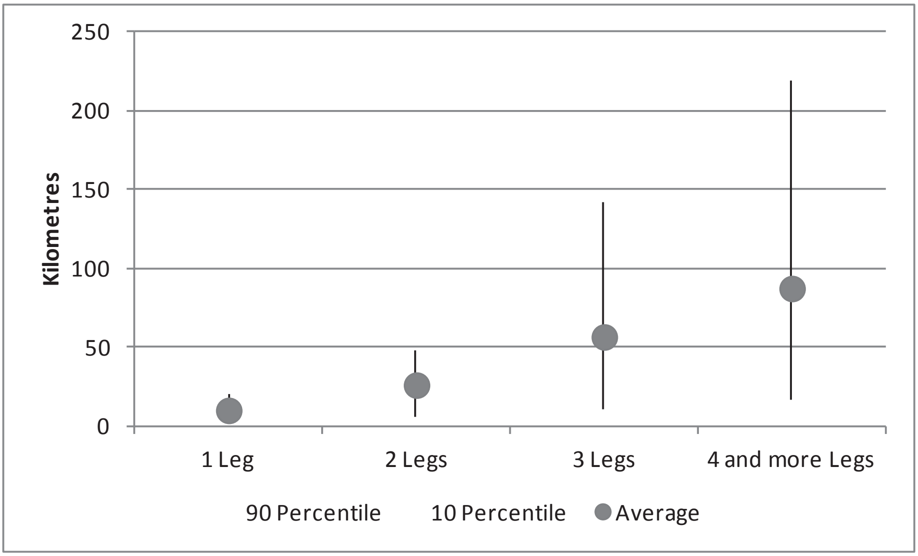

Figure 4-20: Trip distance characteristics related to the number of trip legs for public transport trips – the end points of the lines indicate the 10 and 90 percentile

4.3.3.2 Primary mode

Table 4-13 shows the access and egress modes used for trips with more than one leg in relation to the primary mode. Several access and/or egress modes could be used in a mode chain and thus the numbers do not total 100%. More than half the multiple-leg travellers using the bus as primary mode travel with at least two different bus lines. Bus and walking are among the most preferred access/egress modes to the public primary transport modes. 8% of the multimodal public travellers use the S-train as an access/egress mode to the bus. They travel the longest distance with the bus and a shorter distance with the S-train, even though the S-train is perceived as a more comfortable mode by most travellers. The explanation lies in the design of the public network in the Greater Copenhagen Area. The S-trains are radials leading in straight lines from the city centre and out, while the S-bus lines connect the S-train lines and run in rings around the city centre (see map and explanations about the network in Chapter 2).

Table 4-13: Share of access/egress mode for multiple-legs public transport trips (walking excluded).

Access/egress mode	Primary Mode			
	Bus	S-train	Other train	Metro
Bicycle	29.4	34.2	38.9	24.2
Car driver	1.2	7.4	9.4	2.9
Car passenger	9.6	9.6	24.1	3.1
Bus	54.4	41.4	40.3	48.2
S-train	7.8	14.9	13.1	31.3
Other train	2.1	4.3	8.0	1.0
Metro	5.7	18.6	5.4	4.9
Other	0.5	0.4	1.0	0.0
Totals [%]	110.7	130.9	140.1	115.6
Number	2,384	1,904	841	93

Using the car as driver or passenger is rather attractive as feeder mode to the S-train and other trains. 9% use the car as driver to other trains and 24% use the car as passenger. When driving to the train station the parking facilities at the station are important. Some train stations have large parking lot close to the station possible to use for park-and-ride. These are often found at stations with a great distance to the inner city, to encourage travellers to leave the car instead of bringing it to the most congested areas in the city.

4.3.3.3 Trip purpose

The third important factor in travelling with more than one transport mode is the trip purpose. Table 4-14 shows the shares of unimodal, public multiple-leg, and full multimodal trips for each trip purpose and the percentage of trips with a specific trip purpose that are multimodal.

Table 4-14: Share of all, unimodal and multimodal trips by trip purposes.

Trip Purpose	Total Number	All trips [%]	Unimodal trips [%]	Multiple-leg public [%]	Multimodal public+private [%]	Percentage multimodal public+private
Work	16,243	24.72	21.85	1.37	1.51	6.10
Education	4,243	6.46	5.30	0.68	0.48	1.93
Shopping	21,759	33.11	32.33	0.52	0.26	1.05
Leisure	20,871	31.76	29.54	1.16	1.07	4.32
Business	2,596	3.95	3.71	0.07	0.17	0.68
Totals	65,712	100.00	92.72	3.80	3.48	

Work has the highest share of multimodal trips with 6% of the trips being multimodal. Work travellers often have a destination (work) in a dense and well-served area in terms of public transport and an origin (home) in a less well-served area, so they use a private mode of transport at the home-end and a public mode of transport at the work-end of the trip. The multimodal share of trips to education institutions is surprisingly low compared to work trips (less than one third). One explanation is that students often live close to their education institutions and therefore only use one mode, for example a bicycle or a bus driving between the home and the education institution.

The multimodal share for leisure purposes is the second highest (>4% of all trips) because of the locations of origins and destinations and the low VoT. When visiting friends and family both origin (own home) and location (visiting home) is often placed in areas less well-served by public transport so when using public transport the travellers often use more than one. Approximately one percent of the shopping trips are multimodal, explained by the difficulty in bringing along groceries when transferring between transport modes and by the fact that shopping trips are often short trips, for example from home to a nearby grocery shop.

4.3.3.4 Public Transport Season Ticket

Table 4-15 shows the shares of people with and without a public transport season ticket who travel unimodal, public multiple-leg, and full multimodal. A large number of travellers with season tickets travel with more than one transport mode. The share of season ticket owners is higher for multimodal trips than for the others. This may indicate less knowledge of the transportation network amongst those without season tickets, causing them to travel with just one transport mode. People without season tickets also travel less often with public transport.

Table 4-15: Share of season ticket owners taking all, unimodal, public multiple-leg and full multimodal trips.

Season Ticket	Total Number	All trips [%]	Unimodal [%]	Multiple-leg public [%]	Multimodal public+private [%]	Percentage Multimodal public+private
Yes	11,805	18.0	14.2	70.9	61.3	11.9
No	53,907	82.0	85.8	29.1	38.7	1.6
Totals	65,712	100.0	100.0	100.0	100.0	

4.3.3.5 Bicycle ownership

Table 4-16 shows the choice between multimodal and other trips for owners and non-owners of bicycles. The share of bicycle owners is greatest for the multimodal trips. Actually, 90% of the multimodal travellers own a bicycle compared to 78% of the unimodal travellers, and surprisingly 70% of the public multiple-leg travellers. The bicycle is often an attractive access mode from home to train and bus, given a certain distance (see also next section) to the public service, so the bicycle is often used as a part of a multimodal public transport mode chain. For multiple-leg public the reason for selecting several public transport modes instead of a private-public transport mode combination can be the lower bicycle ownership.

Table 4-16: Share of bicycle owners taking all, unimodal, public multiple-leg and full multimodal trips.

Has bicycle	Total Number	All trips [%]	Unimodal [%]	Multiple-leg public [%]	Multimodal public+private [%]	Percentage Multimodal public+private
Yes	11,805	78.1	78.0	70.3	90.3	4.0
No	53,907	21.9	22.0	29.7	9.7	1.5
Totals	65,712	100.0	100.0	100.0	100.0	

4.3.3.6 Distance to nearest train station

Figure 4-21 shows whether travellers with some distance to the nearest train station make multimodal trips. Travellers living close to a station take multimodal trips less frequently than travellers living further away. Within a half kilometre, 2% of the public transport trips are multimodal and this rises to 3-4% for all travellers with more than 0.5 km to the nearest train station. Travellers close to stations often walk to the station and those with a distance of for example more than 1 km tend to cycle or use the bus to access the train station. The share of unimodal travel increases the farther the traveller lives from the station and the share of public multiple-leg trips decreases.

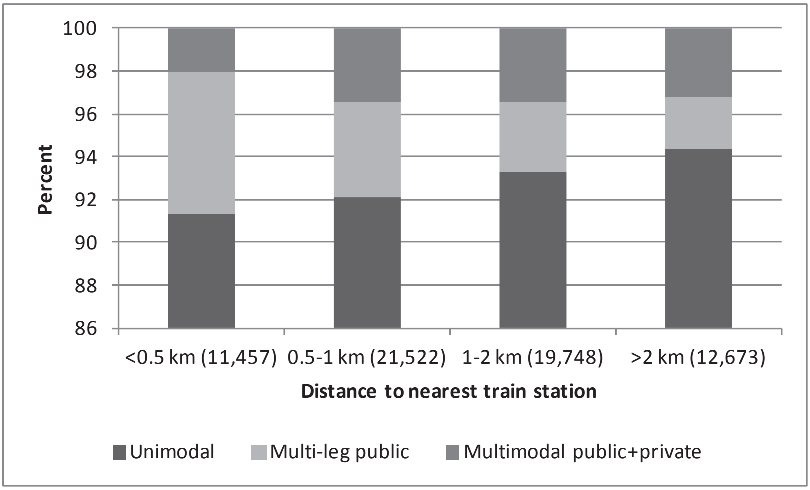

Figure 4-21: Share of unimodal and multimodal public transport trips compared to distance to nearest train station [km] (number of trips in parenthesis) (second axis cut at 86-100%).

4.3.4 Discussion and conclusion on multimodal mode chains

We see in the above sections that the share of multimodal trips increases when *trip distance* increases. The probability of having available one public transport mode that goes from door-to-door decreases with distance and results in the use of one or several access/egress modes.

Bicycle and bus are the most often used *access/egress modes* and their shares differ according to whether the trip is a home-end or activity-end trip.

Work and leisure are the *trip purposes* with the highest shares of multimodal trips (6% and 4%, respectively) and business has the lowest (0.7%).

Public transport season ticket owners more often make multimodal trips than non-owners, for example because they travel with private modes and have less knowledge of the public transport network.

Bicycle owners more often make multimodal trips than non-owners, because a bicycle is an easy integrated part of a multimodal mode chain and because non-owners have to use a car (or borrow or rent a bicycle) in order to make the mode chain multimodal with the given definition.

The *distance to the nearest train station* affects the choice of multimodal transport since the highest share of multimodal travellers has the greatest distance to the nearest train station. Vehicular access modes are more often necessary when the station is far from the point of origin.

The factors which proved to be significant in the binary choice logit model are all implicit to the choice of public transport. Work and leisure travellers use public transport more often than business travellers and therefore also use multimodal chains more often. Public transport season ticket owners by definition use public transport more than those without a season ticket and bicycle owners more often do not own a car compared to the people without a bicycle. For politicians to use these findings the important aspect is to look at possibilities of encouraging people to buy season tickets, or to purchase a bicycle. Also improving conditions for bicycle users, for example improving bicycle parking at bus stops and train stations, is an important aspect for the decision makers to be aware of. Since commuters more often conduct multimodal trips than business travellers the decision makers could focus on improving the conditions for the multimodal trips in peak hours where most commuter trips are performed to attract even more commuters to multimodal trips or improve the conditions for the business travellers to attract more of these trips. The latter is also discussed in section 4.2.2.

4.4 Test of transfer observations in TU

In order to test the travellers' perception of transfers an example from the actual network is isolated and investigated by the use of observed trips from the TU survey.

In the public transport network the travellers are offered many opportunities to transfer between transport modes in order to optimise their trip and arrive at the destination. Often the transfer is between bus and train and if a train line and a bus line follow the approximately same alignment we would assume that the travellers prefer the train and that they would rather transfer from the bus to the train if possible than stay in the bus. In most cases the train would be faster than the bus which is also attractive for the travellers. Also many travellers prefer the train over the bus because of the comfort, the driving patterns, etc. (see also focus group discussion in Appendix 1). These facts would affect the conclusions if the transfers were investigated using such an example.

In the network there are several possibilities to transfer between two train types serving the same train stations but having different routes in between. Investigating such transfers would give a more clear idea of how willing people are to transfer and how they value transfers compared to travel time.

Figure 4-22 shows the schematics of the public transport network (zoom from figure 2-4) close to the city of Copenhagen. In the area surrounded by the black circle we see that the Metro lines (yellow and green) overlap with two S-train lines (C and H – orange and red line in the figure) at Vanløse st. and Flintholm st. (west) and Nørreport st. (east).

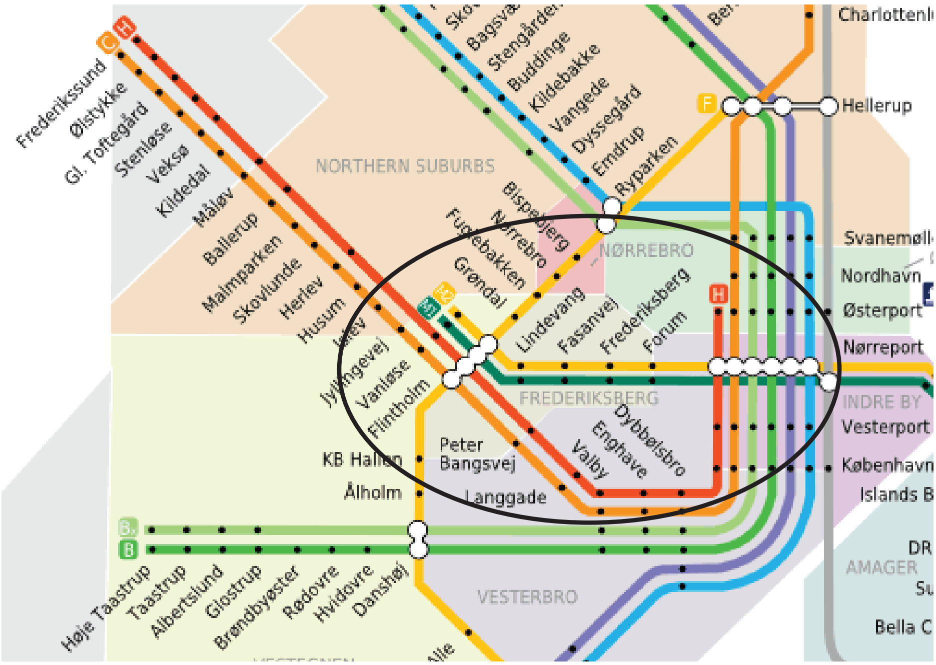

Figure 4-22: Schematics of the public transport network where S-train and metro overlap at Vanløse st, Flintholm st. and Nørreport st. but not in between.

Travelling from west to east in the morning the train route combinations have the lowest total travel time between Ballerup and Nørreport st. (incl. transfer walking and waiting time) in the following order:

- H train + Metro (25.3 min)

- H train alone (27 min)
- C train + Metro (28 min)
- C train alone (31 min)

The travel times in parenthesis are an average of the total travel times of all departures in the hours 8:00 – 9:00. We see that for all departures with S-train (either line H or C) in the morning the traveller can save total travel time by transferring to a metro train (line M1 and M2) at Vanløse or Flintholm st. In the TU survey dataset 23 respondents travel from an S-train station west of Vanløse st. to Nørreport st. in the morning. These trips are investigated and the choices of the travellers are as shown in Table 4-17.

Table 4-17: S-train and metro route combinations from S-train stations west of Vanløse st. to Nørreport st. in the morning 8-9, the travel time between Ballerup and Nørreport st. and number of travellers' choices.

Route	Travel time [min] (Ballerup->Nørreport)	No. respondents
H train + Metro	25.3	5
H train alone	27	8
C train + Metro	28	1
C train alone	31	9

The route choice also depends on the traveller's arrival time to the first train station so it is difficult to compare the H and C trains (and combinations with metro) to each other. The H alone and combined with metro is however useable for comparison since the start of the route (the S-train leg) is exactly the same. We see that the majority of the travellers chooses to stay in the S-train rather than transfer to the metro (8 of 13 and 9 of 10).

For the H train 8 of 13 do not transfer despite the possible decrease in travel time of averagely 2 minutes. This means that they would rather travel 2 minutes or more extra to avoid the transfer. The 5 respondents choose to transfer to the metro meaning that they find the transfer less burdensome than 2 minutes of travel time.

For the C train 9 of 10 choose the C train alone despite the fact that this is in average 3 minutes slower than the combination of train and metro. Only one traveller finds the transfer worth of saving 3 minutes of travel time. The nine travellers using the C train alone perceive the transfer as more burdensome than extra 3 minutes of travel time.

In the afternoon going from east to west the train route combinations are ordered as follows (the fastest first):

Table 4-18: S-train and metro route combinations from Nørreport st. to S-train stations west of Vanløse st. in the afternoon 15-16, the travel time between Nørreport and Ballerup st. and number of travellers' choices.

Route	Travel time [min] (Nørreport->Ballerup)	No. respondents
Metro + H train	27	2
H train alone	28	11
Metro + C train	28.2	4
C train alone	31.5	7

The choice pattern is rather close to the one seen in the last figure. Also going away from Copenhagen in the afternoon the majority chooses to board an S-train at the first station and stay in the train. The routes in this direction differ from the opposite direction by the fact that the traveller has to board either an S-train or a metro at the first boarding station and not transfer to the metro along the way.

The route alternatives only offer a small difference in the total travel time. Between the stations mentioned the possible travel time saving is around 10% of the total travel time. Adding access/egress time etc. to the travel time the possible travel time saving would be even less. We see that most travellers in the observed dataset choose the comfort of staying in the same train, of minimizing the risk of missing a transfer connection, of being exposed to the wind and weather etc. over the benefits of saving a few minutes of travel time.

The statements from the focus group interviews (Appendix 1) saying that some travellers only wish a save in travel time of a few minutes in order to transfer and some wish a save of 10-15 minutes are very much in line with the findings of this transfer data. A few respondents choose the lower total travel time of 2-3 minutes but the majority desire a higher travel time save in order to transfer. The investigations in the following chapters in this thesis, especially the model estimation in chapter 6, will shed more light on the actual desires concerning the transfers and put a value estimate on this.

4.5 Summary and conclusion of the chapter

In the first part of this chapter the data collected in the TU Survey has been analysed using various methods and with various objectives. Two main analyses were carried out: an analysis of the choice between private and public transport and an analysis of the choice between unimodal and multimodal trips. In the chapter the various analyses are finished with a summary and part conclusions.

Several aspects are found to be important for the choice between private and public transport modes, the following resumes the findings:

Trip purpose: the largest share of trips using public transport is found for the trips to education institutions, second most in workplaces, and the fewest for business and errand trips.

Gender: women use public transport more often than men.

Age: the car as the driver is the most often chosen mode for the majority of the age groups but for the youngest and the oldest other modes are equally important. Public transportation use peaks at 16 years, drops continuously until the age of 40 and increase to a local maximum at >the age of 80.

Car ownership: the more cars owned in a household the less frequent the household members chose to use public transport, households with no cars has the highest number of public transport trips.

Income: the use of public transport is higher for the lowest household incomes but the share is stable from 500.000 DKK/month and up.

Distance to nearest train station: the choice between private and public transport in relation to the distance to the nearest train station also depends on the trip purpose at the origin of the trip. The highest number of train trips is conducted by people living less than 2 km from a train station and the bus use is at the highest (5%) trips with less than 1 km to train station or more than 4 km.

Also for the choice between unimodal and multimodal mode chains a number of aspects have been identified:

Trip distance: for short trips (up to 5 km) almost all travellers use only one mode. From trip distances exceeding 5 km the share of multimodal trips increases with the distance. The number of multiple-leg public transport trips is rather stable from 5 km and up.

Primary, access and egress modes: bicycle and bus are the most often used access and egress modes for multimodal trips, the shares vary according to whether the trip is a home- or activity-end trip.

Trip purpose: trips to work and leisure have the highest shares of multimodal trips (6% and 4%, respectively) and business has the lowest (0.7%).

Public transport season ticket ownership: owners of public transport season ticket more often make multimodal trips than non-owners.

Bicycle ownership: bicycle owners more often make multimodal trips than non-owners.

Distance to nearest train station: the distance to the nearest train station affects the choice of multimodal transport since the highest share of multimodal travellers has the greatest distance to the nearest train station.

In the second part of the chapter the effect of the various trip characteristics and socioeconomic variables were investigated for an *all-trips* and a *public transport trips* model. In a logistic regression model six effects were identified as significant in the choice of multimodal travel. The effects are *trip distance*, *primary mode*, *trip purpose*, *public transport season ticket ownership*, *bicycle ownership*, and *distance to nearest train station*. The six effects were examined further using three groups of trips: *unimodal*, *public multiple-leg*, and *full multimodal trips*. The mode chains consist either of only public transport modes (e.g., bus as access/egress mode to/from train) or of a mix of public and private transport modes (e.g., bicycle as access to train). For the public transport trips, one fourth of all travellers use a multimodal mode chain.

The share of multimodal trips increases when *trip distance* increases. The chances of having one public transport mode that goes from door-to-door decreases with distance and results in the use of one or several access/egress modes.

Bicycle and bus are the most often used *access/egress modes* and their shares differ according to whether the trip is a home-end or activity-end trip.

Work and leisure are the *trip purposes* with the highest shares of multimodal trips (6% and 4%, respectively) and business has the lowest (0.7%).

Public transport season ticket owners more often make multimodal trips than non-owners, for example because they travel with private modes and have less knowledge of the public transport network.

Bicycle owners more often make multimodal trips than non-owners, because a bicycle is an easy integrated part of a multimodal mode chain and because non-owners have to use a car (or borrow or rent a bicycle) in order to make the mode chain multimodal with the given definition.

The *distance to the nearest train station* affects the choice of multimodal transport since the highest share of multimodal travellers has the greatest distance to the nearest train station. Vehicular access modes are more often necessary when the station is far from the point of origin.

The findings in this chapter of which factors affect route choice decisions could be used to guide stakeholders (local governmental agencies and public transport agencies) toward effective improvement of public transport services in metropolitan areas and increase the attractiveness with respect to the car.

We see that work and leisure travellers use public transport more often than business travellers and therefore also use multimodal chains more often.

Public transport season ticket owners by definition use public transport more than those without a season ticket and bicycle owners more often do not own a car compared to the people without a bicycle. Through encouraging of people to buy season tickets or to purchase a bicycle more users could be attracted to the public transport network. Also improving conditions for bicycle users, for example improve bicycle parking at bus stops and train stations, is an important aspect for the decision makers to be aware of.

A special effort could be put into improving opportunity of making seamless multimodal trips to attract more travellers to public transport. The data show that commuters more often conduct multimodal trips than business travellers so improving the possibility of uncomplicated multimodal trips in peak hours especially meet the requests of the commuters.

The analyses of the transfer preferences show that many travellers try to avoid transfers. The examples from the TU survey data show that up to six minutes of total travel time can be saved by transferring between two modes in the train and metro system. Most travellers actually choose not to take this transfer and instead stay in the first train they boarded. This shows that they prefer the convenience of staying seated, not having to worry about the transfer etc. over the convenience of saving 4-6 minutes of travel time. This result will be investigated further in Chapter 6 where actual estimations of the travellers' preferences of travel time, transfers and many more attributes are conducted.

5 GENERATION AND QUALITY ASSESSMENT OF ROUTE CHOICE SETS

As modelling route choice behaviour consists of generating relevant routes and estimating discrete choice models, this section focuses on the issue of choice set generation in public transport networks. Specifically, this section describes the generation of choice sets for users of the public transport system of the Greater Copenhagen Area by applying a doubly stochastic path generation algorithm and evaluating the ability to reproduce choices collected in the Danish Travel Survey, the TU Survey.

The following chapter builds on Larsen et al. (2010).

5.1 Introduction

Modelling route choice presumes that travellers maximise their utility by choosing the best option from a set of alternative routes. The key to correct estimation of route choice models and accurate prediction of traffic flows lies in the generation of a choice set including alternatives that travellers would possibly choose and excluding alternatives that travellers would never consider. Recent research has posed increasing attention toward the importance of the size and composition of choice sets in route choice modelling (see for an overview Bovy, 2009; Prato, 2009). Moreover, recent research has shown that path generation techniques have a great impact on route choice model estimates and flow predictions. Prato and Bekhor (2007) compared likelihood values of models estimated with alternative routes generated by constrained enumeration versus alternative paths created by a combination of deterministic path generation techniques. Bliemer and Bovy (2008) calculated choice probabilities and prediction abilities for different models with a synthetic network consisting of 12 alternative routes. Bekhor et al. (2008) evaluated objective function values and convergence times for different choice set sizes in the attempt to evaluate their influence on solutions to the SUE problem. Prato (2012) assessed the errors in model estimates after the same route choice model is estimated on data generated with different techniques for a synthetic network.

All these examples use small samples or synthetic networks, rather than analyse actual behaviour. Moreover, all these examples show the difficulty in finding a good quality measure for routes to be included in choice sets. Bovy (2009) argued that for estimation purposes not all relevant alternatives have to be included in the choice set since a small well-sampled choice set would provide satisfactory results. Bovy (2009) also added that for prediction purposes all relevant routes have to be in the generated choice set and that the inclusion of some unattractive routes is neither critical for the demand predictions nor the computational efficiency. Even though a relevant route may be defined as a route with a high probability of being chosen by travellers (Bovy, 2009), any objective definition of relevant route is actually missing for studies focusing on real-size networks. Accordingly, the assessment of the generated route choice sets is up to the experience and the sensitivity of the analyst rather than to objective measures of choice set quality. Hoogendoorn-Lanser (2005) listed logical, perceptual, feasibility and behavioural conditions for the alternatives in the choice set to fulfil, and assessed that choice sets should reflect travel behaviour knowledge (e.g., from observations).

As no objective definition of relevant routes has been provided in the literature, the present study utilizes actual route choices collected in a travel survey to assess the relevance of the generated choice sets. Specifically, the observed routes (i.e., the travel behaviour knowledge) should be part of the generated choice sets in order to include the most relevant routes to the travellers (i.e., the chosen ones). Also, as only limited information about route choices of public transport users is presented in the literature, the present study does not focus on synthetic environments, but evaluates the quality of generated choice sets with respect to real life choices. It should be noted that it is difficult to collect data on actual route choices in public transport networks, since a lot of information has to be provided to describe the routes. For private transport it is straightforward to use GPS devices to track routes and then map the data to a physical network (see, e.g., Jan et al., 2000; Schönfelder et al., 2002). For public transport, it is not possible to apply the same method, since relevant information about the lines used is not retrievable with these devices, signals may fall out in tunnels (metro and sections of the urban rail system), and information on the trip purpose, which is another fundamental piece of information for uncovering route choice determinants, is not retrievable automatically. These issues are discussed in Chapter 3.

This chapter analyses 4,833 observations of actual route choices in a public transport network, which have been collected by means of a detailed questionnaire gathering all relevant trip information about routes, lines and purposes. Then, with respect to the collected RP data choice sets for public transport users are generated and assessed by using a schedule-based stochastic transit assignment model described by Nielsen (2000) and Nielsen (2004a), and improved by Nielsen and Frederiksen (2006). The assessment is based on the comparison between observed and generated routes in a public network and on the measurement of the coverage, equal to the percentage of observations for which a certain overlap exists between observed and generated paths (Ramming, 2002). Moreover, the assessment is based on considerations by Fiorenzo-Catalano et al. (2004) with respect to the comparison of the generated choice sets with observed route choice sets in the multi-modal context. Two levels of comparison are considered when comparing the coverage of two routes: stop level and link level consider the percentage of route elements in A which are also found in B (Fiorenzo-Catalano et al., 2004).

5.2 Literature

5.2.1 Schedule-based transport modelling

Previously frequency based models were used to describe the public transport networks. These models have the disadvantage of assigning average passenger loads to the transport modes, not taken the heavier loads in peak periods into consideration. Within the last 10 years the schedule-based models have been introduced as alternatives to the frequency based models. In the schedule-based models each public transport line is represented with departure times from bus stops and train stations used. Hence it is possible to follow each traveller through the network and assess the traffic loads at each departure for each of the public transport modes (Nuzzolo and Crisalli, 2004). The schedule-based models are more detailed than the frequency-based but also the requirements of the network input data are higher. The GIS system is often used when working with network data in these models.

Nuzzolo and Crisalli (2004) described three main elements for which the schedule-based transport modelling requires explicit treatment:

- Temporal segmentation of origin/destination matrixes.
- Supply modelling for each single run.
- Choice set generation and assignment models.

In this chapter the third point from the list is of special interest since we are generating choice sets for use in assignment models.

5.2.2 Route choice set generation

The enumeration of alternative routes in the route choice set is an important issue in the route choice modelling. Recent literature has shown that the path generation techniques have a great impact on route choice model estimates and flow predictions.

The number of alternative routes can be very high even for a small network and especially for a large urban network. The choice set can be implicitly or explicitly considered in the stochastic user equilibrium (SUE) algorithms. In principle the implicit methods implies that all routes are considered and in the explicit method a selective approach is used since it is not possible to consider all routes. The selective approach is a heuristic which can either be deterministic or probabilistic and different methods for this are presented in the next sections.

5.2.2.1 Constrained Enumeration algorithms

The constrained enumeration techniques use a set of constraints to generate a full choice set containing all possible routes applying to the given restrictions. A full connection tree is constructed between the origin and destination of the trip by using a branching rule to process sequences of links. The built-in constraints reflect some cognitive, perceptual and behavioural requirements.

Prato and Bekhor (2006) used a number of rules to reflect the behaviour of travellers in a road network. The rules counted a restriction to prevent the route to take the traveller closer to the origin and further from the destination, a restriction to prevent loops, and a restriction to prevent unnecessary long routes from entering the choice set.

In public networks the restrictions are different as seen in Friedrich et al. (2001) who investigated the algorithm in a timetable-based network. In public networks the restrictions in the time dimension are crucial to prevent unrealistic situations with routes going back in time and travellers who change to a line which departs from a stop before the previous used line arrives at the stop. The diachronic time-location graph (Nuzzolo et al. 2001) prevents these situations.

The constrained enumeration techniques are suitable for generation of choice sets for use in model estimation of for example parameters in utility functions since the techniques do not require the parameters of the utility function to be estimated in advance. A significant drawback of the techniques is the fact that it is not possible to represent the heterogeneity in preferences of the traveller since the constraints are not varied.

5.2.2.2 Repeated deterministic shortest path

k-shortest paths

The shortest path techniques are related to the traveller's choice criteria for example minimum time use. The k-shortest path technique often generates choice sets with very similar routes and therefore often gives insufficient results (Bekhor et al., 2006).

Labelling approach

When changing the criterion of the shortest route in a systematic way a number of routes can be generated. The labelling approach was suggested by Ben-Akiva et al. (1984) and the criteria which are labelled are the shortest route, the fastest route, the route that maximizes the use of motorway, etc. Ben-Akiva et al. (1984) found the choice by combining the different criteria and six labels generated approximately 90% of the routes.

Link elimination

In the link elimination technique one or more links are removed from the shortest path and the new shortest path is generated. Azevedo et al. (1993) described an algorithm removing all the links used on the shortest path. This approach is problematic in the disconnection of the network since other possible routes have to be completely different from the first shortest route.

Link penalty

Instead of removing the links as in the link elimination approach links used on the shortest route are assigned with a penalty before generating the new shortest route which reduces the problems of a disconnected network. De la Barra et al. (1993) used a heuristic method to find the shortest paths and find the next shortest route after putting a penalty on the links used in the previously shortest path.

5.2.2.3 Repeated stochastic shortest path

The repeated stochastic shortest path models are based on Sheffi and Powell (1982) and are basically repeated shortest path algorithms.

Monte Carlo simulation

Sheffi and Powell (1982) described the use of the Monte Carlo technique to traffic assignment with a Multinomial Probit model. The link impedance is assumed to follow a distribution describing the traveller's different perception of the cost. The traveller does not have full knowledge of the network and the deviation from the real cost is represented by drawing the cost from a distribution which often has the real cost as the average.

Sheffi (1985) suggested a MNL (error term Gumbell distributed) and an MNP model (error term normal distributed) using Monte Carlo simulation. In MNP variances are drawn at the link level and the variance is often proportional with the deterministic cost on the link piece (travel time in Sheffi, 1985). When drawings from the distributions are carried out at link level the MNL model is not sufficient when aggregating to route level since the sum of Gumbell distributions is not Gumbell distributed. The normal distribution in the MNP model is additive and therefore useful when drawing on link level.

The costs of the links should be positive or zero but the normal distribution does not fulfil this. To solve this Prato and Bekhor (2006) have used a truncated normal distribution but this distribution does not sum to a truncated normal distribution at route level. Nielsen (2004a) suggested a gamma distribution which is additive and limited to avoid negative drawings. Also the sum of the drawings is similar to the sum of the drawings from a normal distribution at a high number of drawings.

The Monte Carlo simulation simulates the link impedance using one iteration. The modeller has to decide on the size of the β parameter and the number of iterations which are crucial for the quality of the generated choice set. A too small β parameter will cause to small variation in the generated route and a too large parameter can cause unrealistic routes. The number of iterations does not directly give the number of routes in the choice set since the same routes can be generated multiple times.

Doubly Stochastic

Nielsen (2000) introduced the doubly stochastic method which is a modification of Sheffi and Powell (1982) and Sheffi (1985). The doubly stochastic method assumes that not only the perceived cost of the network attributes but also the values of the attributes vary because of the travellers' different preferences. In the model also the parameters are drawn from a distribution for each traveller category (as opposed to error terms at the link level). The variation of the parameter is assumed to be proportional to the size of the parameter value and therefore the variation of the parameter value is proportional to the average value.

Bovy and Fiorenzo-Catalano (2007) suggested a trip utility function as a basis for the doubly stochastic function. Network attributes and attribute preferences were stochastically varied to create an optimal search to generate relevant routes.

Prato (2009) described the advantages of the model to be the heterogeneity of the generated alternatives, the relevance of the routes with observed choices, and the computational efficiency in large networks. The disadvantages lie in the calibration of the probability function coefficients since observed choice sets are difficult to collect and use of incorrect values will most likely give unrealistic and irrelevant routes.

5.2.2.4 Probabilistic methods

The probabilistic methods use a probability for each route being generated. Manski (1977) showed that it is not possible to generate all possible alternatives in the choice set, since this number is exhaustive.

Cascetta and Papola (2001) used a simpler method. Instead of assigning a probability of being chosen to all alternatives a probability of being in the final choice set was assigned.

Frejinger (2007) and Frejinger et al. (2009) used shortest path to compute a choice probability for all links in the network. Relevant routes were identified using a random walk in the network. The probability of the route is calculated at destination as the sum of the probabilities of the links of the route. The method is only tested in small synthetic networks and the applicability for real life network is thought to be limited because the random walk algorithm impose unrealistically long routes (also with loops).

5.2.2.5 Choice set generation for multimodal networks

In the literature, only limited research on the generation of route choice sets in multimodal networks is presented.

Lozano and Storchi (2001) used shortest-path searches with path composition constraints to generate routes in a multimodal transit network by creating paths which were feasible with respect to a set of constraints on the paths and a maximum number of transfers.

Bielli et al. (2006) generated choice sets in a multimodal network with private and public transport modes by ranking shortest paths using a k-multimodal shortest paths algorithm by considering the viable path (order of the modes used), time constraints and number of transfers.

By considering delays at mode and link switching points, Ziliaskopoulos and Wardell (2000) computed an intermodal time-dependent least-time path algorithm for a multimodal transportation (transit and freight) network.

Horn (2003) assumed travellers to travel at minimal generalised cost and used Dijkstra's label-setting shortest path algorithm to generate alternatives in a multimodal network (walking, fixed-route public transport, and demand responsive modes such as taxis) by accounting for different cost functions. Florian (2004) also used a labelling shortest path algorithm and introduced the concept of event dominance between two events at the same network element (link or node) which is used to reduce the number of alternatives. Event dominance is defined in terms of time and utility and implemented in terms of time and cost by Florian (2004).

The constrained enumeration technique (branch and bound) for public transport network (Friedrich et al. 2001) was modified by Hoogendoorn-Lanser (2005) who established an implementation for a multimodal network by using constraints suitable for multimodal networks. Choice sets generated with the constrained enumeration technique may be very large containing many irrelevant routes and therefore Hoogendoorn-Lanser (2005) suggested a filtering step. In this step constraints were added to ensure the quality of each route, spatial and functional variety within the choice set and other constraints.

In mixed private and public transport networks Abdelghany and Mahmassani (2001) applied a k-shortest path algorithm to generate choice sets in a multimodal network and Abdelghany and Mahmassani (1999) and Abdelghany (2001) generated choice sets using the labelling approach and the multi-objective shortest path search. Benjamins et al. (2002) generated choice sets for a mixed private and public transport network by using simulation methods. Network attributes were drawn from distributions and travellers divided into user classes to account for differences in preferences and perception of the network. These methods are route-based and an alternative route is found in a network constructed by connecting the unimodal networks from paths, roads, rails, etc.

5.3 Methods

5.3.1 Schedule-based stochastic transit assignment model

The method implemented for path generation is a schedule-based stochastic transit assignment model based on MSA proposed by Nielsen (2000), improved by Nielsen (2004a), and further refined by Nielsen and Frederiksen (2006). The model is a probit-based in order to account for the overlap between alternative routes, and is doubly stochastic in order to account for heterogeneity in both the perceived costs and the perception of the link impedance. The model has the following properties:

- Accounts for overlap across routes in order for correlated alternatives not to be considered independent.
- Considers relevant alternatives in order for possible relevant routes not to be sorted out before calculating the distribution of traffic on routes.
- Is based on stochastic utility theory, and therefore on estimated utility functions from survey interviews, as well as on well-tested theoretical foundations regarding travel behaviour.
- Describes differences in passengers' preferences through error components on the utility functions' coefficients.
- Handles timetables at a level of detail where each run is considered.

The method uses utility functions that are estimated in the East Denmark Model (Nielsen et al., 2001) in three steps:

- A nested logit is estimated due to the number of alternatives and the common features among public transport modes (see equation 1 with error term ε).
- Error components are added to the utility functions and the model is estimated as an error component model.
- The model is transformed to a time error component model to allow for variation between time coefficients (see equation 2).

Error Components are included within the utility function in order to express different preferences across different individuals:

$$U_i = \sum_j \left(\beta_j + \xi_j\right) t_{ji} + \varepsilon_i \tag{5-1}$$

where U_i is the utility of alternative i, t_{ji} are time components of alternative i as perceived by individual j, β_j is the coefficient expressing individual preferences for a certain time component, ξ_j is the error component expressing variation in these individual preferences, and ε_i is a Gamma distributed error term for alternative i. It should be noted that the utility of any alternative path is given by the sum of the utility of its links, and hence equation (1) is valid at the link level.

Entering the details of the model, for each link it is possible to write the coefficients for a certain time component and their variation as values of the different time components for the individuals:

$$\begin{aligned} U_i = {} & \beta_{walktime} \cdot TT_{walktime,i} + \beta_{waittime} \cdot TT_{waittime,i} + \beta_{changepen} \cdot N_{changes,i} + \\ & \beta_{conntime} \cdot TT_{conntime,i} + \beta_{waitzone} \cdot TT_{waitzone,i} + \beta_{IVT,train} \cdot TT_{IVT,train,i} + \\ & \beta_{IVT,ICtrain} \cdot TT_{IVT,ICtrain,i} + \beta_{IVT,STrain} \cdot TT_{IVT,STrain,i} + \beta_{IVT,bus} \cdot TT_{IVT,bus,i} + \\ & \beta_{IVT,Metro} \cdot TT_{IVT,Metro,i} + \varepsilon_i \end{aligned} \tag{5-2}$$

where $\beta_{walktime}$ is the value of walking time to the public transport mode (i.e., train, S-train, metro, bus), $\beta_{waittime}$ is the value of time for waiting the public transport vehicle, $\beta_{changepen}$ is the penalty for change, $\beta_{conntime}$ is the value of time spent on the connector between zone and public transport stop/station, $\beta_{waitzone}$ is the value of time for the waiting associated with the zone, $\beta_{IVT,train}$ is the value of in-vehicle time for regional and local trains, $\beta_{IVT,IC\text{-}train}$ is the value of in-vehicle time for IC-trains, $\beta_{IVT,S\text{-}train}$ is the value of in-vehicle time for S-trains, $\beta_{IVT,bus}$ is the value of in-vehicle time for buses, and $\beta_{IVT,metro}$ is the value of in-vehicle time for the metro.

The solution algorithm of the assignment model is through MSA (Nielsen and Frederiksen, 2006), where the stochastic part of the utility function is simulated through a Monte Carlo simulation.

5.3.2 Path generation

The route choice sets are generated through the randomisation of the VOT-terms in equation (5-2) for each link, and summing over links to obtain routes.

This chapter considers three different formulations for the utility function in order to test the effect of the variation of the different time components. Moreover, all three formulations for the utility function are tested with six values for the variances of the distributions of both error components and error term.

Referring to equation (2), the three formulations are defined as follows:

- *ErrTermOnly*: the β's are fixed (i.e., not randomly distributed across the population), and only the ε_i is randomly distributed according to a Gamma distribution.
- *ErrCompAll:* all β's are randomly distributed according to a Log-Normal distribution, and the ε_i is not considered in the utility function (no variation on the error term).
- *ErrCompErrTerm*: all the β's are distributed across the population according to a Log-Normal distribution, and the ε_i is randomly distributed according to a Gamma distribution.

These formulations cover from the simple stochastic generation technique with different variations on error terms, error components on in-vehicle travel times, error components on time components other than in-vehicle ones, to the doubly stochastic generation technique where every single term is distributed across the population. For each of the utility function formulations, the error term is considered with six different values of the scale parameter (0.025, 0.05, 0.1, 0.2, 0.5, 1.0), and the VOT-terms are examined with three difference variance scale parameters (0.05, 0.10, 0.15, 0.2, 0.5, 1.0).

5.3.3 Comparison between choice sets and observed routes

Routes are generated by simulating the utility functions for 200 times for each of the six scenarios (i.e., one for each scale parameter value) of the three formulations. The routes are produced as sequences of links within the network and are then compared to the observed routes.

For each generated route *r*, the overlap O_r with the observed route of individual *j* is measured as follows (Ramming, 2002):

$$O_r = \frac{L_{jr}}{L_j} \tag{5-3}$$

where L_{jr} is the percentage of overlapping attributes between *r* and *j* and L_j is the sum of the attributes of the observed route. The attributes that are compared correspond to the link and the stop elements. On link level the elements correspond to the public transport lines which means that only when the same line is used for both the observed and the generated route a match is found (on LineVariantElements level, see Section 5.3.3.2). On the stop level, the lines using the same links and stops are considered to be matching elements.

Considering overlap thresholds from 0% to 100%, it is possible to calculate the coverage relative to each threshold by counting the number of observations for which the generated route *r* overlaps for at least the overlap threshold with the observed route. The coverage for each overlap threshold is equal to this number divided by the number of total observations.

5.4 Results

The present study examines the collected routes in order to test the effectiveness of the doubly stochastic path generation algorithm. A total of 4,833 observed routes has been examined under the 18 scenarios (i.e., the three combinations of parameters, each with six different scale parameters), and 200 iterations are calculated for each scenario.

The variation of the coverage as a function of the overlap threshold is presented in Figure 5-1 at the link level. Considering the definition of coverage and overlap, an observed route is perfectly reproduced for 100% overlap, and the coverage for this overlap value expresses the percentage of observations perfectly reproduced. It is evident that this value is low for the ErrTermOnly scenario, where it is only around 58% for the lowest variance and 75.3% for the highest, while it is higher for the ErrCompErrTerm scenario, with 66.4% for the lowest and 77.7% for the highest values of the variance of the error components and error term. For the ErrCompAll scenario, the coverage is higher for the lowest variances (0.1 and below) and lower for the high variances (0.2, 0.5, 1.0). The increase of the variance does not have as high effect to the ErrCompAll as to the ErrTermOnly formulation.

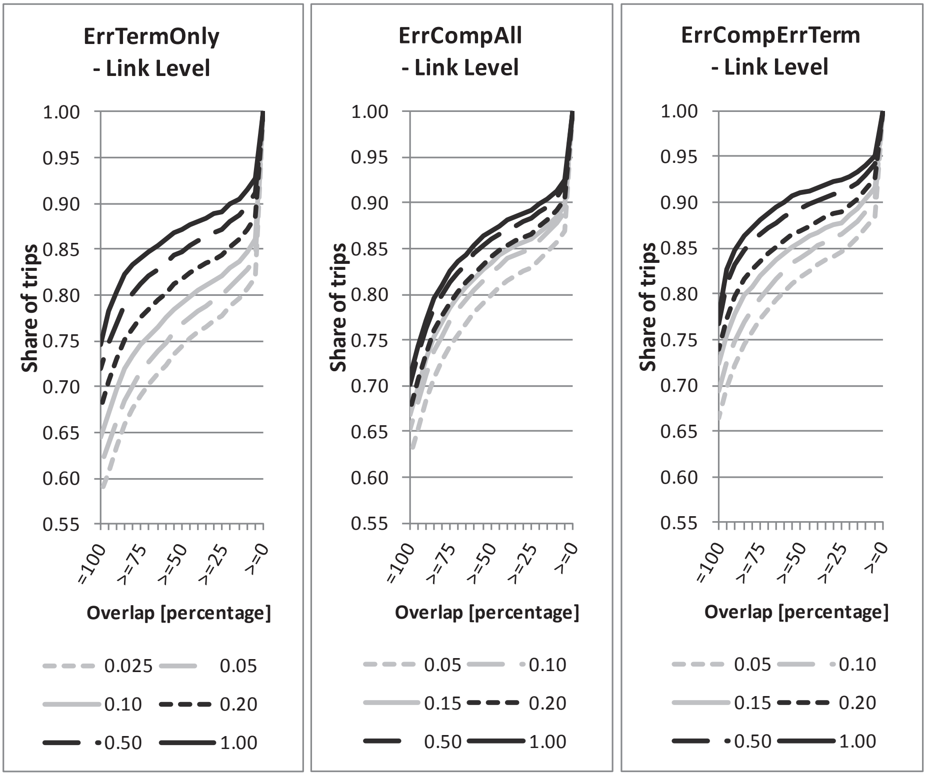

Figure 5-1: Coverage for the three formulations – calculated at link level.

These results suggest that at low variances the coverage becomes higher when the VOT-terms are randomly distributed across the population with a higher variance relative to the time components, and in particular to the in-vehicle time components. At the low scale parameter values both the error term and the error components contribute to the reproduction of the observed routes. At higher variances the error term contributes most to the reproduction.

Figure 5-2 shows the coverage for the highest variance for each scenario and the order of the coverage shows to be ErrCompErrTerm and ErrTermOnly outperforming the ErrCompAll for the highest coverages. At low coverages the ErrCompAll performs equally as well as the ErrComTerm.

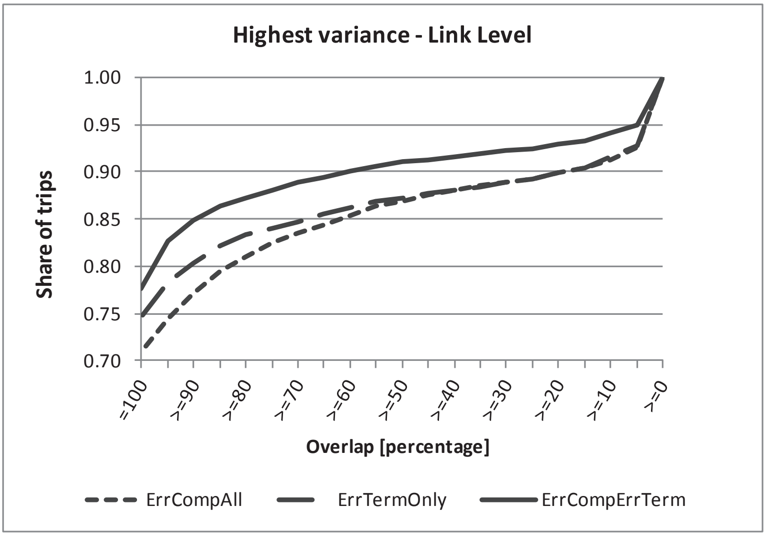

Figure 5-2: Coverage for link level, highest variance for each scenario.

Figure 5-3 shows the coverage at the link level for the three scenarios for variances equal to 0.2. At high coverages (90-100%) the ErrTermOnly and the ErrCompAll scenarios perform equally well and for all coverages below 60% the coverage for the ErrCompAll is constant at 2 percentage points better than the ErrTermOnly scenario.

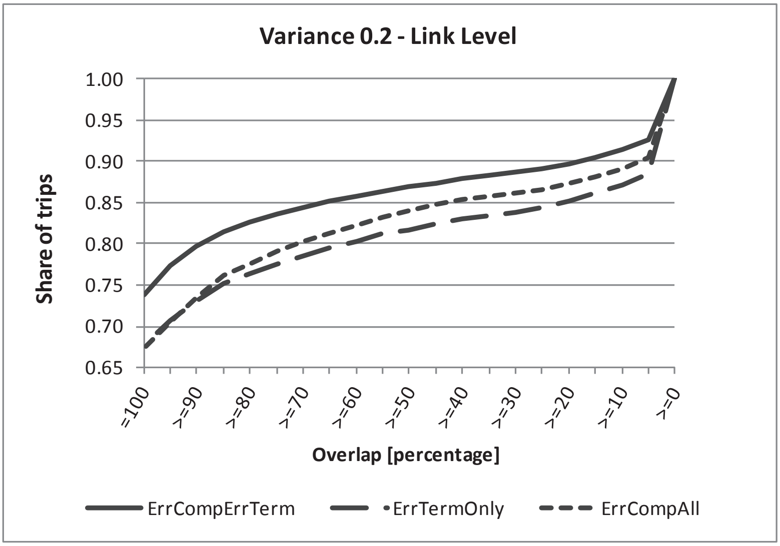

Figure 5-3: Coverage for link level, variance 0.2 for each scenario.

When relaxing the overlap threshold, the coverage is obviously higher for all six scenarios and Figure 5-4 shows the coverage calculated at the stop level. A similar interpretation of the results can be drawn from these graphs, even though with better coverage values since all scenarios perform the same or better for overlap measured at the stop level rather than at the link level. In fact, some trains and bus lines share exactly the same level of service and run on the same links, but have different names and hence are not considered to be the same in the link level coverage. This parallels the well-known blue bus – red bus problem and leads to the consideration that the stop level coverage is more precise because it considers the overlap between lines.

Overall, the curves appear shifted toward indicating better coverage for roughly 8-15% and the trend described still holds, as the lowest variation of the VOT-terms for the in-vehicle time are the ones contributing the most to generating routes similar to the observed ones and the ErrTermOnly at high variances. For an overlap threshold of 80%, a value often considered in the literature as a good limit to assess the behavioural consistency of an algorithm (e.g., Ramming, 2002; Prato and Bekhor, 2007), the coverage is over 99% for the doubly stochastic ErrCompErrTerm scenarios, while it is over 80% in the traditional stochastic generation approach in the ErrTermOnly scenarios. Again, at high variances, most of the coverage seems to be reached because of the contribution of the error terms but at low variances the error components for the value of in-vehicle time contribute equally as much as the error term variance.

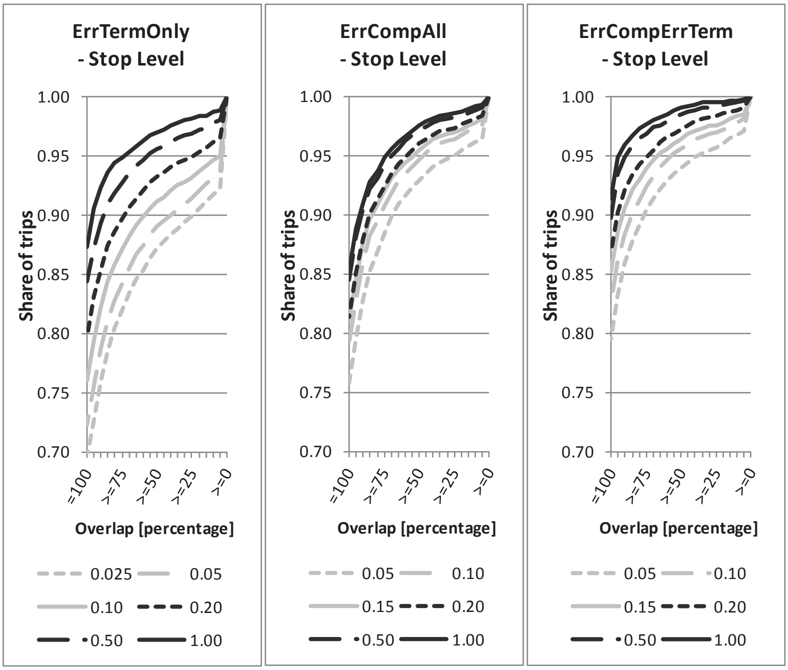

Figure 5-4: Coverage for the three formulations – calculated at stop level.

Figure 5-5 shows the coverage for the highest variance for the six scenarios. On stop level the coverage for the ErrCompAll is almost exactly equal to the ErrTermOnly and for coverage under 70% the ErrCompAll performs better. The ErrCompErrTerm is by far the best performing scenario at all coverage levels.

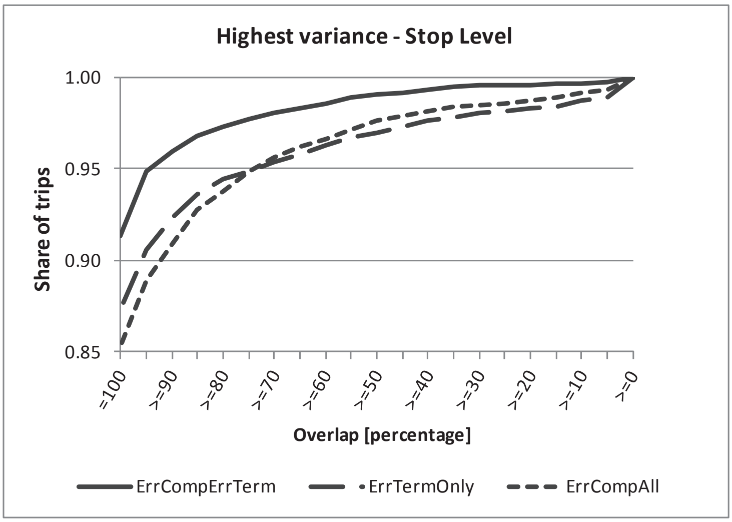

Figure 5-5: Coverage for stop level, highest variance for each scenario.

For comparison, in Figure 5-6 is shown the coverage at stop level for the three scenarios with variance 0.2. For this level of variance the ErrCompAll scenario performs much better than the ErrTermOnly scenario and for low coverage the ErrCompAll is the scenario that contributes by far most to the coverage of the best performing scenario, ErrCompErrTerm.

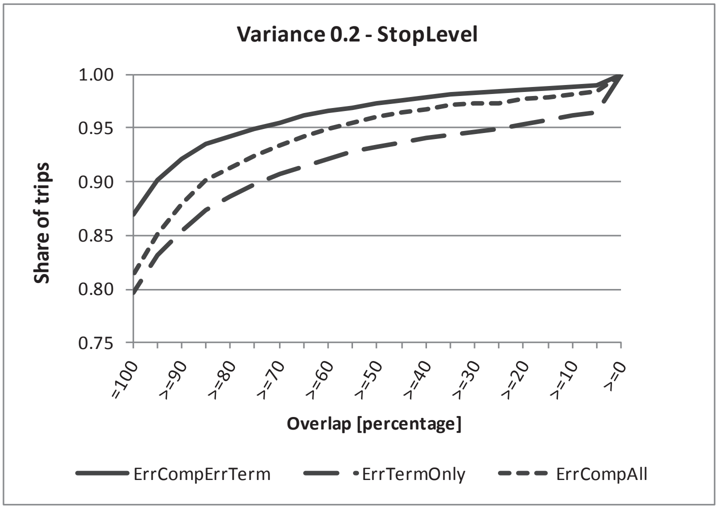

Figure 5-6: Coverage for stop level, variance 0.2 for each scenario.

In Figure 5-7 the coverage at the link and stop level are compared for the best performing scenario ErrCompErrTerm for the highest variance. The stop level has the highest coverage. At both levels the scenarios perform better than the 80% overlap threshold. At 92% the coverage on the two levels differs with 55 percentage points and at a 95% threshold the difference is more than 90 percentage points.

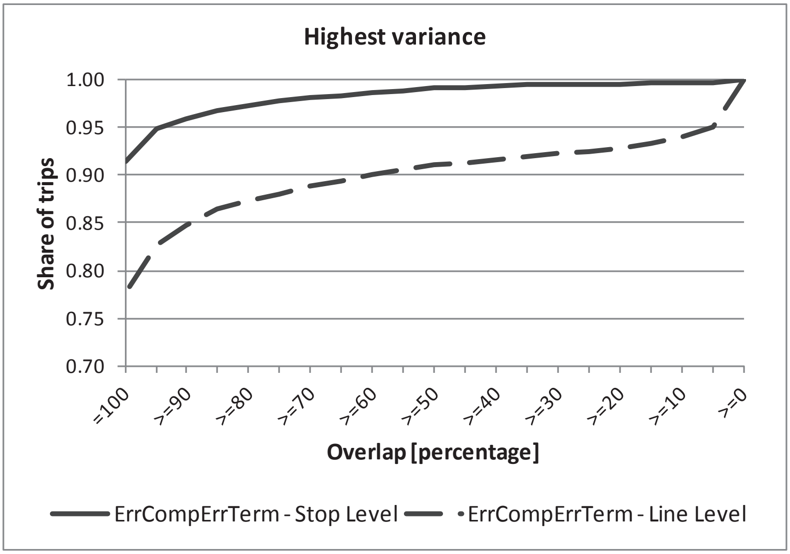

Figure 5-7: Coverage for link and stop level, highest variance for each scenario.

Even though the doubly stochastic generation algorithm produces very promising results, still there are observed routes that are not even remotely reproduced (2.8% at stop level) and are classifiable into three categories:

- The observed route seems reasonable, although the model does not generate it.
- The observed route does not seem reasonable or realistic, and for this reason the model does not generate it.
- The observed route does not seem rational, and is not generated by the model that does not consider this aspect, although from an activity-based perspective the route might be rational.

The number of unique generated routes at Link Level is summarised in Table 5-1. When more stochasticity is added to the parameters and/or the error term in average more routes are generated. The highest number of routes is generated in the ErrTermOnly and ErrCompErrTerm scenarios and for the highest variance.

Table 5-1: Total, maximum, minimum, average number of routes for the three scenarios at Link Level.

Configuration	Scale Par	Total	Min	Max	Average
ErrTermOnly	0.025	16,480	1	26	3.4
	0.05	20,882	1	32	4.3
	0.1	27,710	1	45	5.8
	0.2	39,294	1	61	8.2
	0.5	68,930	1	80	14.3
	1	114,734	1	95	23.8
ErrCompAll	0.05	22,162	1	30	4.6
	0.1	29,804	1	36	6.2
	0.15	36,044	1	41	7.5
	0.2	41,426	1	49	8.6
	0.5	61,039	1	59	12.7
	1	77,968	1	61	16.2
ErrCompErrTerm	0.05	29,305	1	39	6.1
	0.1	42,695	1	54	8.9
	0.15	57,671	1	69	12.0
	0.2	76,994	1	74	16.0
	0.5	134,469	1	91	28.0
	1	194,352	2	96	40.4

Among the observed routes not generated in the choice sets are different tendencies and Table 5-2 lists these tendencies along with remarks and explanations to why the routes are not generated.

Table 5-2: Tendencies for the observed routes not generated in the choice sets and remarks and explanations to these.

Tendency	Remarks/explanations
The travellers choose a public transport mode (typically bus) at a point where it makes no sense intuitively.	The route choice can be said to be adaptive, since the traveller on route sees an opportunity to save time/cost by using an arriving bus and therefore chooses to change route choice en route.
Travelling which can clearly be carried out by a less costly route.	In some cases the travellers travel along with fellow travellers on a part of the trip and the route choice is adapted to the route of the other traveller.
Among the registered data, information on a trip leg for a great part of the trip seems to be missing.	Can be due to wrong matching of the stated address, a wrong stated address or missing information in the data set.
Irrational choice of train station or bus stop or transport mode choice for access/egress mode.	During a day several trips are performed and the route and mode choices in each of the trips are mutual depended. This dependency is not modelled in the existing model but could be a part of an activity-based model.
Some trips use routes passing greater train stations or shopping areas even though faster alternatives exist.	Some travellers prefer well-served or centrally located stops where it is possible to spend the waiting time on e.g., shopping. This cannot be taken into consideration using the existing model for generation of route choice set. An ongoing project at DTU Transport is modelling the access transport mode and station choice in a discrete choice model since such station specific attributes can be used in these types of models.

In the following the choice set generated with the highest variances for the scale parameters and error terms are used.

5.5 Choice set visualisation

In this section a few choice sets for observed trips are visualised and discussed. Figure 5-8 shows a map of the choice set for a trip from Kgs. Lyngby to Værløse. 72 route alternatives are in the choice set calculated at the link level and consist of a total of five public transport modes (bus, E-bus, S-bus, local train and S-train). The routes are concentrated around the buses connecting the two points (no train serves both origin and destination). In the choice set are also routes using the suburban S-train (green lines on the map) to travel via Copenhagen. Also bus routes with buses to the north and south are present. It seems as if a few of the routes contain very long detours but some travellers might use these routes to avoid the local bus travelling more direct between the origin and destination.

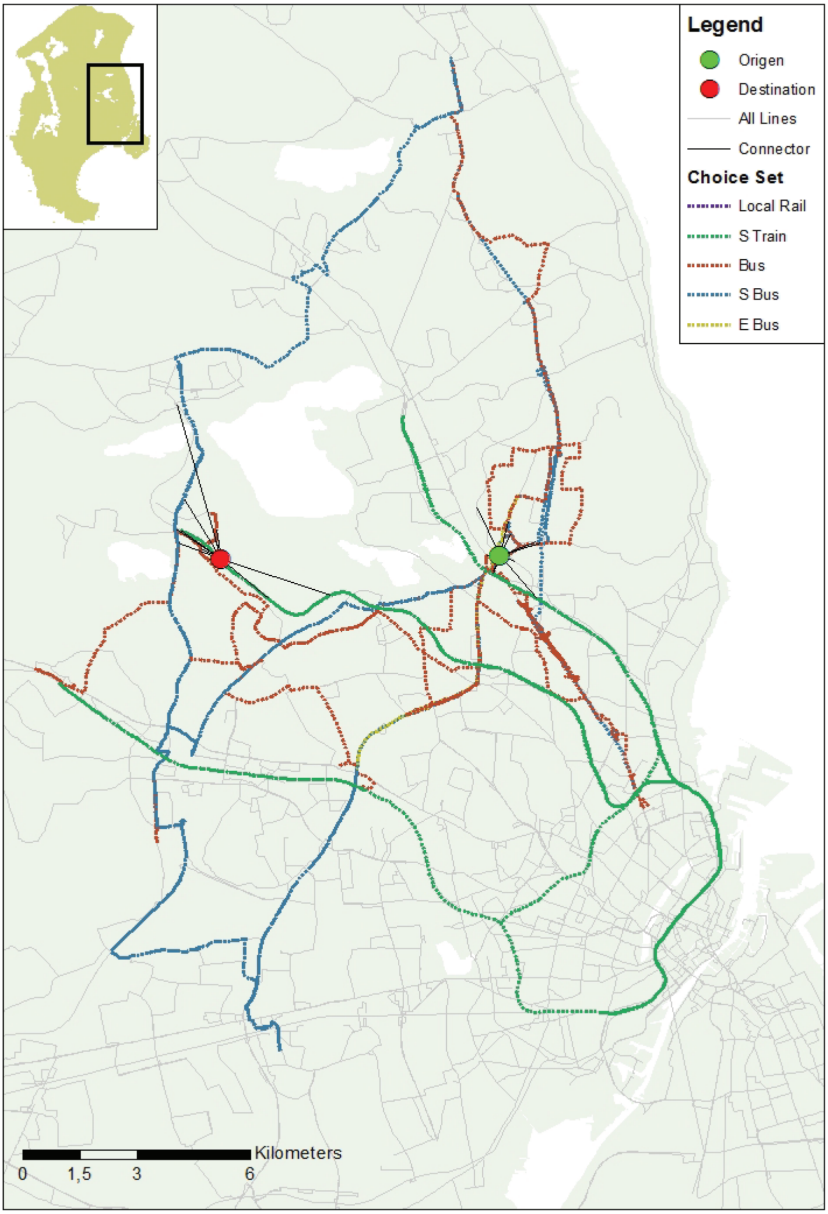

Figure 5-8: Choice set for a trip from Lyngby to Værløse using five public transport modes (ID1102459).

Figure 5-9 shows a map of the choice set for a trip from Copenhagen to Roskilde. In this choice set all public transport modes are present (Bus, S-bus, E bus, local train, metro, S-train, regional and IC-train). The alternatives are either using direct routes or more high class public transport modes (trains, high class buses) at longer routes leading south or north of the destination point. This choice set consists of 87 unique alternatives.

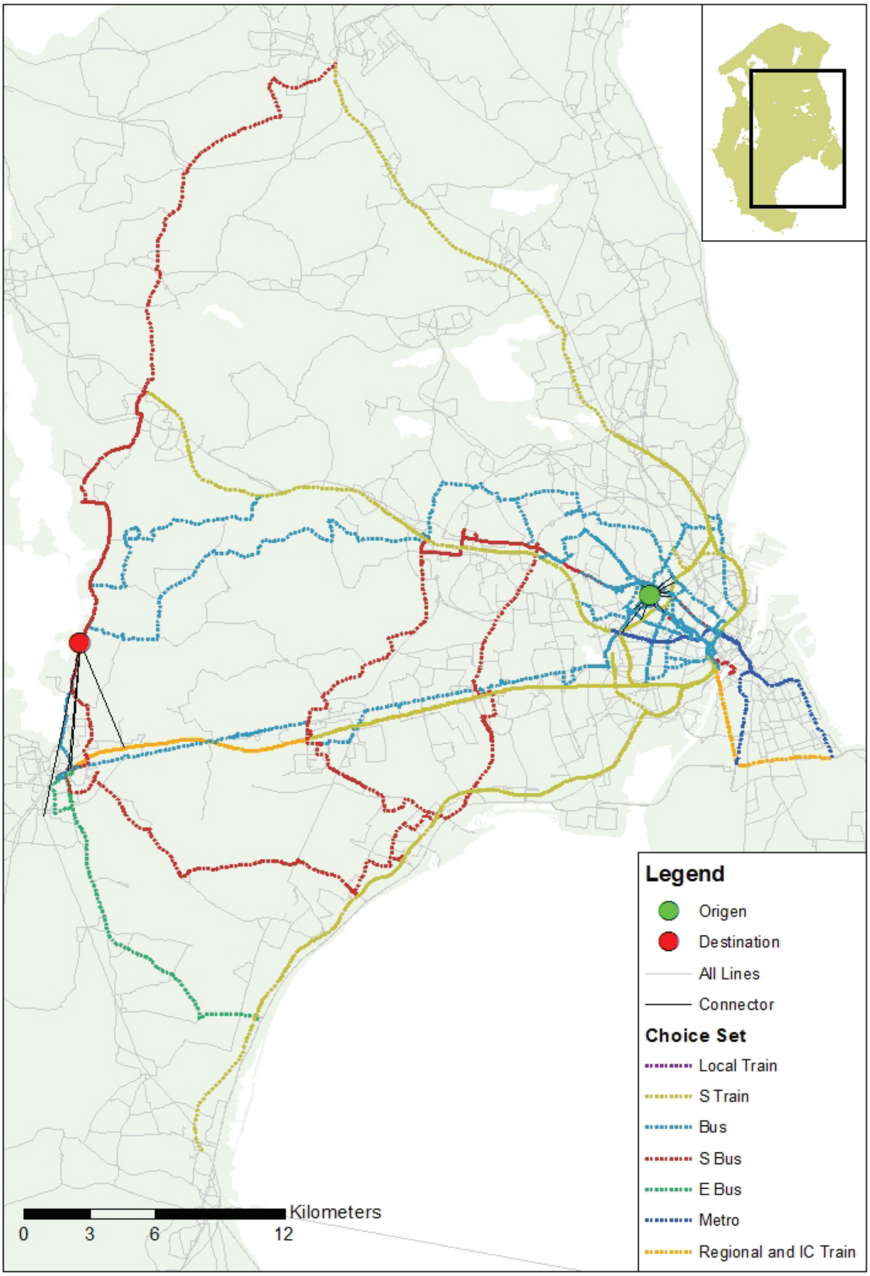

Figure 5-9: Choice set for a trip from Copenhagen using seven public transport modes.

In Figure 5-10 the choice set of a local trip in Copenhagen is visualised.

Figure 5-10: Choice set for a trip within Copenhagen centre using four public transport modes.

The choice set contains four public transport modes (Bus, S-bus, metro and S-train). The alternatives mainly use the bus, but also alternatives using the train are present, although these provide longer detours but minimise the walking/biking distances if combined with a bus from/to the train station. Actual loops are revealed at the eastern part of the where a route uses a metro, shifts to an S-bus and returns to the same point (but another station type though). The choice set for this OD pair consists of 35 routes.

The three choice sets visualised in the above shows that there is a great variety in the choice alternatives created using the doubly stochastic simulation method. Because of the variation which changes the traveller's perception of the trip attributes a large number of different, but still plausible routes are presented in the choice sets.

5.6 Summary and conclusions

In this chapter we have investigated the generation of choice sets in the public transport route choice context. As literature focuses mainly on private transport route choice and evaluates path generation techniques with small examples or synthetic experiments, the present study investigates actual choices of public transport users and assesses choice set quality against these RP data.

We implemented a schedule-based stochastic transit assignment model based on MSA for path generation. The model is probit-based, in order to account for similarities across alternatives, and is doubly stochastic, in order to account for heterogeneity in both the perceived value of time components and the perception of the link impedance. The value of time components concerns in-vehicle time for the various public transport modes available in the Greater Copenhagen Area (i.e., bus, metro, trains), waiting time at the stations and at the zone level, walking time from and to the stations, and connecting time relative to the structure of the network. 4,833 observations are examined in a GIS network that reproduces all the path and road network of the study area.

Results show that the best coverage results are obtained from the doubly stochastic generation function where error components for the value of time of individuals are drawn from a log-normal distribution alongside an error term from a Gamma distribution. Single stochastic generation functions are outperformed, and both the variation of the VOT-terms and the variation of the error term contribute to reach a good coverage. The study builds on Larsen et al. (2010) by adding higher variances to the error components and to the error terms. Higher variance produced more unique routes, as the number of unique routes generated shows that the same alternatives are continuously generated, with consequent low efficiency in the production of alternatives. Ideally, any stochastic generation technique should produce a variety of unique routes to allow for better coverage to be reached in a reasonable number of iterations but also it is important to create choice sets with a high variety of route choice alternatives.

When adding route choice alternatives to the route choice sets it is important that most routes are actually routes which will be considered by the traveller. But also routes which are not likely to be chosen can be added to the choice sets since the presence of these routes implies which attributes the traveller finds important and which preferences the traveller have for the different routes. If all routes are very similar and close to the optimal route of the traveller the preferences are more difficult to determine since small differences in the routes will be the only reason for choosing one alternative over another. This is also discussed in Chapter 6 in relation to the path size factor.

Further research could direct efforts into (i) reaching higher coverage by implementing the suggestions mentioned in Table 5-2, and (ii) the comparison of the results with alternative algorithms (e.g., deterministic techniques combined with heuristic rules).

6 ESTIMATION OF PUBLIC TRANSPORT ROUTE CHOICE MODELS

Route choice models are a very important part of traffic assignment and have been a research topic for decades. The understanding of the route choice behaviour is important to understand the preferences of travellers and to predict traffic flows in future scenarios. Route choice of car passengers has been the focus for the majority of the research within the route choice literature, but in recent years more literature has taken up the challenge of describing route choice for public transport passengers. Description of the multimodal structure of the public transport network is an important challenge to meet, making the public transport network very different from the car road network.

The following chapter introduces the literature on previous route choice models estimated or both car drivers and public transport passengers. A literature review concerning the development of discrete choice models with random utility maximization forms the basis for the presentation of various models suggested for describing the route choice.

6.1 Literature review

The following review of literature on both discrete choice and route choice models in general and more specific for public transport is partly based on Train (2003), Prashker and Bekhor (2004), Prato (2005) and Prato (2009).

6.1.1 Discrete Choice and Random Utility Models

Discrete choice models consider decision makers choosing among alternatives in a choice set. Train (2003) described three important attributes of the choice set:

- The alternatives must be mutually exclusive from the perspective of the decision maker; the decision maker chooses one and only one alternative from the choice set.
- The choice set must be exhaustive, so all (relevant) alternatives are included in the choice set.
- The number of alternatives must be finite.

Discrete choice models usually assume utility maximization behaviour by the decision maker. Utility is a measure of the benefit the decision maker gains by choosing a specific alternative from the choice set. The concept originates from Thurstone (1927) who described psychological stimuli and Marschak (1960) developed the concept into random utility derived from utility optimisation with the individual choice being probabilistic.

In a transportation network the traveller will select a route through the network from his route choice set. According to Ben-Akiva and Lerman (1985) the choice set consists of the routes feasible and known to the traveller and is a finite set of mutually exclusive alternatives. The choice is a discrete choice based on random utility optimization.

The decision maker obtains a certain amount of utility from each route in the choice set and he will maximise his utility by choosing the route with the highest amount of utility. This choice is modelled by using a random utility model (RUM).

It is assumed that the utility of each alternative is known to the decision maker and that he chooses the alternative from which he obtains the highest utility. The modeller determines the utility of the traveller by dividing the utility in two components: the deterministic component representing the observed attributes and the random component representing the uncertainty causes by the imperfect knowledge of the modeller and the stochasticity involved in choice behaviour. Manski (1973) identified the random part of the utility to be influenced by: unobserved attributes for alternative, unobserved taste variations, errors in the measurement, and instrumental (proxy) variables.

The traveller n maximizes his utility of choosing alternative k from the set of alternatives routes C_n:

$$U_{kn} = V_{kn} + \varepsilon_{kn} \ \forall \ k \in C_n \tag{6-1}$$

Where V_{kn} is the deterministic component of U_{kn} and ε_{kn} is the random component describing uncertainties in the model as described above. This means that the utility is expressed as follows:

$$U_{kn} = V(\beta_n;X_{kn}) + \varepsilon_{kn} \tag{6-2}$$

Where β_n is a vector of parameters representing the preferences of decision maker n and X_{kn} is a vector of attributes connected to route k as seen from the perspective of decision maker n.

In the random utility model it is assumed that the traveller chose the alternative with the highest utility in the choice set and the probability is calculated by:

$$P_n(k)= P(U_{kn} = \max(U) = P(V_{kn} + \varepsilon_{kn} > V_{jn} + \varepsilon_{jn}) \qquad \forall \ j \neq k, \text{with } j,k \in C_n \tag{6-3}$$

The random terms defined in the probabilistic model determines the structure of the choice model and the various model types are estimated from this. A Gumbel distribution of the error term returns a logit model and a normal distribution generates a probit model.

6.1.2 UE and SUE

Historically, the deterministic User Equilibrium (UE) defined by Wardrop (1952) has been the most studied approach in traffic assignment. The UE assumes that the traveller has perfect knowledge of the network and the route costs and selects the route which will minimise his travel costs. The UE is defined as the state in which no traveller can reduce his travel cost only by changing route. Obviously, the traveller does not have perfect knowledge of the network and this approach needed to be improved.

Daganzo and Sheffi (1977) extended the UE to the Stochastic User Equilibrium (SUE) which accounts for uncertainty in perception of the travel costs. The SUE is defined as the state in which no traveller can reduce his *perceived* travel cost only by changing route. Sheffi and Powell (1982) defined an operational solution algorithm to the SUE called Method of Successive Averages (MSA). To represent the difference in route cost caused by the traveller's different perceptions of the costs the coefficients in the utility are drawn from a distribution which typically has the real cost as an average.

The Multinomial Logit model is one of the most often used discrete choice models. MNL has the disadvantage that it cannot capture the similarities among alternatives and is therefore not suitable for route choice and this problem is dealt with in different ways in the various model specifications. The C-logit and Path Size logit use the MNL structure include a correction factor in the deterministic part of the utility function to deal with overlapping. Generalized Extreme Values models (such as Paired Combinatorial Logit, Cross-Nested Logit) represent similarities among routes in the error component of utility. Models with a probit-structure use both the deterministic and the random part of the utility function to account for behavioural differences.

The Method of Successive Averages (MSA) was developed by Sheffi and Powell (1982) to use in the search for an improved feasible solution. The algorithm in based on a predetermined step size along the descent direction and is therefore not determined on the basis of the current solution or of the objective function. Sheffi (1985) showed that the MSA SUE can be used for route choice models in both analytical and simulation based methods.

6.1.3 Models with MNL structure

MNL

An early route choice model specification is the Multinomial Logit (MNL). Luce (1959) developed the first Logit model with Gumbel distributions for the error terms. The model is widely used for route choice models despite of some disadvantages. Train (2003) discussed the model extensively.

The multinomial logit choice probability is expressed by:

$$P_n(k) = \frac{exp(V_{kn})}{\sum_{j \in C_n} exp(V_{jn})} \tag{6-4}$$

The model has a simple analytic form and is easy to estimate. Manski (1977) described the model as an Independent and Identically Distributed Random Utility model (IIDRU) and this characteristic prevents the use of random taste variations across decision makers.

The MNL has the disadvantage that it cannot capture the similarities among alternatives and is therefore not suitable for route choice. Luce (1959) was the first to derive the logit formula from the assumptions of the characteristics of choice probabilities, most importantly the Independence from Irrelevant Alternatives (IIA). In transport network many routes will be overlapping to some extent. The IIA claims that the choice probability for an alternative is independent of the attributes of other alternatives. Thereby if a new alternative is added to the network the probability of selecting the new alternative will be drawn equally from probabilities of existing alternatives (and the opposite if removing an alternative) (see e.g., McFadden, 1973). McFadden (1973) showed that the unobserved component of the utility is Type 1 Extreme Value (Gumbel) distributed.

The Gumbel distribution allows for use of cumulative distributions and enhanced route choice models have been developed from this model containing the simplicity of the logit structure.

C-Logit

Cascetta et al. (1996) introduced a modification of the MNL model called C-Logit. The C-Logit keeps the closed analytical structure of the MNL but overcomes the problems with overlapping. The model deals with similarities among overlapping routes using an additional cost attribute (the Commonality Factor) in the MNL utility function.

The model proposed is:

$$P_n(k) = \frac{exp(V_{kn} - \beta_{CF} \cdot CF_k)}{\sum_{j \in C_n} exp\left(V_{jn} - \beta_{CF} \cdot CF_j\right)} \tag{6-5}$$

Where CF_k is the commonality factor of route k.

The commonality factor corrects the utility function for similarities among routes and the factor measures to what degree the given route is overlapping with other routes in the route choice set.

Four different definitions of the commonality factor are suggested by Cascetta et al. (1996):

$$CF_k = ln \sum_{j \epsilon C_n} \left(\frac{L_{kj}}{\sqrt{L_k L_j}}\right)^{\gamma} \tag{6-6}$$

$$CF_k = ln \sum_{a \epsilon \Gamma_k} (w_{ak} N_a) \tag{6-7}$$

$$CF_k = \sum_{a \epsilon \Gamma_k} \left(w_{ak} ln(N_a)\right) \tag{6-8}$$

$$CF_k = ln\left[1 + \sum_{j \epsilon C_n, k \neq j} \left(\frac{L_{kj}}{\sqrt{L_k L_j}}\right)\left(\frac{L_k - L_{kj}}{L_j - L_{kj}}\right)\right] \tag{6-9}$$

Where:

- L_{kj} is the length of links shared between routes k and j.
- L_k and L_j are the length of route k and j respectively.
- Γ_k is the set of links belonging to route k.

- w_{ak} is proportional weight of link a for route k, defined for example as the fraction the length of link a holds of route k, weights sum to 1 for all links on a route.
- N_a is the number of routes between each OD pair sharing link a.
- δ_{al} is the link-path incidence dummy equal to one if route j uses link a and zero otherwise.

The commonality factor should always be positive or zero (if the route does not share links with any other route in the choice set). A positive value assigns lower flows to overlapping routes compared to predictions by MNL. A new route almost equal to an existing imposes a decrease of the probability of the existing but the sum of the two should be at least equal to the original probability of the existing route.

Path Size Logit

Ben-Akiva and Bierlaire (1999) introduced the Path Size Logit which uses a size variable in the utility of the route to correct for overlapping alternatives.

$$P_n(k) = \frac{exp(V_{kn} + \beta_{PS} \cdot lnPS_{kn})}{\sum_{j \in C_n} exp\left(V_{jn} + \beta_{PS} \cdot lnPS_{jn}\right)} \tag{6-10}$$

Where PS_k is the Path Size factor of route k and is defined by:

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{j \epsilon C_n} \delta_{aj} \frac{L^*_{C_n}}{L_j}} \tag{6-11}$$

Where:

- L_a, L_k and L_j are the length of link a, route k and j respectively.
- Γ_k is the set of links belonging to route k.
- δ_{al} is the link-path incidence dummy equal to one if route j uses link a and zero otherwise.
- L^*_{Cn} is the length of the shortest route in C_n.

The C-logit and the Path Size logit both add a correction term to the utility function in the MNL model but the correction terms differ. The commonality factor is always equal to or greater than zero decreasing the utility when overlapping with other routes. The Path Size factor is always between zero and one providing the amount of overlapping with another route. Ben-Akiva and Bierlaire (1999) observed that the Path Size factor is equivalent to the Commonality Factor of the C-logit when two routes are either completely overlapping or completely non-overlapping.

For the Path Size logit the ratio between the L_a and L_k is equal to the route impedance from link a. The second part of the formula holds information on how many routes are using the link and the length of the link and the chosen route. The fraction equals 1 when link a is used by one route only. If more than one route uses the link the term depends on the ratio of the length of the shortest path and the alternative route.

Ramming (2002) suggested an expansion of the Path Size formula to account for different contributions from routes with different lengths:

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{j \in C_n} \left(\frac{L_k}{L_j}\right)^{\gamma} \delta_{aj}} \tag{6-12}$$

Where γ is a parameter to be estimated.

Bovy et al. (2009) determined the Path Size Correction (PSC) factor based on theoretical arguments. They derived the probability of choosing an alternative *k* to be:

$$P(k) = \frac{exp(V_k + \beta_{PSC} \cdot PSC_k)}{\sum_{j \in C_n} exp\left(V_j + \beta_{PSC} \cdot PSC_j\right)} \tag{6-13}$$

Where the Path Size correction factor for route *k* is calculated as:

$$PSC_k = -\sum_{a \in \Gamma_k} \left(\frac{l_a}{L_k} ln \sum_{j \in K_n} \delta_{aj} \right) \tag{6-14}$$

The values of the PSC logit factor vary between $-\infty$ and 0 whereas the values of the original PS logit factor are between 0 and 1. The estimates of the PSC show similar performances to the original PS proposed by Ben-Akiva and Bierlaire (1999). Bovy et al. (2009) included a parameter β_{PSC} whereas no β_{PS} were included in the formulation of Ben-Akiva and Bierlaire (1999).

In the suggested Path Size formulations for car one Path Size factor is calculated for each alternative to measure the amount of overlap the specific route alternative has with all other routes in the choice set. Hoogendoorn-Lanser and Bovy (2007) suggested separate PS factors for different parts of the trip in a multimodal transport network with only the train as the main transport mode. They identified three trip parts *R*: home-end, activity-end, and train part. They proposed three formulations and found the Path Size factor PS_{irn} based on the total length L_i of the full route *i* and the total number N_{na} of unique routes using leg *a* to capture the most correlation between trip parts.

$$PS_{irn} = \frac{1}{L_i} \sum_{a \in \Gamma_{ir}} \frac{l_a}{N_{na}} \tag{6-15}$$

For a car network with unrestricted choice set size, Fosgerau et al. (2013) proposed a link size attribute which is similar to the path size attribute but additive.

6.1.4 GEV

The Generalized Extreme Value (GEV) models were presented in order to solve the problems with the MNL models and are based on generalisations of the extreme value distributions. The generalisation allows correlation in the unobserved factors over alternatives and equals the multinomial logit model when the correlation is zero.

Nested Logit

The Nested Logit model is an extension of the MNL to capture some overlapping between alternatives. The choice set is divided into nests and the utility function of each alternative is added an alternative specific term and a term associated with the nest the alternative belongs to. The nested logit is not useable for route choice since it assumes that an alternative belongs to one nest only and all alternatives in a nest are required to be similar. Considering the links as nests implies that each route should be completely different from any other alternative which is not realistic. See also Ben-Akiva and Lerman (1985) for a description of the nested logit model.

Other logit-based models developed from the GEV theorem of McFadden (1978) can be used to model route choice.

Generalized Nested Logit

The GEV type model Cross Nested Logit, CNL, was developed from the NL by Vovsha (1997) particularly for mode choice modelling and Vovsha and Bekhor (1998) extended it to route choice. The Cross Nested Logit uses links shared by several routes as a basis for nesting and the overlapping of routes is handled by a nesting parameter. The CNL allows for a route to belong to more than one nest. Cascetta (2001) showed that for realistic size networks and especially complex multimodal networks the nesting structure will be extraordinarily complex.

Wen and Koppelman (2001) further developed the CNL to the Generalized Nested Logit, GNL.

The Generalized Nested Logit model also allows for a route to belong to more than one nest which in the route choice context is equal to a link. The probability of choosing a route *k* is:

$$P(k) = \sum_{m} P(j)P(k|m) \tag{6-16}$$

Where the conditional probability of choosing route *k* in nest *m* is

$$P(k|m) = \frac{\left(\alpha_{km} exp(V_k)\right)^{\frac{1}{\mu}}}{\sum_j \left(\alpha_{jm} exp(V_j)\right)^{\frac{1}{\mu}}} \tag{6-17}$$

Where α_{km} is the inclusion coefficient defined as $0 \leq \alpha_{km} \leq 1$ and the μ is the nesting coefficient defined as $0 \leq \mu \leq 1$.

The marginal probability of choosing nest *m* is:

$$P(m) = \frac{\left(\sum_k \left(\alpha_{km} exp(V_k)\right)^{\frac{1}{\mu}}\right)^{\mu}}{\sum_b \left(\sum_k \left(\alpha_{kb} exp(V_k)\right)^{\frac{1}{\mu}}\right)^{\mu}} \tag{6-18}$$

The probability of choosing route *k* depends on the deterministic part of the utility function, the inclusion coefficients in the nests *m* that form route *k*, and the nesting coefficients. The GNL allows a route to belong to more than one nest.

The CNL is also a special case of the GNL where all nests have the same nesting coefficient μ, which is a parameter to be estimated. Vovsha and Bekhor (1998) defined the inclusion coefficient from the links on the route:

$$\alpha_{ka} = \frac{L_a}{L_k}\delta_{ka} \tag{6-19}$$

Where δ_{ka} is the link-path incidence dummy equal to 1 if link *a* is on route *k* and 0 otherwise.

The inclusion coefficient μ describes the degree of nesting. When μ is 1 the CNL is equal to the simple MNL model. When the degree of nesting increases and $\mu \rightarrow 0$ the model turns probabilistic at the higher (link) level and deterministic at the lower (nest) level.

This extreme case is suitable for route choice since fully overlapping routes would be considered as one route and mutually exclusive routes are distributed only according to route utilities.

Paired Combinatorial Logit

Chu (1989) proposed the GEV-type model Paired Combinatorial Logit (PCL) with pairs of alternatives making up a nest for each. When having *J* alternatives each alternative is a member of *J*-1 nests and the unobserved utility can be estimated. Koppelman and Wen (2000) developed the PCL model further and described the derivation, structure, properties and estimation of the model.

Since the PCL allows for differential correlations in the unobserved utilities between pairs of alternatives the PCL overcomes the problem of the Nested Logit formulation where all alternatives in a nest are required to be similar.

The PCL model was adapted to route choice models by using a similarity coefficient which is correlated to the network topology. In the PCL routes are chosen among a pair of alternatives in the choice set. The PCL model allows for differential correlation between the pairs of alternatives in the choice set and the probability of choosing an alternative *k* is:

$$P(k) = \sum_{k \neq j} P(kj)P(k|kj) \tag{6-20}$$

Where *P(kj)* is the marginal probability of choosing the pair of alternatives *k*,*j* from the choice set of *J*(*J*-1)/2 pairs given as:

$$P(kj) = \frac{(1-\sigma_{kj})\left(exp\left(\frac{V_k}{1-\sigma_{kj}}\right) + exp\left(\frac{V_j}{1-\sigma_{kj}}\right)\right)^{1-\sigma_{kj}}}{\sum_{p=1}^{n-1}\sum_{q=p+1}^{n}(1-\sigma_{qp})\left(exp\left(\frac{V_q}{1-\sigma_{qp}}\right) + exp\left(\frac{V_p}{1-\sigma_{qp}}\right)\right)^{1-\sigma_{qp}}} \quad \textbf{(6-21)}$$

Where σ_{kj} is a similarity index between alternatives *k* and *j*.

P(*k*|*kj*) is the conditional probability of choosing alternative *k* given the chosen pair of alternatives *k,j* and is given as follows:

$$P(k|kj) = \frac{exp\left(\frac{V_k}{1-\sigma_{kj}}\right)}{exp\left(\frac{V_k}{1-\sigma_{kj}}\right) + exp\left(\frac{V_j}{1-\sigma_{kj}}\right)} \quad \textbf{(6-22)}$$

The paired combinatorial definition of the model is given by the double summation including *J*(*J*-1)/2 elements which is the number of different pairs of alternatives in the choice set of *J* alternatives.

Prashker and Bekhor (1998) defined a functional form of the similarity index similar to the C-logit commonalty factor as:

$$\sigma_{kj} = \frac{L_{kj}}{\sqrt{(L_k L_j)}} \quad \textbf{(6-23)}$$

Where L_{kj} is the length of the path common for route *k* and *j*. The similarity index takes on values between 0 and 1 which also is necessary for the PCL to be consistent with random utility optimisation. For similarity index values approaching 1 the routes are very similar with each other and if it is zero the two routes do not have any links in common. Because of the pairwise comparison at the upper level in the nest the number of nests increases quickly with the size of the network and the approach has only been applied to small networks on synthetic data.

6.1.5 Models with probit structure

Unlike logit models, probit models can handle random taste variation and allow any pattern of substitution. To solve the behavioural realism problems, the probit models modify both the deterministic and the random parts of the utility function. With the parameters randomly distributed the utility for choosing an alternative will be $U_{nk} = \beta_n' x_{nk} + \varepsilon_{nk}$

Where β_n is a vector of coefficients representing the tastes of traveller *n*.

The probit model describes the unobserved parts of utility as following normal distributions. A cost coefficient cannot be assumed to follow a normal distribution since this would mean that some people have a positive price coefficient. Ben-Akiva and Bolduc (1996) and McFadden and Train (2000) suggested models with Probit structure using combinations of Gaussian and Gumbel error terms.

Multinomial Probit

The Multinomial Probit (MNP) was proposed as an alternative to model route choice by Daganzo and Sheffi (1977) by describing the random coefficients by a normal distribution.

In a multinomial probit model with J alternatives the joint density of the error terms has a vector of means of length J and a covariance matrix of length J x J. The calculation of the probit choice probability is not straightforward when calculating for a large number of routes since no closed form exits for the cumulative normal distribution and numerical techniques must be used. The covariance matrix is used to calculate the choice probabilities but the specification of this is difficult since the covariance matrix has to relate the error terms to the network attributes. The unrestricted covariance matrix will have $[(J\text{-}1)J/2]\text{-}1$ covariance parameters when normalized and $J(J+1)/2$ when not normalised.

In Sheffi and Powell (1982) the variance was assumed to be proportional to a fixed link attribute for example length or free flow travel time. Yai et al. (1997) estimated a model where the covariance matrix depends on overlapping between routes in terms of measurable attributes such as length or free flow time similar to the first mentioned model.

The computational effort for estimating MNP route choice models is high and therefore alternative formulations are often preferred to this.

Error Component, Mixed Logit or Logit Kernel with random coefficients

The Error Component (EC) model is a Normal Mixture of the MNL model (MMNL) and was described by McFadden and Train (2000) and by Ben-Akiva and Bolduc (1996) who referred to it as Logit Kernel (LK). The EC model splits the unobserved part of utility into a component containing correlation and heteroscedasticity and an i.i.d. Gumbel random part.

The EC can amongst others be used to capture the overlap between routes. The flexibility in the configuration is high since the taste coefficients can be specified to be distributed randomly over individuals, hereby specifying the unobserved heterogeneity of individuals.

In the Error Component model based on random coefficients the traveller has a choice set of alternatives C_n and the probability of traveller n choosing route k is described by

$$P_{nk}(\beta_n) = \frac{exp(\beta_n' X_{nk})}{\sum_j exp(\beta_n' X_{nj})} \tag{6-24}$$

Where β_n is a vector of random coefficients representing the tastes of traveller n and X_{nk} is the observable variables relating traveller n to alternative k. The modeller cannot observe β_n and cannot condition on β_n and the unconditional choice probability is therefore calculated as the integral of $P_{nk}(\beta_n)$ over all values of β_n

$$P_{nk} = \int \frac{exp(\beta' X_{nk})}{\sum_j exp(\beta' X_{nj})} f(\beta) d\beta \tag{6-25}$$

Where *f(β)* is the distribution density of *β* over the population. Simulation is used to compute the unconditional probability:

$$P_{nk} = \frac{1}{D}\sum_{d=1}^{D}\frac{exp(\beta_d' X_{nk})}{\sum_j exp(\beta_d' X_{nj})} \tag{6-26}$$

Where β_d represents a draw *d* from the distribution of *β* and *D* is the number of draws.

The distribution for the parameters has often been specified as normal or log-normal (McFadden and Train, 2000 and Ben-Akiva et al., 1993). Nielsen (2000) showed that preferences are well simulated by log-normal and gamma distributions and he proposed a stochastic traffic assignment model with differences in passenger's utility functions. Nielsen et al. (2002) tested normal and log-normal distribution coefficient for travel time and cost for different traveller categories.

The log-normal distribution is intuitively better suited to represent the coefficients of costs than the normal distribution since the latter implies that some people have a positive cost coefficient and therefore prefer longer and more costly trips. Nielsen et al. (2002) showed the log-normal distribution to give large variation and sometimes illogical results and according to Han et al. (2001) it does not produce satisfactory results. Nielsen (2000) found that gamma distributions for coefficients produce reproducible and non-negative results and is therefore preferable to log-normal.

The research of Mabit and Nielsen (2006) showed that the randomisation of the parameters in the choice model improves the model fit.

6.1.6 Route choice models in public transport

For public transport networks and multimodal networks a number of studies are presented in the following.

Van der Waard (1988) estimated a multinomial model for route choice in public transport by the use of revealed preference (RP) data. He investigated the impact of the route attributes such as walking time, waiting time, in-vehicle-time (split on trip purposes) and number of transfers. The cost was not taken into account because of the fare structure (see also Chapter 2 for a description of the fare structure in the Greater Copenhagen Area).

The C-logit approach was applied to a public transport network by Nuzzolo et al. (1997).

Koppelman and Wen (2000) investigated a number of model formulations (MNL, PCL and NL) for the high speed train corridor between Toronto and Montreal, and specifically discussed the use of the paired combinatorial logit for route choice. Benjamins et al. (2002) implemented the paired combinatorial logit model for route choice in a super-network approach.

Lo et al. (2004) applied a three-level nested logit approach for a multimodal travel network. The three levels accounted for the different decisions in the transport network. The first two levels dealt with combined-mode choice and transfer location choice and the third level with route choice.

Nielsen (2004a) tested different error component model configurations for travellers in the Danish public transport system of Copenhagen-Ringsted and ended up deriving a model with only one error component, thereby assuming full correlation between time coefficients.

Hoogendoorn-Lanser (2005) estimated several specifications for route choice models in a multimodal transport network for trips with train as the main transport mode. These models included home-activity MNL and access-egress MNL (relevance of trip attributes), PS logit and GNL for route choice and recommended PS model over the MNL model and the GNL model over the PS models.

Hoogendoorn-Lanser and Bovy (2007) estimated a PS logit model for a multimodal network with separate PS factors for different parts of the trip.

Vrtic and Axhausen (2002) estimated route choice attributes for both separate and joint route and mode choice nested logit model in a regional and long distance public transport network.

Raveau et al. (2011) introduced network topological attributes in the estimation of a multinomial logit model in the Santiago metro network using data collected amongst travellers in the metro.

Eluru et al. (2012) estimated a transit route choice model for the multimodal public transport network of Montreal including buses, metros and trains and walking as access/egress mode. They collected data on commuting trips amongst the staff at the McGill University and estimated a mixed multinomial logit model assuming Normal distributions for travel time in train, walking time, and number of transfers.

Definitions of overlap in public transport

The similarity calculations explained in the above are in most cases derived for private transport route choice or for spatial choices, which could be rather different from public transport. The road similarity factors are often measured in spatial dimension and the spatial dimension are not as important in the public transport networks. Hoogendoorn-Lanser and Bovy (2007) found the use of common trip legs to be describing the overlap well in multimodal route choice when using a hub-and-spoke network. Cascetta and Papola (2003) found similarities in departure time to be more important than similarities between public transport modes. Also fare was important. Friedrich et al. (2001) defined a similarity measure building on the transfer choices.

Bovy (1996) defined the overlap for road networks to be calculated in terms of distance. When the overlap i calculated as the distance of the overlapping legs, the overlap is independent of the conditions of the road network such as congestion, weather conditions, etc. If defined in terms of time the overlap between two alternatives increases if using a congested road. Instead of actual travel time Ramming (2002) suggests using the free-flow time to overcome the dependency of network conditions. Ramming (2002) shows that the model results using overlap expressed in free-flow time are better (in terms of log-likelihood) compared to the results when using overlap expressed in travel distance.

In public transport networks the scheduled travel times have the same characteristics as the free-flow travel times in road networks and could therefore be used to measure the overlap between alternatives. Hoogendoorn-Lanser et al. (2005) analysed three different formulations of overlap measures in multimodal transport networks, that is overlap based on: number of legs, scheduled travel time, and distance. They estimate a number of models using the PS formulation of Ramming (2002) and conclude that overlap based on time and distance does not improve the model estimates compared to the best MNL setup. The models are specified with fixed and estimated parameter β_{PS} and with two different levels for the scale parameter γ (0 and 20). The models with fixed PS parameter perform worse than MNL and specifications with number of legs for both γ values outperform the time and distance specifications.

6.1.7 Estimated public transport route choice parameters in literature

The following section presents the work of other authors who have investigated the route choice in public transport and estimated public transport route choice parameters. The tables presented are used for comparison to the findings in the results sections of this chapter.

Van der Waard (1988)

Van der Waard (1988) estimated multinomial logit models for route choice based on a sample of 1,095 public transport travellers. The most detailed model included in vehicle times for various public transport modes. Table 6-1 reports the parameter estimates presented by van der Waard (1988) scaled to in-vehicle time for the bus.

Table 6-1: Parameter estimates from van der Waard (1988) scaled to bus in-vehicle time (=1.0).

Parameter	Scaled to bus IVT
Access time	2.3
Egress time	1.2
Waiting time at first stop	1.4
In-vehicle time (all modes)	
In- vehicle time Bus	1.0
In- vehicle time Tram	1.0
In- vehicle time Rapid Transit	0.9
Walking time at transfer	2.2
Waiting time at transfer	1.2
Number of transfers	5.9

According to the final model the travellers perceive the time spent in the bus and the tram equally and prefer the rapid transit over the others. Access time is valued double as much as egress time and the waiting time at the first stop is only slightly worse than the waiting time at transfers along the trip.

Vrtic and Axhausen (2002)

Vrtic and Axhausen (2002) investigated the importance of a variety of factors for regional and long distance public transport by using data from a SP survey which was built on the continuous RP survey about travel behaviour of the Swiss population.

They investigated the attributes of in-vehicle time, number of transfers, transfer time, headway, fare and comfort and they tested for differences between trip purposes and demographic characteristics.

Table 6-2: Selected parameter estimates from Vrtic and Axhausen (2002) scaled to train in-vehicle time (=1.0).

Parameter	All	Commuters	Shopping	Leisure/Vacation	Business
In-vehicle Time (train)	1.0	1.0	1.0	1.0	1.0
Headway	0.3	0.3	0.2	0.2	0.0
Transfer time	0.5	0.6	1.2	0.4	0.8
Number of transfers	18.9	7.6	6.0	22.4	8.6

The relative low headway is explained by the fact that the railway services investigated mostly have half hour frequencies and is therefore not as important for the route choice as the other components.

Bovy & Hoogendoorn-Lanser (2005)

Bovy and Hoogendoorn-Lanser (2005) studied 235 multimodal trips all consisting of three trip parts: a main trip part (one or more train legs) using train and access and egress legs. Bovy and Hoogendoorn-Lanser (2005) considered a number of dummies and variables and found the parameters shown in Table 6-3 to be significant for a nested logit model and for a multi-Nested GEV model.

Table 6-3: Parameter estimates from Bovy & Hoogendoorn-Lanser (2005) scaled to in-vehicle train time (=1.0).

Parameter	NL	MN-GEV
Access IVT – private modes	1.6	1.6
Access IVT – public transport modes	0.8	0.8
Train IVT	1.0	1.0
First wait time	2.2	2.2
Waiting time at transfers (train–train)	2.2	2.2
Walking time at transfers	2.0	1.9
Number of high frequency transfers	5.7	5.1
Number of low frequency transfers	11.4	11.4

The models show that the access time by private transport modes is assessed more cumbersome for the travellers in the survey than access travel time in public transport modes. For both models the public transport mode access time is more attractive than IVT for the train trip part. Waiting/walking transfer time is assessed twice as bad as train IVT. The authors divided the transfers in high and low frequency transfers and showed a higher preference for the transfers to a public transport mode with high frequency since a missed transfer affects the total travel time less if the public transport line has high frequency.

Nielsen and Frederiksen (2006)

In their paper Nielsen and Frederiksen (2006) used more than 8,500 observations from a number of SP and RP studies to estimate an error component model. They segmented the data according to three trip purposes and estimated a number of trip specific parameters as presented in Table 6-4.

Table 6-4: Selected parameter estimates from Nielsen and Frederiksen (2006) scaled to bus in-vehicle time (=1.0).

Parameter	Commuters	Business	Education/leisure
Bus in-vehicle time	1.0	1.0	1.0
S-train in-vehicle time	0.8	0.8	0.8
Regional + IC-train in-vehicle time	-	0.7	0.8
IC (>60 min) in-vehicle time	1.9	-	-
Access/egress time	1.3	0.9	1.7
Hidden waiting time	0.5	0.3	0.6
Wait + transfer time	1.1	0.9	2.3
Delay at the destination	1.4	1.4	2.7
Penalty if no seat	4.5	4.0	15.0
Transfer penalty	15.1	13.5	20.0

For all trip purposes train in vehicle time is perceived less negatively than the bus in-vehicle time. For commuters and leisure/education access/egress time is more burdensome than bus and for business travellers almost equally negative. The penalty for no seat and for transfers are perceived from 4 to 20 times worse than one bus in-vehicle time both worst for the education/leisure travellers.

Raveau et al. (2011)

Raveau et al. (2011) presented a topological route choice model for metro trips in the city of Santiago. The observed routes were collected among travellers at the metro stations and observed routes with one or more “reasonable” alternative routes were used for estimation (16,029). By reasonable the authors meant routes chosen by other travellers in the survey. The route choice set then consisted of the observed routes of others and was limited to two routes (observed and alternative) for 97% of the trips.

Raveau et al. (2011) investigated the importance of adding factors based on the topology of the network to the model. They tested if the users preferred the most direct route, if they prefer the better known or most heavily travelled routes, and if the travellers consider variables besides the traditional (travel time, waiting time, fare, etc.) such as factors relating to comfort, reliability and physical characteristics from the modes and the stations.

The metro system has (as the fare system in the public network of the Greater Copenhagen Area) a flat fare structure where the fare depends on the origin and destination of the trip and therefore they do not estimate a fare parameter.

The estimated parameters to compare to the parameters estimated in the route choice models above are listed in Table 6-5:

Table 6-5: Selected parameter estimates from Raveau et al. (2011) scaled to metro in-vehicle time (=1.0).

Parameter	Base model	Proposed model
In-vehicle travel time metro	1.00	1.00
Waiting time	1.41	0.93
Walking time		2.02
Number of transfers	8.47	3.77

As can be seen in the table the inclusion of the variables relating to topological factors reduces the importance of the waiting time and the number of transfers significantly.

Abrantes and Wardman (2011)

Abrantes and Wardman (2011) collected British evidence on the values of travel time from numerous studies. The 226 studies investigated included both car and public transport trips and amongst the studies were found both RP and SP studies, and mode choice and route choice studies. From the studies the authors derived a table presenting the overall time multipliers compared to in-vehicle time (all modes).

Table 6-6: Selected parameter estimates from Abrantes and Wardman (2011) scaled to in-vehicle time (=1.0).

Parameter	Scaled to IVT
In-vehicle time	1.00
Walking Time	1.65
Out of vehicle time	1.43
Waiting Time	1.70
Headway	0.78

Travellers prefer the in-vehicle time over all other attributes than headway.

Eluru et al. (2012)

Eluru et al. (2012) estimated a transit route choice model for the multimodal public transport network of Montreal. The public transport network includes buses, metros and trains and walking was registered as access and egress mode. The estimations were carried out using 1,228 observations of actual route choices for commuting trips collected amongst the staff at the McGill University. Eluru et al. (2012) estimated a mixed multinomial logit model assuming Normal distributions for travel time in train, walking time, and number of transfers. The estimated parameters scaled to bus in-vehicle time are presented in Table 6-7.

Table 6-7: Selected parameter estimates from Eluru et al. (2012) scaled to bus in-vehicle time (=1.00).

Parameter	Scaled to Bus IVT
Transit alternative has bus	0.88
Transit alternative has metro	-2.37
Transit alternative has train	5.82
The alternative with the earliest arrival time	-0.88
Travel time in bus	1.00
Travel time in metro	0.60
Travel time in train	0.65
Standard Deviation	-0.18
Total Walking time	1.32
Total Walking time squared	0
Standard Deviation	-0.48
Number of transfers	9.29
Standard Deviation	-3.63
Waiting Time per transfer	0.28

The parameter estimates show that the travellers prefer metro over bus and bus over train. The authors explain this by the inconvenience of train modes because of fewer train stations compared to metro stations and bus stops.

National Transport Model

During recent years the Danish National Transport Model (NTM) has been developed (see for example Rich et al., 2010 for a description of the overall models in the NTM). The model has amongst others modules for demand modelling and route choice in both private, public and freight transportation. By use of the TU survey data, public transport traffic counts and other data sources the authors estimated preferences for public transport route choice. The NTM model version 1.05 has parameters as presented in Table 6-8.

Table 6-8: Selected parameter estimates from Nielsen and Johansen (2012) scaled to bus IVT (=1.0).

	Trip purpose		
Parameter	**Commute**	**Business**	**Leisure**
In-vehicle time			
Bus	1.0	1.0	1.0
Local Train	0.9	0.9	0.9
Metro	0.7	0.7	0.7
S-train	0.7	0.7	0.7
IC-train	0.9	0.9	0.9
Regional train	0.9	0.9	0.9
Access/Egress time	1.1	1.0	1.1
Hidden waiting time	0.5	0.5	0.6
Transfer			
Waiting time	1.1	1.0	1.1
Walking time	1.1	1.0	1.1
Penalty	4.0	4.0	6.0

Relative to bus IVT all other mode in-vehicle parameter estimates are lower so the travellers prefer the train modes over the bus. The access/egress time is a little worse than the bus time and so are the transfer waiting and walking attributes. The penalty for transferring is from 4-6 minutes meaning that the traveller would accept the difficulties of transferring to save 4-6 minutes in the bus if no transfer waiting and walking time were included in the transfer. The transfer waiting and walking time is added to the inconvenience of travelling.

6.2 Data and Methods

6.2.1 Statistical analysis of the generated choice sets

In the following the data used for the public transport route choice model estimation is described. The data is output from the choice set generation explained in Chapter 5. In this chapter is used only results from a setup equal to the ErrCompErrTerm with the highest variances. When running the assignment model used for the choice set generation the data is written to csv.-files for each iteration. The first 100 iterations are used and this gives 100 alternative routes for each OD pair. However many alternative routes are identical and the output data is therefore cleaned to only contain the unique routes.

In this section 5,767 OD combinations are used (the final estimations use 5, 461 since the routes with the longest access/egress trips are sorted out according to the findings in Chapter 3). Among these choice sets the largest has 82 alternative routes and the smallest has 2 routes. All choice sets have a mean of 37 alternatives and a median of 38. Figure 6-1 shows the cumulative distribution of the sizes of the choice sets. Only 0.4% of the observations have a choice set of less than 10 routes, and 15% have 50 or more routes. 90% of the choice sets have between 20 and 60 routes.

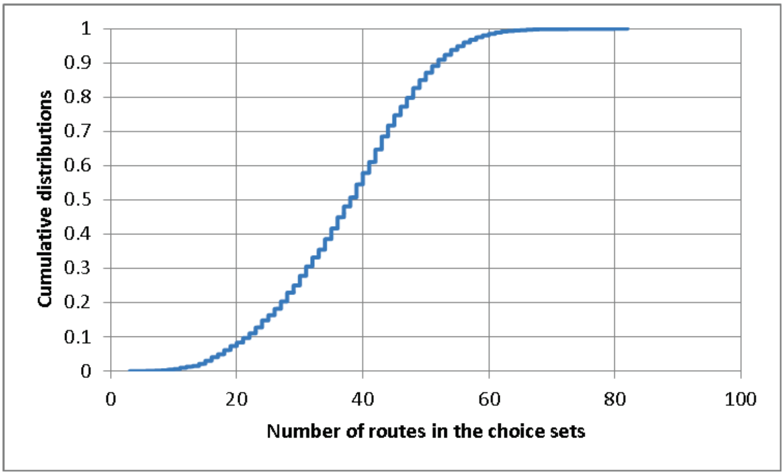

Figure 6-1: Cumulative distribution of the choice set size, 5,767 OD pairs.

Hoogendoorn-Lanser (2005) and Hoogendoorn-Lanser and van Nes (2006) used a rule-based approach and a diachronic-graph representation of the multi-modal transport system to generate route choice sets for 189 observations. These objective choice sets had an average of 48 alternatives (median 39) and for the traveller with the largest choice 278 alternative routes were generated. They limited the choice set to 50 alternatives for computational reasons.

In the following graphs, the cumulative distribution of specific variables is investigated and compared for the observed routes (from TU) and the generated routes (from the choice sets). The comparisons present a statistical overview of travel distance, in vehicle time and number of trip legs. It is important to take into consideration that this is merely a summary of the data and that for the choice sets the number of generated routes affects the graphs a lot.

Figure 6-2 shows the travel distance for the observed trips from the TU Survey data as well as for the generated routes in the choice sets. There is a higher share of the shortest routes in the choice sets but the total trip distance of the observed trips increases faster from 5 km and up. This implies as expected that the travellers choose among the shortest routes in their choice sets but also other factors play a role. 79% of the observed trips are 20 km and shorter which is the case for only 65% of the sampled routes. This shows that travellers in public transport travel (among other things) do try to minimise the distance.

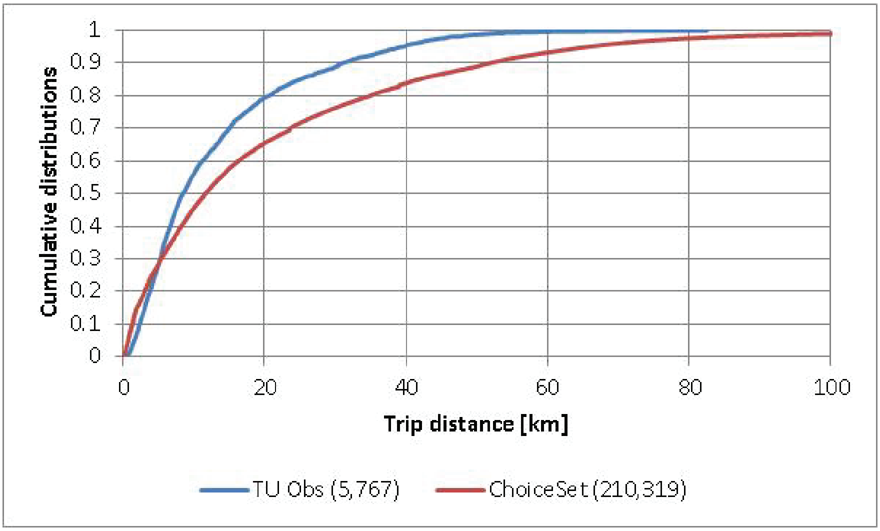

Figure 6-2: Cumulative distribution of the trip distances in observed TU trips and in generated choice sets.

The trip length and the travel time for public transport trips are often strongly correlated and in Figure 6-3 the cumulative distributions for the observed and generated routes are compared.

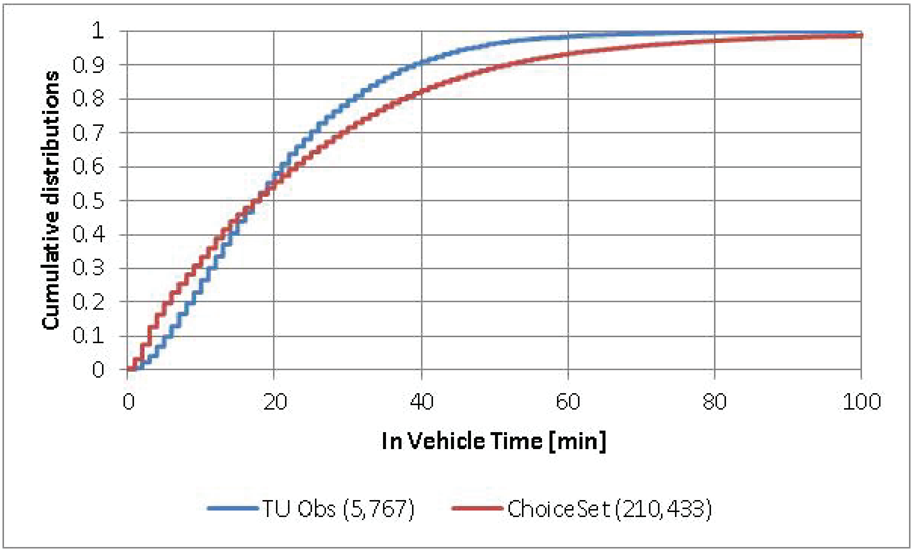

Figure 6-3: Cumulative distribution of the in vehicle travel time (in public transport vehicles) in observed TU trips and in generated choice sets.

Also for the in vehicle travel time a higher share of sampled routes are short. The set of observed routes and the generated routes both have 50% of the trips shorter than 18 min. Over 18 min, the share of observed routes increases faster than the generated routes and the sampled routes are also longer than the observed routes.

The graph shows that 90% of the observed trips are 40 min or below while this is the case for 82% of the trips in the choice set. Often the traveller chooses among the shorter routes and it is possible to find routes in the network which has a significantly higher in vehicle time than the chosen one. Since the graphs are rather close to each other the alternatives sampled in the choice set are not too unrealistic but presents alternatives which has a lower utility than the chosen route.

Figure 6-4 shows the number of vehicular trip legs for the set of observed and generated trips. 60% of the observed trips use one public transport line only for the trip whereas this is only the case for 47% of the generated trips in the choice set. The maximum number of trip legs is 5 for the observed trips and 9 for 99.9% of the generated trips. The last 0.1% of the observed trip has from 8-31 different lines used to complete a trip within the Greater Copenhagen Area. These trips are most likely never chosen but they play an important role in the model estimation.

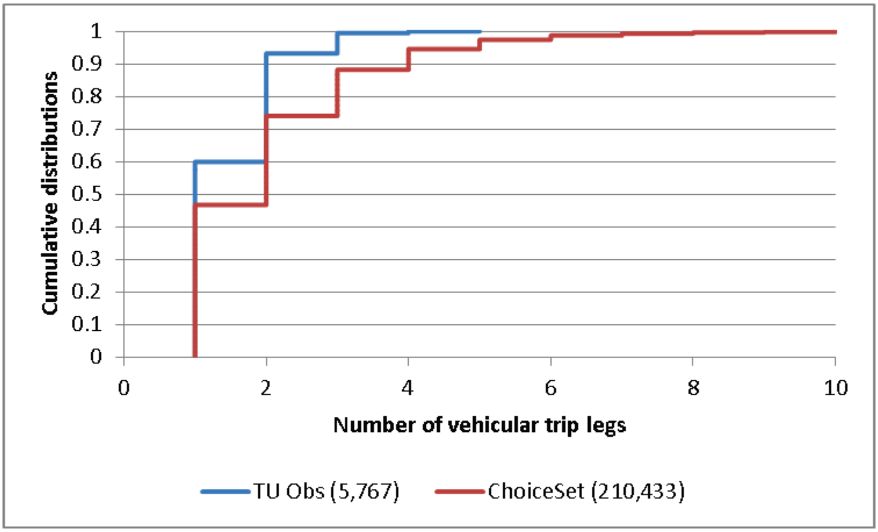

Figure 6-4: Cumulative distribution of the number of vehicular trips in observed TU trips and in generated choice sets.

From the graph can be noticed that the travellers in 2/3 of the trips use only one leg and thereby no transfers and another 33% have one transfer (two legs). This points to the fact that travellers try to minimize the number of transfers. They are presented with a variety of routes but choose among those with few transfers (and in many cases also short travel time). The graph shows that it is possible to find routes with a large number of transfers of which some are more relevant for the traveller than others.

6.2.2 Data in choice sets

The following presents the data in the generated choice set, which is the output from the traffic assignment model.

Walking and biking

In the public transport network, the traveller has several connectors from the point of origin/to the point of destination, describing the access to and egress from the public transport services. The public network is always accessed using a private mode, e.g., walking, biking or using a car (private car as driver or passenger, taxi passenger, etc.). The travellers are assumed to use either bicycle or walking to reach the first stop on the trip. As described in the last chapter connectors are created to the bus stops and the train stations closest to the origin, the distance is calculated by finding the shortest path in a path and road network, and the connector speed is calculated as described in chapter 3.

In the choice set the distance travelled and the timed used on the connector is defined and are used for the estimation.

Car

In the TU data some travellers use the car to get to the train and a few uses the car to access a bus. When the car is used as access mode to a bus the traveller is most often passenger in a private car and the choice of the bus stop is most likely very dependent on the route and destination of the car driver. Most likely the car passenger is dropped off at a bus stop where it is convenient to stop the car and this is not necessarily the stop closest to home. For many trips the car leg is the longest in time and distance and the public transport leg(s) is the shortest, probably used to avoid a detour for the car driver. The choice of access/egress points to the public transport network is difficult to reproduce and therefore the trips with car as access mode are not used in the model route choice estimations.

Bus and train

In the choice set calculation output the trip is split in information for each LineVariantElement used. Each part of the trip using a new line variant consists of a number of links (LineVariantElements) leading between the stops the public transport lines serve.

Travel time

The travel time is an important measure for public transport travellers and this can be split in in-vehicle times for the different public transport modes. The in-vehicle-time consist of both driving and time used in the vehicle at stops and these elements are both considered to be in vehicle time since the passenger is on board the vehicle.

The travel distance is also listed but this is assumed not to be important for the public transport passenger since only travel time is important to them. Also the travel cost is highly correlated with the travel time and are not used for model estimation.

In the Greater Copenhagen Area the fare depends on the location of the origin and destination and no other measures as time and distance. Since the decision of origin and destination is prior to the route choice the fares are not used as a part of the route choice model estimation.

Transfers

From the output data information about the number of transfers can be derived. A transfer can be uncomfortable for the traveller because he has to disembark the public transport vehicle, perhaps wait outside in all kinds of weather, there could be an issue of uncertainties about getting in time to reach the following public transport mode if short transfer time/delays, etc. On the other hand some might like the transfers time over travelling time because of the break in driving, possibility to catch fresh air, etc. The transfer penalty is additional to the transfer time and can be seen as an extra disutility of transferring (refer to e.g., Nielsen et al., 2001 and Florian, 2004)

Numbering of alternatives

The same number of alternatives is generated for each observation. Afterwards the choice sets are prepared for use in the model estimation by sorting out alternatives which are completely similar. By this the number of alternatives is reduced considerably (from 100 down to two) for some observations and by a small percentage for others (refer to Chapter 5 for further details). The alternatives are renamed with id's beginning from 1 and these numbers are used in the estimation model files. In this way all choice sets have an alternative 1 and 2 (two is the lowest number of alternatives), and only very few have more than 70 alternatives. As can be seen from this alternative 1 for an observation has not necessarily any route attributes overlapping with alternative 1 for another observation.

The route choice sets contain a great amount of information on the alternatives and many different parameters can be created and added to the model specification.

6.3 Route choice model estimation

The following specifies different specifications for the model estimation using the observed route choice data and the generated choice sets. The specifications of suggested random utility models in the above are used as inspiration for the choice of models. First, a multinomial logit is estimated to use for comparison with the more advanced models.

Afterwards a MNL model is estimated including Path Size factor to account for similarities among alternatives. Also a model including the Path Size commonality factor is estimated. Finally mixed logit models are investigated to take the heterogeneity into account.

6.3.1 Parameters included in the model specification

The data available in the route choice sets and the observed data make a number of parameters possible for use in estimation of public transport route choice model, for example:

- TripID.
- Route alternative.
- Time:
  - Walking Time (transfer time).
  - Waiting time (transfer time).
  - Connector walking time.

  - In vehicle travel time.
- Transfers:
  - Number of transfers.
  - Bus->Bus, Train->Train, Train->Bus, Bus->Train.
  - Waiting time at Bus stop/Train station.
- Path Size.
- Headway:
  - o First/Longest
- Transport mode (service type, line variant).

Some parameters are directly transferable from the data but most need some or much data preparation before they can be used as input for the model.

Dummies

In the following the dummies included in some of the parameters in the model specifications are presented and explained.

Mode specific constants

Since alternative 1 in a given choice set most likely shares no attributes with alternative 1 in another choice set, the inclusion of alternative specific constants are not reasonable in the utility functions of the model. Instead mode specific dummies are investigated for whether they are significant in the model. The mode specific dummies are 1 if the mode type is used on the trip alternative and 0 otherwise. The mode specific dummies are specified for the following five transport mode types:

- Bus.
- S-train.
- Regional and IC-train.
- Metro.
- Local train.

This specification allows the dummies to sum to more than one for each alternative but the sum is at least one.

The mode specific dummies can be thought as an assessment of the travellers' attitudes towards the specific mode types. If the coefficient is estimated to be positive the travellers prefer the mode type over the ones with estimates closer to zero or with negative estimates.

The question is which effect the mode type specific dummies have in the utility function. One possible way of understanding the dummies is as a measure of the travellers' willingness to board a specific mode type vehicle, or more precise the reluctance against boarding the vehicle. This measure is not much different from the characteristics of the transfer variable, which will be discussed later.

Service type specific dummies

An alternative to specifying the mode dummies is to include service type specific dummies in the utility function. By service type is referred to the classes of public transport types in this case defined as bus and train. As for mode dummies it is possible for a traveller to have a dummy equal to one for more than one entry - that is if the traveller uses both bus and train. All travellers have at least one service type specific dummy equal to one.

The service type specific dummy is a measure for the willingness/reluctance to board a bus or a train. The estimates of the coefficients can be compared to obtain an indication of which service type the travellers prefer. In this way the dummies can be used as a measure for the so-called Rail Factor – that is a preference for using rail assuming all service factors being equal. Axhausen et al. (2001) referred to this as the rail bonus and derived it considering behavioural changes when upgrading the public transport system in Dresden. Scherer and Dziekan (2012) found a percentage measure for the rail factor through studies in Germany and Switzerland.

As before these dummies can also be assessed as a transfer penalty and the estimates of the service type dummies can therefore indicate several effects.

Time measures

In the following the parameters related to time used for the different model setups are explained.

In-Vehicle Time - mode types

In the models, the public transport modes are split in five mode types mentioned above. The preliminarily tested models split the modes into seven modes types; the five mentioned above, and Bus split in A-bus, E- and S-bus and Other bus but the effect of this was very low. For model purposes, the in-vehicle time (IVT) for each of the mode types are summed for each route alternative.

The in-vehicle time is the time spent in the vehicle and includes the time from boarding to alighting. The in-vehicle time includes both driving time and time spent at stops and stations in the vehicle.

The in-vehicle times in the various mode types can be assessed differently by the travellers. The difference between bus and train is obvious but also differences between train types might occur. The trains have differences in the stop frequencies, the comfort of the driving, seating, regularity, and so forth and the travel time might be assessed differently by the travellers.

In-Vehicle Time – Public Transport service

For investigation purposes the above mentioned IVT for mode types are summed into two public transport service categories: bus and train. The IVT for the public transport services are used to include IVT in models with the mode dummies, not to have full correlation between mode dummy and in-vehicle time measure.

Access/egress time

The time spent on the connectors is defined as the access and egress time. The connectors lead from the origin point of the trip to the entrance point to the public transport network (bus stop or train station) and opposite at the destination end of the trip. The length of each connector is found using a path and road network and defined as the shortest path through the network.

Several private modes can be used for the access and egress parts of the trip and the exact chosen mode is not taken into consideration in the section. The observations included in the model estimation contain both walking and biking as access/egress modes. Cars are not included since it is, as previously mentioned, very difficult to define the choice of entrance point to the public transport network when using the car as access mode. The time (length) is not the most important factor but parking and kiss and ride facilities play an important role. When including observations with car access/egress legs the car availability is important and should be modelled separately. This is an important task and could be looked into in further work and is not a part of this thesis.

Walking Time (between stops)

Walking trip parts are located at different points in the network. The travellers always walk from their point of origin to the first stop/station in the public transport network and from the last stop/station to their destination. Further the travellers can walk between two stops/stations and mentioned above.

The walking times between points in the public transport network are defined in this category since the characteristics of the walking are closely related to each other.

The walking is defined in terms of time to be comparable to the in-vehicle times.

Waiting Time

The waiting time variable refers to the time spent at a stop or a station waiting for the bus/train to arrive. The waiting time does not get the traveller closer to the destination and is just extra time added to the total travel time so travellers could be assumed to try to minimize this. On the other hand is the waiting time might not assessed as negatively as the walking time since the traveller is at the point where the public service departs from and is therefore not unsure if he reaches the public transport service in time (as could be the case when walking to/between/from stops/stations.

Hidden waiting time

The waiting time in this model only covers waiting times at stops in the public network. In real life there will also be waiting times (called hidden waiting time) at origin and destination. At the origin, the traveller might have to wait longer than his preferred time of departure because of the public transport time schedules not matching his desires completely. At the destination, the traveller might arrive before the desired time of arrival in order to be there on time.

The hidden waiting time is not part of this model since the corresponding data is not collected for the trip observations. In order to get such information the traveller would have to be asked specifically when he wanted to leave the point of origin and that might not be easy for the traveller to answer. This would bias the entry of the hidden waiting time in the model. At the origin, the question might be easier to answer if for examples a student or an employee has to be at school or work at a fixed time.

In the sampled alternatives a measure of the waiting time in the origin zone is registered but since there is no such information for the observed trip the desired departure time does not build on any reliable data. In the choice set generation the start time of a trip is defined using a 5 minute interval before and after the departure time stated for the observed trip. The departure time is defined as a launch at a random point in time within this time interval.

It is difficult to know whether the stated time of departure is in fact the desired departure time so stating this in the model estimation could bias the model.

Trip purposes

To take into consideration heterogeneity between various classes of travellers the observations are divided into groups by trip purposes. The defined purposes are:

- Work (comparing with the trip purposes defined in Chapter 4, Table 4-2 this group includes commuters, business travellers and trip for educational purposes).
- Leisure and other (leisure, errands, and other).

The definition is based on the trip end purposes and not on the trip purposes as in Table 4-2 which redefine all home purposes to the other categories.

A dummy defining the trip purpose could be specified but such a dummy would not vary over the alternatives for a specific traveller since the trip purpose is the same for all trips for one observation. Instead purpose specific dummies and variables are defined in combination with mode specific dummies and in-vehicle travel time for modes.

Estimating using the trip purpose specific variables provides an indicator of the behaviour of the travellers in each specific category. In a multinomial logit model the travellers in each group are all assumed to have the preference for mode (dummy or in-vehicle time) as the model estimates. In a mixed logit model the tastes of each traveller group for a specific mode are assumed to follow the specified distribution with the estimated mean and variance. Both mean and variance (and distribution) can vary between trip purposes. For example it is assumed that business travellers have a higher disutility for in-vehicle time than leisure travellers.

Hensher (2001) explained the presence of individual specific random effects (heterogeneity) for the value of travel time, and Vrtic and Axhausen (2002) confirmed this by showing that price parameters were higher for the commuters than for the other three defined purposes (shopping, leisure, and business).

The mean and variance estimated for the travellers in each trip purpose is estimated the travellers tastes for a mode are assumed to follow the distribution

Mode and trip purpose specific dummies

Mode specific dummies are created for each of the seven modes in combination with the four trip purposes, so 24 mode and purpose specific dummies are created. The estimated coefficients can show whether travellers from different groups have the same preference for each defined mode group.

As before the mode specific dummies can either be perceived as an attraction measure or as an additional transfer penalty.

In-Vehicle Time - Mode and trip purpose specific

Also in-vehicle times are calculated for combinations of transport mode type and trip purpose giving 24 combinations.

In general travellers with some purposes are thought to have higher IVT than others. Business travellers are very restricted on their time and therefore put high values on time measures. Leisure travellers are perhaps more willing to travel longer if this results in a more spectacular route, etc.

Trip distance

For some model specifications the importance of the length of the trip is tested. The trip distance can influence the choice of transport modes, the transfer attributes etc. Maybe the traveller prefers specific public transport modes over others for the longer trips; maybe the traveller prefers longer travel time to avoid transferring or the opposite.

Also the traveller might perceive the in-vehicle time in the various public transport mode IVT's differently depending on the length of the trip. Often the Regional and IC-trains are used for longer trips between cities and travellers on shorter trips tend to avoid these. The stop patterns for the train types are very different (also refer to map in figure 2-4):

- The metro serves the CBD of Copenhagen and have stations for every 900 meters.
- The S-train serves the city and suburbs (station for every 2 km).
- The local rails connect the suburbs (stations for every 1.6 km).
- The regional and intercity trains have few stops in the city and serve other larger cities at the Zeeland (stations every 6 km).

When testing the importance of trip distance the observations are divided in distance bands. The distance bands tested are:

- <10km, >=10 km.
- <10, 10-25km, >=25 km.
- <20 km, >=20 km.

The number of observations in each of the data sets is presented in Table 6-16 in the result section.

Transfers

Transfers are thought to be very important for public transport travellers (van der Waard, 1988 and others). In the model estimation various formulations of the transfer disutility are tested.

Number of transfers

In a multimodal network transfers are of great importance. The transfers in this chapter are defined as transfers between two public transport lines. The transfer can have different characteristics such as:

- Transfer between two buses serving the same stop.
- Transfer between two buses serving different stops which are within walking distance.
- Transfer between two trains serving the same platform at a train station.
- Transfer between two trains serving the different platforms at the same train station.
- Transfer between buses and trains serving different stops and train stations within walking distance.

The possible/preferred walking distance is defined by the traveller and differs according to the characteristics of the traveller. In the sampled route a traveller can change between stops and stations connected by a transfer link, and the travel time on the transfer link is considered when the traveller chooses his optimal route.

Transfers are often thought of to be uncomfortable (Vrtic and Axhausen, 2002) and often travellers try to avoid transfers. The traveller determines a trade-off between longer in-vehicle/longer waiting time at stops/etc. and the disutility connected to transferring.

Vrtic and Axhausen (2002) showed that the number of transfers is more important than the transfer time but emphasise the importance of separating the two.

The number of transfers is used as a part of the models to estimate the mentioned trade-offs by comparing the disutility of transfer to the other components of the utility function. In this thesis the type of transfers is tested in the utility function either as the number of transfers (disrespecting the type) or as depending on from and to service type as described below.

Transfers between public transport mode types

The type of transfer can also be important to the perception of the transfer and models including variables explaining which transport modes the traveller transfers between are also estimated. Chapter 2 presented the statement that travellers prefer transfers at high-order bus stops and train stations over low-order stops, both because of frequency and because of the shelter from rain and wind.

The public transport modes are divided in the two service types: Trains and Buses and four categories are defined, as follows:

- Transfer from Bus to Bus.
- Transfer from Bus to Train.

- Transfer from Train to Bus.
- Transfer from Train to Train.

This can show whether the travellers prefer to stay on the same network (rail or road) and how strong the preference for one public transport service type is compared to the other. Transferring from train to bus will often mean that the traveller has to change level if the rails or road in on a bridge passing the other. A transfer between two buses is often (but not always) at the same level but often the traveller has to walk to another bus stop and a bus-bus transfer can be very affected by the regularity of the buses if one or both of the buses has a low frequency.

Transfer location

When transferring between two public transport lines the importance of the station type is also tested for in the route choice model estimation. The transfer locations are split according to whether the traveller waits at a train station or at a bus stop to board the next public transport mode in his trip chain. Some travellers may prefer to wait at a train station because these often have shops, are well-lit, have other travellers waiting, etc. Also train stations are less exposed to the weather because they more often have a roof over and other weather covers.

Headway

The importance of the headway is also tested. The headway for the number of minutes between two departures of a specific line can be important for the traveller since this may determine when it is most optimal to leave the point of origin or arrive at destination. This point in time may not match the preferred time exposing the traveller to hidden waiting time.

The headway is tested in two definitions:

- First headway – headway for the first public transport line used in the given route alternative.
- Highest headway - headway for the public transport line with the highest headway used in the given route alternative.

The *first headway* is the headway for the first public transport line used for a trip. This may be important for the hidden waiting time at the origin and impacts when it is possible to reach the next vehicle if several modes are used.

The *highest headway* is the headway for the line on the trip which has the lowest number of departures in the time of travel. If a traveller uses several modes along the trip and therefore has transfers included in his trip the frequency can be of great importance in defining the chosen route alternative. If the traveller is travelling by a high frequency train line as the main mode and uses a low frequency bus line from the train station to the destination point the frequency of the low frequent bus mode could be of significant importance to the traveller’s other choices along the route. The low frequent line can be critical for the start time of the trip and for the hidden waiting time at the origin.

For both First and Highest Headway the variable enters the utility function as half the headway. If the travellers are assumed to arrive at a stop or station following a uniform distribution the average hidden waiting time will be half the headway.

Additionally variables are created to test whether the travellers perceive an additional minute of headway differently for short and long headways. The headway is split at 6 minutes since vehicles with a headway of up to 6 minutes defined by the printed schedules to be arriving very often. As presented in chapter 2 the A-busses have headways up to 6 minutes in the peak hours and the schedules for these often inform the arrival time as “Every 6th minute” and not the exact arrival time. Two headway variables are made to represent:

- Headway up to 6 minutes.
- Headway exceeding 6 minutes.

6.4 Model estimations

In this section the development and estimation of the various models for route choice in the public transport network are defined and estimated. Not all model results are described, but the most important results will be shown.

In the result tables the estimated parameter estimates are presented alongside with the t-test to show whether the estimated values are significantly different from zero. The estimated parameters are also presented as scaled to Bus in-vehicle time. No fare payment is part of the models because of the fixed fare system and therefore it is not possible to estimate value-of-time parameters for the models. The scaled parameters, however, are a valuable measure for the preferences and for their relationship within and are used as a foundation for the discussions of the route choice preferences.

6.4.1 Modelling results concepts

The public transport route choice models are estimated using the research freeware “BIerlaire Optimization toolbox for GEv Model Estimation” (BIOGEME 1.8).

The basic model results are described in the following. Bierlaire (2009) described the contents of the report file from BIOGEME containing the results of the maximum log-likelihood estimation of the model.

Null log-likelihood is the log-likelihood of the sample for a discrete choice model with all *β* parameters being 0 and is calculated as:

$$\mathcal{L}_0 = \sum_{n \in sample} \omega_n ln \frac{1}{\# C_n} \quad \textbf{(6-27)}$$

Where $\#C_n$ is the number of alternatives available to traveller n and ω_n is the weight.

Init log-likelihood is the initial log-likelihood of the sample for the model defined in the model file.

Final log-likelihood is the log-likelihood of the sample for the estimated model.

The *Likelihood ratio test* is

$$-2(\mathcal{L}^0 - \mathcal{L}^*) \tag{6-28}$$

Where $\mathcal{L}^0$ is the log-likelihood for the sample for a discrete choice model where all β parameters are 0 and $\mathcal{L}^*$ is defined by the log-likelihood of the sample of the estimated model.

Rho-square is

$$\rho^2 = 1 - \frac{\mathcal{L}^*}{\mathcal{L}^0} \tag{6-29}$$

And the *Adjusted rho-square* is

$$\rho^2 = 1 - \frac{\mathcal{L}^* - K}{\mathcal{L}^0} \tag{6-30}$$

Where *K* is the number of parameters estimated. ρ^2 is used to normalise for the increasing of the log-likelihood when adding parameters to the model.

Log-likelihood is a measure of how well the model describes the data and the closer it is to zero the better the model performs. Adjusted rho-square increases the better the model describes the data. The log-likelihood and the adjusted rho-squared describe how the model overall perform.

Also, results for the estimated parameters are reported. The *estimated value* of the parameter describes how the corresponding variable affects the choice. The magnitude and the sign of the estimated parameter are important and the sign shows whether a change in the variable value causes an increase or decrease of the choice probability. The *standard error* indicates the preciseness of the parameter estimate. A high standard error is either caused by the fact that there are too few observations of the variable or problems with the identification of the model. The *t-value* shows the results of a t-test for the significance of the parameter estimate. The value shows the test of whether the parameter is different from a known value, often zero. The significance can be checked at different levels and the critical values are the percentiles of a standardised normal distribution.

6.4.2 MNL

The multinomial logit model is estimated in different setups with the above mentioned parameters.

Mode specific models

Three models using mode specific dummies are reported here. One model includes mode specific dummies, wait time, walk time, and transfer penalty. The deterministic term of the utility function is for traveller *n* for alternatives $k \in 1,..,K$ for the choice set C_n with *K* alternatives given by:

$$V_{kn} = \beta_{Mode,Bus} \cdot Mode_{Bus,kn} + \beta_{Mode,LocalTrain} \cdot Mode_{LocalTrain,kn} + \beta_{Mode,Metro} \cdot Mode_{Metro,kn} + \beta_{Mode,STrain} \cdot Mode_{STrain,kn} + \beta_{Mode,RegIC} \cdot Mode_{RegIC,kn} \tag{6-31}$$

Where $Mode_{xxx,k}$ is the mode dummy for mode type *xxx* in alternative *k* and $\beta_{Mode,xxx}$ is the coefficient for mode *xxx* to be estimated.

The second model also includes the transfer specific attributes waiting time, walking time and number of transfers:

$$V_{kn} = \beta_{Mode,Bus} \cdot Mode_{Bus,kn} + \beta_{Mode,LocalTrain} \cdot Mode_{LocalTrain,kn} + \beta_{Mode,Metro} \cdot Mode_{Metro,kn} + \beta_{Mode,STrain} \cdot Mode_{STrain,kn} + \beta_{Mode,RegIC} \cdot Mode_{RegIC,kn} + \beta_{WalkTime} \cdot T_{Wait,kn} + \beta_{WaitTime} \cdot T_{Wait,kn} + \beta_{ChangePen} \cdot N_{Change,kn} \tag{6-32}$$

Where the $T_{WalkTime}$ and $T_{WaitTime}$ are walking times between stops/stations and waiting time at stops/stations, respectively. The N_{Change} is the number of changes.

The third model also has *IVT* for aggregated service types: bus and train.

$$V_{kn} = \beta_{Mode,Bus} \cdot Mode_{Bus,kn} + \beta_{Mode,LocalTrain} \cdot Mode_{LocalTrain,kn} + \beta_{Mode,Metro} \cdot Mode_{Metro,kn} + \beta_{Mode,STrain} \cdot Mode_{STrain,kn} + \beta_{Mode,RegIC} \cdot Mode_{RegIC,kn} + \beta_{IVT,QBus} \cdot IVT_{QBus,kn} + \beta_{IVT,QTrain} \cdot IVT_{QTrain,kn} + \beta_{WalkTime} \cdot TT_{Wait,kn} + \beta_{WaitTime} \cdot TT_{TWait,kn} + \beta_{ChangePen} \cdot N_{Change,kn} \tag{6-33}$$

The models have significant values for almost all the estimated coefficients except for the bus dummy in model formulation (6-33).

Table 6-9: Estimated parameter coefficients scaled to Metro dummy (=1.0) and (robust t test) for the MNL models including mode dummies.

	MNL Model		
	Mode dummies	**Dummies + Transfer attributes**	**Dummies + Transfer +IVT**
Parameter	(6-31)	(6-32)	(6-33)
Mode Dummies			
Bus	-0.875 (-23.0)	0.108 (1.48)	-0.031 (-0.41)
Local Train	0.323 (3.81)	1.290 (7.87)	1.060 (5.43)
Metro	0.346 (8.06)	2.760 (32.0)	2.790 (31.4)
Regional + IC-train	-0.558 (-10.8)	1.480 (14.3)	1.060 (8.50)
S-train	0.809 (22.2)	2.560 (33.4)	2.630 (30.8)
In Vehicle Time			
Bus	-	-	-0.070 (-23.9)
Train	-	-	-0.081 (-20.8)
Transfer			
Waiting Time	-	-0.075 (-12.9)	-0.077 (-13.1)
Walking Time	-	-0.281 (-58.2)	-0.326 (-58.4)
Number of transfers	-	-2.530 (-45.8)	-2.360 (-42.1)
Number of estimated parameters:	5	8	10
Number of observations:	5,641	5,641	5,641
Null log-likelihood:	-20,172	-20,172	-20,172
Final log-likelihood:	-19,028	-12,019	-11,490
Likelihood ratio test:	2,287	16,306	17,362
Adjusted rho-square:	0.056	0.404	0.430

The model only including the mode dummies (formula (6-31)) has a very poor model fit and the model fit is improved considerably by adding the transfer attributes to the model. Since the metro dummies obtain positive parameter estimates in the above mentioned models the travellers are attracted to the metro mode. In the first model the attraction to the S-train is even higher (2.3 times) and when adding transfer attributes to the model the estimates are close to each other. In the mode dummies only model the local train also has a positive parameter estimate but the bus and regional train have negative parameter dummies and are therefore not assessed as attractable.

In the model considering mode dummies and transfer attributes (formula (6-32)) the estimates show that travellers prefer using modes over transferring since all mode dummies are positive and all transfer attributes are negative.

Adding parameters for in-vehicle times for bus and train (all transport modes on rail together – formula (6-33)) some of the explanatory power is removed from the mode dummies and the bus dummy is no longer significant.

The parameters for transfer waiting time and walking time are negative in both models but a minute of walking is perceived worse than a minute of waiting. According to these models the most important transfer attribute is the number of transfers. With these model specifications a transfer is perceived approx. 10 times worse than a minute of walking, approx. 30 times worse than a minute of waiting at transfer locations and up to 34 times worse than train/bus in-vehicle time. This implies that the travellers would rather wait 30 minutes at a stop/station or travel 34 minutes by bus than transferring between two modes. This is not comparable to the literature and to how the travellers are expected to behave.

The coefficient estimate for the transfer penalty (number of transfers) is at the same level as the mode dummies but with a negative sign which means that these cancel out each other. The mode dummy is assigned a value when using the specific mode so if a traveller uses both the metro and the train the transfer to train would be penalised but the penalty would be cancelled out by the train dummy. This is not intuitively correct and therefore different model setups are considered in the following sections.

Mode Specific IVT models

Since the models including mode dummies show counterintuitive results the following models consider instead mode specific in-vehicle time. In this section, five models including in-vehicle time for each mode are reported. The first model contains only *in-vehicle time* for each of the defined public transport modes.

$$\begin{aligned} V_{kn} &= \beta_{IVT,Bus} \cdot IVT_{Bus,kn} + \beta_{IVT,LocalTrain} \cdot IVT_{LocalTrain,kn} \\ &+ \beta_{IVT,Metro} \cdot IVT_{Metro,kn} + \beta_{IVT,STrain} \cdot IVT_{STrain,kn} \\ &+ \beta_{IVT,RegIC} \cdot IVT_{RegIC,kn} \end{aligned} \qquad \textbf{(6-34)}$$

Where the β's are parameters to be estimated. The IVT_{xxx}'s are In-Vehicle Time for the various public transport modes (Bus, Local train, Metro, Regional and InterCity train, and S-train, see Chapter 2 for further explanation of the public transportation modes).

The second model includes the *access time* from origin to the first public transport mode and the *egress time* from the last public transport mode to the destination (Access/Egress time).

$$\begin{aligned} V_{kn} &= \beta_{IVT,Bus} \cdot IVT_{Bus,kn} + \beta_{IVT,LocalTrain} \cdot IVT_{LocalTrain,kn} \\ &+ \beta_{IVT,Metro} \cdot IVT_{Metro,kn} + \beta_{IVT,STrain} \cdot IVT_{STrain,kn} \\ &+ \beta_{IVT,RegIC} \cdot IVT_{RegIC,kn} + \beta_{Access/egress} \cdot TT_{Access/egress,kn} \end{aligned} \qquad \textbf{(6-35)}$$

The third formulation also contains *waiting times* at stops and stations and *walking time* between two stops/stations when transferring between two public transport modes.

$$\begin{aligned} V_{kn} &= \beta_{IVT,Bus} \cdot IVT_{Bus,kn} + \beta_{IVT,LocalTrain} \cdot IVT_{LocalTrain,kn} \\ &+ \beta_{IVT,Metro} \cdot IVT_{Metro,kn} + \beta_{IVT,STrain} \cdot IVT_{STrain,kn} \\ &+ \beta_{IVT,RegIC} \cdot IVT_{RegIC,kn} + \beta_{Access/egress} \cdot TT_{Access/egress,kn} \\ &+ \beta_{WalkTime} \cdot TT_{Walk,kn} + \beta_{WaitTime} \cdot TT_{Wait,kn} \end{aligned} \qquad \textbf{(6-36)}$$

The $TT_{WalkTime}$ and $TT_{WaitTime}$ are walking times between stops and waiting time at stops, respectively.

Fourthly also the *number of transfers* is considered as part of the utility function:

$$\begin{aligned} V_{kn} &= \beta_{IVT,Bus} \cdot IVT_{Bus,kn} + \beta_{IVT,LocalTrain} \cdot IVT_{LocalTrain,kn} \\ &+ \beta_{IVT,Metro} \cdot IVT_{Metro,kn} + \beta_{IVT,STrain} \cdot IVT_{STrain,kn} \\ &+ \beta_{IVT,RegIC} \cdot IVT_{RegIC,kn} + \beta_{Access/egress} \cdot TT_{Access/egress,kn} \\ &+ \beta_{WalkTime} \cdot TT_{Walk,kn} + \beta_{WaitTime} \cdot TT_{Wait,kn} \\ &+ \beta_{TransferPen} \cdot N_{Transfer,kn} \end{aligned} \tag{6-37}$$

Where $N_{Transfer}$ is the number of transfers.

The fifth MNL model including IVT for public transport modes also includes *service type specific dummies* for bus and train.

$$\begin{aligned} V_{kn} &= \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\ &+ \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\ &+ \beta_{IVT,RegIC} * IVT_{RegIC,kn} + \beta_{Access/egress} \cdot TT_{Access/egress,kn} \\ &+ \beta_{Mode,QBus} * Q_{Bus,kn} + \beta_{Mode,QTrain} * Q_{Train,kn} + \beta_{WalkTime} * TT_{Walk,kn} \\ &+ \beta_{WaitTime} * TT_{Wait,kn} + \beta_{TransferPen} * N_{Transfer,kn} \end{aligned} \tag{6-38}$$

Where Q_{Bus} and Q_{Train} are mode dummies for all buses and all trains respectively.

The parameter estimates for the five MNL models are presented in Table 6-10 and the parameter coefficients scaled to Bus IVT (=1.0) are presented in Table 6-11.

Table 6-10: Estimated parameter coefficients and (robust t test) for the MNL models including in-vehicles times.

Parameter	IVT (6-34)	+ Acc/Egr Time (6-35)	MNL model + Wait & Walk Time (6-36)	+ No. Transfers (6-37)	+ Mode Const (6-38)
In Vehicle Time					
Bus	-0.021 (-17.6)	-0.133 (-41.0)	-0.173 (-41.3)	-0.153 (-34.3)	-0.120 (-27.9)
Metro	-0.001 (-0.71)	-0.114 (-21.2)	-0.105 (-16.7)	-0.025 (-3.18)	-0.035 (-4.06)
Local train	0.016 (6.13)	-0.107 (-14.4)	-0.130 (-14.7)	-0.112 (-10.9)	-0.134 (-12.7)
Regional + IC-train	-0.001 (-0.51)	-0.148 (-24.9)	-0.161 (-20.3)	-0.134 (-15.9)	-0.157 (-18.1)
S-train	0.014 (18.4)	-0.100 (-29.3)	-0.111 (-25.1)	-0.085 (-16.9)	-0.109 (-20.6)
Access/Egress	-	-0.166 (-60.9)	-0.292 (-58.5)	-0.368 (-55.7)	-0.376 (-55.4)
Mode Dummy					
Bus	-	-	-	-	-1.080 (-13.5)
Train	-	-	-	-	0.723 (7.36)
Transfers					
Waiting Time	-	-	-0.140 (-16.8)	-0.084 (-13.8)	-0.083 (-13.7)
Walking Time	-	-	-0.507 (-32.7)	-0.136 (-10.4)	-0.078 (-6.04)
No. Transfer	-	-	-	-2.140 (-43.7)	-2.050 (-41.2)
Number of estimated parameters	5	6	8	9	11
Number of observations	5,641	5,641	5,641	5,641	5,641
Null log-likelihood	-20,172	-20,172	-20,172	-20,172	-20,172
Final log-likelihood	-19,812	-16,420	-13,538	-11,947	-11,637
Likelihood ratio test	719	7,503	13,267	16,450	17,068
Adjusted rho-square	0.018	0.186	0.328	0.407	0.423

Only describing the choice of a route using public transport mode IVT (formula (6-34)) the model fit is very poor. The bus is assessed as the worst mode (low, negative parameter coefficient). The parameter for local train and S-train IVT have a positive sign meaning that local train and S-train are so attractively that the traveller prefers to travel as long as possible in the given modes. The parameter for *metro IVT* is not significant.

When adding the *access/egress time* (walking, biking and shorter car trips to and from stops and stations) to the model (formula (6-35)) the model fit increases considerably. All parameter estimates are negative as expected and significantly different from zero. In this model access and egress time are thought of as the most onerous time spent on the route. The *S-train IVT* is preferred over the other time attributes.

Adding waiting and walking time at transfers to the model (formula (6-36)) further improves the model fit. All IVT's for public transport modes loose importance relative to *bus IVT* (compared to model formulation (6-35)) and the parameter for *access/egress time* increases relative to *bus*

IVT. The *transfer walking time* is the most onerous (travellers would rather travel 3 minutes longer in a bus to avoid 1 minute of *transfer walking*).

Table 6-11: Estimated parameter coefficients scaled to Bus IVT (=1.0) for the MNL models including in-vehicles times.

			MNL Models		
	IVT	+ Acc/Egr Time	+ Wait & Walk Time	+ No. Transfers	+ Mode Const
Parameter	(6-34)	(6-35)	(6-36)	(6-37)	(6-38)
In Vehicle Time					
Bus	1.0	1.0	1.0	1.0	1.0
Metro	-	0.9	0.6	0.2	0.3
Local Train	-0.7	0.8	0.8	0.7	1.1
Regional + IC-train	-	1.1	0.9	0.9	1.3
S-train	-0.7	0.8	0.6	0.6	0.9
Access/Egress	-	1.2	1.7	2.4	3.1
Mode Dummy					
Bus	-	-	-	-	9.0
Train	-	-	-	-	-6.0
Transfers					
Waiting Time	-	-	0.8	0.5	0.7
Walking Time	-	-	2.9	0.9	0.7
No. Transfer	-	-	-	14.0	17.1

When the *transfer penalty* is added to the model (formula (6-37)) the attribute is assigned a value 14 times greater than the *bus IVT*. This means that travellers would travel 14 minutes longer by bus to avoid a transfer. Adding the *transfer penalty* increases the negative perception of the *metro IVT* (relative to *bus IVT*) but the value is still very low and the size of the value suggests that a traveller would rather travel 5 minutes in the metro than 1 minute in a bus. The importance of the *transfer walking time* parameter is decreased when adding the *transfer penalty* and the *access/egress time* is even more important when the *transfer penalty* is part of the model.

Finally, adding a *mode dummy* for the bus and one for the train (local train, S-train and regional + IC-train) the positive parameter estimate shows that travellers prefer train over all other variables. However, the dummies affect the *train IVT* relative to *bus IVT* so that the IVTs for all the train modes increase (relative to *bus IVT*) and the local train and regional + IC-train are now worse than *bus IVT*. This is an expected consequence of adding the *mode dummies* but since the effect is behaviourally difficult to explain this model setup is not used in the following.

Therefore the best fit for MNL models is obtained in the model formulation (6-37) with *public transport mode in-vehicle times*, *access/egress time* and *transfer attributes* and the model will be the foundation of the following models.

6.4.3 Path Size Logit models

In order to account for the similarity across routes in the multimodal network, the *Path Size (PS) Factor* is added to the model:

$$\begin{aligned} V_{kn} = {} & \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\ & + \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\ & + \beta_{IVT,RegIC} * IVT_{RegIC,kn} + \beta_{WalkTime} * TT_{Walk,kn} \\ & + \beta_{WaitTime} * TT_{Wait,kn} + \beta_{ChangePen} * N_{Change,kn} + \beta_{PS} * \ln(PS_{kn}) \end{aligned} \tag{6-39}$$

Where PS_{kn} is the Path Size Factor calculated according to formula (6-12) and β_{PS} is a coefficient to be estimated.

Also the Path Size Correction (PSC) factor from formula (6-14) has been calculated and added to the utility and estimated according to the formula:

$$\begin{aligned} V_{kn} = {} & \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\ & + \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\ & + \beta_{IVT,RegIC} * IVT_{RegIC,kn} + \beta_{WalkTime} * TT_{Walk,kn} + \beta_{WaitTime} * TT_{Wait,kn} \\ & + \beta_{ChangePen} * N_{Change,kn} + \beta_{PSC} * PSC_{kn} \end{aligned} \tag{6-40}$$

The parameter estimates for the Path Size Logit models are presented in Table 6-12 alongside the estimates for the selected MNL model.

Table 6-12: Estimated parameter coefficients and (robust t test) for the MNL, PS and PSC models including in-vehicles times, transfer attributes and PS factors.

Parameter	MNL model + No. Transfers (6-37)	Path Size models LN(PS) (6-39)	PSC (6-40)
In Vehicle Time			
Bus	-0.153 (-34.3)	-0.141 (-30.9)	-0.143 (-31.4)
Metro	-0.025 (-3.18)	-0.024 (-2.92)	-0.101 (-9.07)
Local train	-0.112 (-10.9)	-0.099 (-8.94)	-0.027 (-3.21)
Regional + IC-train	-0.134 (-15.9)	-0.110 (-12.7)	-0.113 (-13.0)
S-train	-0.085 (-16.9)	-0.074 (-14.0)	-0.076 (-14.5)
Access/Egress	-0.368 (-55.7)	-0.340 (-50.4)	-0.343 (-50.9)
Path Size Factor			
LNPS	-	-0.747 (-14.8)	-
PSC	-	-	-0.755 (-14.0)
Transfers			
Waiting Time	-0.084 (-13.8)	-0.083 (-13.7)	-0.083 (-13.7)
Walking Time	-0.136 (-10.4)	-0.131 (-10.1)	-0.132 (-10.1)
No. Transfer	-2.140 (-43.7)	-1.970 (-40.9)	-2.020 (-42.0)
Number of estimated parameters	9	10	10
Number of observations	5,641	5,641	5,641
Null log-likelihood	-20,172	-20,172	-20,172
Final log-likelihood	-11,947	-11,740	-11,760
Likelihood ratio test	16,450	16,862	16,824
Adjusted rho-square	0.407	0.417	0.417

The coefficients for the PS and PSC factors are both negative. The Path Size factor enters the utility function with a negative sign and hence a value of zero means that the alternative is unique and the higher absolute value the more the alternative is overlapping with other alternatives in the choice set. The negative estimated coefficient therefore means that the travellers prefer the alternatives with high overlap. This issue is investigated further in the discussion section of this Chapter.

The inclusion of the Path Size Factor does not affect the estimates relative to each other but it gives a small improvement in the model fit. Since the Path Size Factors take the overlapping of routes into consideration and since the Path Size Correction formulation is better theoretically founded the formula (6-40) is used as foundation for the following models.

Path Size Logit model for trip purposes

In order to account for heterogeneity between different segments of travellers different segmentations of the observations into trip purposes are tested. The differences in behaviour and perception of the variables between some traveller groups are not as great as might be expected and finally two traveller groups are defined based on trip purposes:

- Work related trips – 2,952
- Leisure and other trips – 2,689

The work related trips include commuting trips, trips to educational institutions and business trips. The number of trips for each category is shown in Table 6-13. The Work Trips segment has the highest number of public transport legs with an average 1.45 public transport legs per trip and leisure trips have 1.30 legs.

Table 6-13: Number of public transport legs for public transport mode for trip purpose.

	Traveller group					
	All Trips		Work Trips		Leisure Trips	
Mode	No.	[%]	No.	[%]	No.	[%]
Bus	3,403	43.7	1,718	40.2	1,685	48.1
S-train	2,488	32.0	1,463	34.2	1,025	29.3
Metro	1,086	14.0	589	13.8	497	14.2
Regional + IC-train	636	8.2	414	9.7	222	6.3
Local Train	169	2.2	94	2.2	75	2.1
Sum	7,782	100.0	4,278	100.0	3,504	100.0

The observations are grouped in the two trip purpose groups and a model is estimated with each of the data sets. Table 6-14 shows the estimated coefficients for the estimations compared to the estimation with all trips.

Table 6-14: Estimated coefficients and (robust t test) for each trip purpose for the PS logit model including in-vehicles times.

	Traveller group		
Parameter	All Trips	Work Trips	Leisure Trips
In vehicle Time			
Bus	-0.143 (-31.4)	-0.166 (-25.3)	-0.115 (-18.7)
Local Train	-0.101 (-9.07)	-0.136 (-10.6)	-0.062 (-3.54)
Metro	-0.027 (-3.21)	-0.046 (-4.00)	-0.003 (-0.27)
Regional + IC-train	-0.113 (-13.0)	-0.135 (-11.5)	-0.084 (-6.60)
S-train	-0.076 (-14.5)	-0.101 (-13.7)	-0.042 (-5.89)
Access/Egress	-0.343 (-50.9)	-0.365 (-36.6)	-0.320 (-35.0)
Path Size Factor			
PSC	-0.755 (-14.0)	-0.754 (-9.78)	-0.764 (-10.1)
Transfers			
Waiting Time	-0.083 (-13.7)	-0.082 (-14.1)	-0.085 (-7.05)
Walking Time	-0.132 (-10.1)	-0.137 (-7.39)	-0.132 (-7.35)
No. Transfer	-2.020 (-42.0)	-2.020 (-29.7)	-2.060 (-28.9)
Number of estimated parameters:	10	10	10
Number of observations:	5,641	2,952	2,689
Null log-likelihood:	-20,172	-10,724	-9,442
Final log-likelihood:	-11,760	-6,082	-5,645

Likelihood ratio test:	16,824	9,284	7,595
Adjusted rho-square:	0.417	0.432	0.401

All coefficients are significant except for the metro IVT which is not significantly different from zero for leisure trips. All parameters have the expected signs and most parameters have the expected size. The model fit is much higher for work travellers than for leisure travellers.

Table 6-15: Estimated coefficients scaled to bus IVT (=1.0) for each trip purpose for the PS logit model including in-vehicles times.

	Trip purpose		
Parameter	**All**	**Work**	**Leisure**
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.7	0.8	0.5
Metro	0.2	0.3	-
Regional + IC-train	0.8	0.8	0.7
S-train	0.5	0.6	0.4
Access/Egress	2.4	2.2	2.8
Transfers			
Waiting Time	0.6	0.5	0.7
Walking Time	0.9	0.8	1.1
No. Transfer	14.1	12.2	17.9

The tables show that leisure travellers find all aspects of transferring more burdensome compared to a bus IVT minute than travellers with work related purposes do. Work related travellers would rather spend additional 12 minutes on the bus than transferring whereas leisure travellers would rather spend 18 extra minutes on the bus. The leisure travellers find bus IVT, local train IVT, S-train IVT and Access/Egress travel time less burdensome than work travellers do.

For leisure trips the coefficients for the IVT in buses are higher than the IVT for all train modes. For work related travellers the regional train IVT is equal to the local train IVT and preferred over the bus IVT and only the metro and S-train IVTs are assessed to be more attractive.

Trip purposes and distance bands

To investigate the variation in preferences for travellers on trips of different lengths the model data are divided into distance bands. Several distance bands are investigated to determine the most descriptive division for the trips in the public transport network of the Greater Copenhagen Area. The distance bands are tested on the full data set and on the traveller groups work trips and leisure trips.

Three distance bands definitions are tested:

- Cut at 10 km.

- Cut at 10 and 25 km.
- Cut at 20 km.

The number of trips for the trip purposes in each distance band is shown in Table 6-16 (see Appendix 3 for the cut at 10 and 25 distance band).

Table 6-16: Number of public transport legs for public transport mode for trip purpose and distance band.

Trip purpose		Trip Distance Band				
	Mode	All	<10km	>10km	<20km	>20km
All Trips	Bus	3,403	2,143	1,260	2,909	494
	S-train	2,488	924	1,564	1,763	725
	Metro	1,086	694	392	924	162
	Regional + IC-Train	636	76	560	192	444
	Local Train	169	41	128	88	81
	Sum	7,782	3,878	3,904	5,876	1,906
Work Trips	Bus	1,718	932	786	1,401	317
	S-train	1,463	451	1,012	997	466
	Metro	589	336	253	485	104
	Regional + IC-train	414	40	374	111	303
	Local Train	94	17	77	47	47
	Sum	4,278	1,776	2,502	3,041	1,237
Leisure Trips	Bus	1,685	1,211	474	1,508	177
	S-train	1,025	473	552	766	259
	Metro	497	358	139	439	58
	Regional + IC-train	222	36	186	81	141
	Local Train	75	24	51	41	34
	Sum	3,504	2,102	1,402	2,835	669

Of the 7,782 public transport trip legs for all trips 55% is used for work related trips and 45% for leisure trips. The table shows that leisure trips generally have fewer trip legs (and also trips) in the long trip distance bands.

The table shows that the choice of transport modes along the routes depends very much on the length of the trip. At the cut-at-20 km distance band most travellers use bus, S-train and metro for the shorter trips, local train is equally divided between the distance bands and the regional and IC-trains are more often used for longer trips than for shorter trips.

The three mentioned cuts of distance bands are tested in the model estimation and the cut at 10 km are assessed to be the best representation of the data. The parameter estimates for the “All trips” model for the distance bands with cut at 10 and 20 km are presented in Table 6-17. The model estimates for the 10+25 km distance cut can be seen in Appendix 3.

Table 6-17: Estimated parameter coefficients and (t tests) for All Trips for the cut-at-10km and cut-at-20km distance bands.

	Trip Distance Band			
Parameter	**<10 km**	**>10km**	**<20 km**	**>20km**
In vehicle Time				
Bus	-0.179 (-24.5)	-0.116 (-21.8)	-0.153 (-28.7)	-0.111 (-14.5)
Local Train	-0.154 (-4.05)	-0.076 (-7.57)	-0.115 (-6.43)	-0.069 (-5.78)
Metro	-0.045 (-3.62)	-0.022 (-1.96)	-0.036 (-3.69)	-0.007 (-0.43)
Regional + IC-train	-0.196 (-6.73)	-0.089 (-10.4)	-0.144 (-8.76)	-0.087 (-8.52)
S-train	-0.089 (-8.29)	-0.063 (-11.01)	-0.071 (-10.93)	-0.077 (-9.46)
Access/Egress	-0.386 (-36.9)	-0.297 (-35.68)	-0.358 (-44.8)	-0.284 (-23.9)
Path Size Factor				
PSC	-0.798 (-12.1)	-0.659 (-7.13)	-0.803 (-13.7)	-0.523 (-3.93)
Transfers				
Waiting Time	-0.166 (-13.1)	-0.051 (-8.84)	-0.118 (-17.4)	-0.036 (-4.91)
Walking Time	-0.165 (-8.34)	-0.113 (-6.40)	-0.137 (-9.38)	-0.129 (-4.43)
No. Transfer	-2.010 (-29.9)	-1.890 (-29.0)	-2.000 (-38.7)	-1.850 (-17.7)
Number of estimated parameters:	10	10	10	10
Number of observations:	3,185	2,456	4,507	1,134
Null log-likelihood:	-11,108	-9,070	-15,908	-4,272
Final log-likelihood:	-6,351	-5,255	-9,170	-2,479
Likelihood ratio test:	9,513	7,631	13,477	3,586
Adjusted rho-square:	0.427	0.420	0.423	0.417

All parameter estimates are significant except metro in-vehicle time for trip distances exceeding 20 kilometres and all models improve the model fit from the model without distance bands.

Table 6-18 presents the estimated parameter coefficients for the Work and Leisure trips for the distance band with cut at 10 km. All coefficients have the expected signs and all coefficients are significant except the IVT metro for the leisure trips in both distance bands. The model fits are high for the work related trips. The model fit for the leisure trips for the above 10 km distance band is low but the adjusted rho squares are all higher for the distance bands models than for the models without distance bands.

Table 6-18: Estimated parameter coefficients and (t tests) for Work and Leisure Trips for the cut-at-10km distance band.

	Trip purpose			
	Work		Leisure	
Parameter	**<10 km**	**>10km**	**<10 km**	**>10km**
In vehicle Time				
Bus	-0.209 (-17.8)	-0.139 (-18.7)	-0.154 (-16.6)	-0.082 (-11.5)
Local Train	-0.176 (-2.87)	-0.111 (-9.03)	-0.134 (-2.73)	-0.041 (-2.84)
Metro	-0.061 (-3.44)	-0.046 (-3.09)	-0.032 (-1.87)	0.012 (0.74)
Regional + IC-train	-0.223 (-4.98)	-0.113 (-9.53)	-0.174 (-4.56)	-0.053 (-4.52)
S-train	-0.111 (-6.24)	-0.089 (-11.3)	-0.072 (-5.60)	-0.021 (-2.84)
Access/Egress	-0.427 (-23.8)	-0.320 (-28.0)	-0.356 (-28.1)	-0.269 (-22.3)
Path Size Factor				
PSC	-0.773 (-8.00)	-0.715 (-5.84)	-0.825 (-9.19)	-0.583 (-4.17)
Transfers				
Waiting Time	-0.128 (-8.40)	-0.063 (-11.0)	-0.219 (-10.8)	-0.035 (-3.58)
Walking Time	-0.166 (-5.79)	-0.123 (-5.02)	-0.179 (-6.56)	-0.100 (-4.13)
No. Transfer	-2.120 (-19.3)	-1.880 (-22.0)	-1.920 (-22.2)	-1.960 (-18.9)
Number of estimated parameters:	10	10	10	10
Number of observations:	1,392	1,560	1,793	896
Null log-likelihood:	-4,951	-5,777	-6,153	-3,293
Final log-likelihood:	-2,766	-3,265	-3,551	-1,955
Likelihood ratio test:	4,371	5,024	5,203	2,675
Adjusted rho-square:	0.439	0.433	0.421	0.403

Table 6-19 presents the estimated parameters scaled to bus IVT. For most travel time and transfer attribute parameters there is a difference between the estimates relative to bus IVT for the trips in the short and long distance bands. The results show that for the “All trips” model the regional and IC-train IVT is preferred over the bus at long trips and more onerous than the bus for short trips. For the shorter trips the number of transfers is considered worse (relative to bus IVT) than for the longer trips.

The Path Size factor coefficients are different for work travellers on short and long trips compared to the size of the other parameter estimates in the same model. Relative to other parameter estimates the estimates for the long trips are higher which means a higher preference for overlapping paths at long trips. This phenomenon is discussed in the discussion section of this chapter.

Table 6-19: Estimated parameter coefficients scaled to Bus IVT for All, Work and Leisure Trips for the cut-at-10km distance band.

	Trip purpose					
	All trips		Work Trips		Leisure Trips	
Parameter	**<10 km**	**>10km**	**<10 km**	**>10km**	**<10 km**	**>10km**
In vehicle Time						
Bus	1.0	1.0	1.0	1.0	1.0	1.0
Local Train	0.9	0.7	0.8	0.8	0.9	0.5
Metro	0.2	0.2	0.3	0.3	-	-
Regional + IC-train	1.1	0.8	1.1	0.8	1.1	0.6
S-train	0.5	0.5	0.5	0.6	0.5	0.3
Access/Egress	2.2	2.6	2.0	2.3	2.3	3.3
Transfers						
Waiting Time	0.9	0.4	0.6	0.5	1.4	0.4
Walking Time	0.9	1.0	0.8	0.9	1.2	1.2
No. Transfer	11.2	16.3	10.1	13.5	12.5	23.9

For Leisure trips the metro IVT is not significant in this distance band (this is the case for all leisure trip distance band cuts, see Appendix 3).

There is a general trend of the transfer preferences for all three trip purposes. Compared to bus IVT the travellers have a lower transfer penalty at short trips than at long trips. For work trips this is partly compensated by the waiting time at transfer being considered more burdensome (compared to bus IVT) at the shorter trips but only for high waiting times the waiting time difference is high enough to compensate for the transfer penalty difference. The transfer issues in the distance bands models are discussed in Section 6.5.

Path Size Logit model including headway

To add more explanatory power to the models the importance of the headway for public transport route choice is tested by adding half the headway in minutes to the models. Two definitions of the headway are tested:

- First headway.
- Highest headway.

The headway is a measure of the number of minutes between each departure (run) of a specific line serving the same stops. The *first headway* is the headway for the first public transport vehicle on the trip and the *highest headway* is the headway for the public transport vehicle on the route with the highest headway/lowest frequency. The utility model is estimated following:

$$\begin{aligned} V_{kn} = {} & \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\ & + \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\ & + \beta_{IVT,RegIC} * IVT_{RegIC,kn} + \beta_{WalkTime} * TT_{Walk,kn} + \beta_{WaitTime} * TT_{Wait,kn} \\ & + \beta_{ChangePen} * N_{Change,kn} + \beta_{PSC} * PSC_{kn} + \beta_{Headway} * T_{Headway,half,kn} \end{aligned} \quad \textbf{(6-41)}$$

Where $T_{Headway,half,kn}$ is the half of the *headway* measured in minutes and $\beta_{Headway}$ is a parameter to be estimated.

The models are estimated for All, Work and Leisure trips disregarding the distance bands. The parameter estimates for the Highest Headway are presented in Table 6-20 and the estimation results for the First Headway are presented in Appendix 3.

Table 6-20: Estimated parameter coefficients and (robust t test) for All, Work and Leisure Trips for PS Logit model with the *Highest Headway* parameter.

Parameter	Model considering *highest* headway		
	All	Work	Leisure + Other
Headway			
½ of Highest	-0.052 (-8.34)	-0.051 (-5.90)	-0.053 (-6.10)
In vehicle Time			
Bus	-0.137 (-30.5)	-0.160 (-24.5)	-0.110 (-18.0)
Local Train	-0.073 (-6.93)	-0.117 (-8.85)	-0.046 (-3.28)
Metro	-0.029 (-3.51)	-0.047 (-4.09)	-0.009 (-0.75)
Regional + IC-train	-0.112 (-12.7)	-0.131 (-10.5)	-0.091 (-7.05)
S-train	-0.077 (-14.8)	-0.102 (-13.8)	-0.045 (-6.34)
Access/Egress	-0.353 (-50.7)	-0.376 (-37.1)	-0.330 (-34.3)
Path Size Factor			
PSC	-0.682 (-12.3)	-0.690 (-8.68)	-0.676 (-8.80)
Transfers			
Waiting Time	-0.080 (-13.3)	-0.079 (-13.7)	-0.082 (-6.75)
Walking Time	-0.126 (-9.75)	-0.133 (-7.20)	-0.125 (-7.03)
No. Transfer	-1.990 (-40.3)	-1.980 (-28.7)	-2.020 (-27.4)
Number of estimated parameters:	11	11	11
Number of observations:	5,641	2,952	2,689
Null log-likelihood:	-20,172	-10,724	-9,442
Final log-likelihood:	-11,659	-6,032	-5,593
Likelihood ratio test:	17,025	9,385	7,699
Adjusted rho-square:	0.421	0.437	0.407

The Highest Headway parameters are significant for the whole data set and for the two trip purposes. The metro IVT parameter for leisure trips is not significantly different from zero. The model fits improve compared to the model without headway (see Table 6-14).

Table 6-21 shows the parameter estimates scaled to bus IVT.

Table 6-21: Estimated parameter coefficients scaled to Bus IVT (=1.0) for All, Work and Leisure Trips for PS Logit model with the *Highest Headway* parameter.

	Model considering highest headway		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
½ of Highest	0.4	0.3	0.5
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.5	0.7	0.4
Metro	0.2	0.3	-
Regional + IC-train	0.8	0.8	0.8
S-train	0.6	0.6	0.4
Access/Egress	2.6	2.4	3.0
Transfers			
Waiting Time	0.6	0.5	0.7
Walking Time	0.9	0.8	1.1
No. Transfer	14.5	12.4	18.4

Relative to the bus IVT the route choice for leisure trips is more affected by the highest headway parameter than the work related trips. The scaled parameter estimates are all very similar to the scaled estimates for the models without the highest headway (see Table 6-15).

The influence of the headway variable on the Path Size factor is discussed further in the discussion in Section 6.5.

PS Logit model with transfer characteristics

In this section the estimation model is added new variables describing the transfers more detailed. The variable *No. of transfers* is replaced by four variables describing which public transport service types the transfer is between. Four variables are defined:

- Transfer from Bus to Bus.
- Transfer from Bus to Train.
- Transfer from Train to Bus.
- Transfer from Train to Train.

The model is specified according to the formula:

$$
\begin{aligned}
V_{kn} = {} & \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\
& + \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\
& + \beta_{IVT,RegIC} * IVT_{RegIC,kn} + \beta_{WalkTime} * TT_{Walk,kn} + \beta_{WaitTime} * TT_{Wait,kn} \\
& + \beta_{ChangePen,Bus-Bus} * N_{Change,Bus-Bus,kn} \\
& + \beta_{ChangePen,Bus-Train} * N_{Change,Bus-Train,kn} \\
& + \beta_{ChangePen,Train-Bus} * N_{Change,Train-Bus,kn} \\
& + \beta_{ChangePen,Train-Train} * N_{Change,Train-Train,kn} + \beta_{PSC} * PSC_{kn} \\
& + \beta_{Headway} * T_{Headway,half,kn}
\end{aligned}
\tag{6-42}
$$

The model estimation will reveal whether the travellers find some transfers more burdensome than others. The estimation results are presented in Table 6-22 and Table 6-23 presents the parameters scaled to bus IVT (=1.0).

Table 6-22: Estimated parameter coefficients and (robust t test) for All, Work and Leisure Trips for PS Logit model with *Transfer Characteristic.*

	Trip purpose		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
½ of Highest	-0.050 (-7.94)	-0.049 (-5.60)	-0.051 (-5.90)
In vehicle Time			
Bus	-0.131 (-29.7)	-0.154 (-23.4)	-0.106 (-17.9)
Local Train	-0.080 (-8.14)	-0.122 (-9.04)	-0.051 (-3.95)
Metro	-0.061 (-7.20)	-0.084 (-7.02)	-0.035 (-2.94)
Regional + IC-train	-0.122 (-14.3)	-0.141 (-11.7)	-0.100 (-8.13)
S-train	-0.090 (-16.8)	-0.115 (-15.1)	-0.057 (-7.85)
Access/Egress	-0.364 (-51.0)	-0.391 (-37.1)	-0.337 (-34.5)
Path Size Factor			
PSC	-0.684 (-12.2)	-0.684 (-8.48)	-0.688 (-8.88)
Transfers			
Waiting Time	-0.080 (-13.1)	-0.079 (-13.4)	-0.081 (-6.67)
Walking Time	-0.154 (-9.14)	-0.151 (-6.54)	-0.166 (-6.75)
No. Transfer			
Bus->Bus	-2.580 (-32.4)	-2.670 (-22.0)	-2.490 (-23.7)
Bus->Train	-1.940 (-20.6)	-2.020 (-15.6)	-1.860 (-13.2)
Train->Bus	-2.140 (-22.5)	-2.260 (-17.3)	-2.010 (-14.1)
Train->Train	-1.230 (-20.6)	-1.180 (-14.8)	-1.350 (-14.4)
Number of estimated parameters:	14	14	14
Number of observations:	5,641	2,952	2,689
Null log-likelihood:	-20,172	-10,724	-9,442
Final log-likelihood:	-11,563	-5,957	-5,573
Likelihood ratio test:	17,217	9,536	7,739
Adjusted rho-square:	0.426	0.443	0.408

All parameter estimates have the expected sign and all are significantly different from zero (also IVT for metro for leisure travellers). All parameter estimates for the train IVTs increase relative to bus IVT. For travellers with both work and leisure trip purposes the local train IVT and regional train IVT are now greater than the bus IVT. The metro IVT increases to approximately ½ the bus IVT. For work travellers the transfer walking and waiting times are hardly affected but for leisure travellers the importance of transfer walking time increases 17% to 1.2 minutes of bus IVT.

Table 6-23: Estimated parameter coefficients scaled to Bus IVT (=1.0) for All, Work and Leisure Trips for PS Logit model with *Transfer Characteristic.*

Parameter	All	Work	Leisure + Other
Headway			
½ of Highest	0.4	0.3	0.5
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.6	0.8	0.5
Metro	0.5	0.5	0.3
Regional + IC-train	0.9	0.9	0.9
S-train	0.7	0.7	0.5
Access/Egress	2.8	2.5	3.2
Transfers			
Waiting Time	0.6	0.5	0.8
Walking Time	1.2	1.0	1.6
No. Transfer			
Bus->Bus	19.7	17.3	23.5
Bus->Train	14.8	13.1	17.5
Train->Bus	16.3	14.7	19.0
Train->Train	9.4	7.7	12.7

Addressing the issue of comparing the transfer types it can be seen that the travellers do in fact distinguish between the transfer types. For both trip purposes the travellers prefer the transfer types in the following order (most preferred first):

1. Train -> Train
2. Bus -> Train
3. Train -> Bus
4. Bus -> Bus

As expected the travellers prefer the transfers between two trains, bus->train->bus is in between and bus-> bus is the less attractive transfer combination. This pattern is discussed in more detail in the discussion section below.

The segmentation of the transfer attributes imposes all train IVTs to increase compared to bus IVT. This model specification affects the metro IVT parameters to be significant and increase for all purposes to more intuitively correct values (half of the bus IVT) and the model is therefore used in the following.

Correlation of transfer walking and waiting time and transfer penalties

A correlation matrix has been made to investigate the correlation between the transfer characteristics and the correlation matrix in Table 6-24 shows that there is indeed correlation between the coefficients. The maximum correlation of the Transfer Waiting Time coefficient is a correlation with the Transfer Bus->Bus coefficient of -0.132. The correlation with waiting time is rather low for all transfer coefficients so from this it is difficult to say that a specific transfer type involves more waiting time than the other transfer types.

Table 6-24: Correlation matrix between transfer waiting time, transfer walking time and transfer penalties.

Coefficient	Transfer Bus ->Bus	Transfer Bus ->Train	Transfer Train ->Bus	Transfer Train ->Train	Transfer Walk Time	Transfer Wait Time
Transfer Bus->Bus	-					
Transfer Bus->Train	0.257	-				
Transfer Train->Bus	0.250	0.310	-			
Transfer Train->Train	(-0.110)	0.218	(-0.166)	-		
Transfer Walk Time	(-0.070)	-0.451	-0.463	(-0.027)	-	
Transfer Wait Time	(-0.121)	(-0.132)	(-0.083)	(-0.081)	(-0.060)	-

The correlations for all coefficient pairs of wait/walk time and transfer are negative. This means that when the coefficient for time increases the coefficient for number of transfers decreases.

The Transfer Walking Time coefficient is correlated with two transfer types, transfers from bus to train and train to bus. This shows that especially the transfers between the train and bus systems include a high walking time.

For correlation matrixes there should be values above +/- 0.3 but no values should exceed 0.7 and this is fully fulfilled by the matrix above.

PS Logit model including split of Headway and Regional + IC-train in-vehicle time

In the final model the variables for headway and regional and IC-train in-vehicle time are split in two variables each to representing the differences in perceptions of the variables found in the earlier models.

The headway variable is split into:

- Headway (½ of the highest) up to 6 minutes.
- Headway (½ of the highest) exceeding 6 minutes.

The regional and IC-train variable are split in two variables:

- Regional and IC-train in-vehicle time up to 20 km.

– Regional and IC-train in-vehicle time exceeding 20 km.

The new variables are added to the model (replacing the original variables) and the model is specified as follows:

$$
\begin{aligned}
V_{kn} = {} & \beta_{IVT,Bus} * IVT_{Bus,kn} + \beta_{IVT,LocalTrain} * IVT_{LocalTrain,kn} \\
& + \beta_{IVT,Metro} * IVT_{Metro,kn} + \beta_{IVT,STrain} * IVT_{STrain,kn} \\
& + \beta_{IVT,RegIC\_upto20km} * IVT_{RegIC\_upto20km,kn} \\
& + \beta_{IVT,RegIC\_above20km} * IVT_{RegIC\_above20km,kn} + \beta_{WalkTime} * TT_{Walk,kn} \\
& + \beta_{WaitTime} * TT_{Wait,kn} + \beta_{ChangePen,Bus-Bus} * N_{Change,Bus-Bus,kn} \\
& + \beta_{ChangePen,Bus-Train} * N_{Change,Bus-Train,kn} \\
& + \beta_{ChangePen,Train-Bus} * N_{Change,Train-Bus,kn} \\
& + \beta_{ChangePen,Train-Train} * N_{Change,Train-Train,kn} + \beta_{PSC} * PSC_{kn} \\
& + \beta_{Headway\_upto6min} * T_{Headway\_upto6min,half,kn} \\
& + \beta_{Headway\_above6min} * T_{Headway\_above6min,half,kn}
\end{aligned}
\qquad \textbf{(6-43)}
$$

Table 6-25 presents the estimation results for the model.

Table 6-25: Estimated parameter coefficients and (robust t test) for All, Work and Leisure Trips for PS Logit model with *Transfer Characteristic, Split headway* and *Regional + IC-train in-vehicle time.*

	Trip purpose		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
Up to 6 min	-0.279 (-7.12)	-0.261 (-4.91)	-0.302 (-5.23)
Above 6 min	-0.044 (-7.16)	-0.044 (-5.11)	-0.045 (-5.24)
In vehicle Time			
Bus	-0.132 (-29.6)	-0.156 (-23.6)	-0.105 (-17.6)
Local Train	-0.080 (-8.14)	-0.124 (-9.10)	-0.051 (-4.09)
Metro	-0.075 (-8.29)	-0.100 (-7.86)	-0.048 (-3.72)
Regional + IC-train			
Up to 20km	-0.167 (-10.1)	-0.196 (-8.39)	-0.142 (-5.75)
Above 20km	-0.084 (-7.01)	-0.099 (-6.70)	-0.060 (-2.94)
S-train	-0.095 (-17.1)	-0.123 (-15.1)	-0.061 (-8.24)
Access/Egress	-0.366 (-51.0)	-0.394 (-36.9)	-0.339 (-34.7)
Path Size Factor			
PSC	-0.680 (-12.2)	-0.673 (-8.30)	-0.687 (-8.94)
Transfers			
Waiting Time	-0.080 (-13.2)	-0.079 (-13.5)	-0.082 (-6.72)
Walking Time	-0.146 (-8.74)	-0.142 (-6.20)	-0.159 (-6.49)
No. Transfer			
Bus->Bus	-2.560 (-32.3)	-2.660 (-21.9)	-2.470 (-23.5)
Bus->Train	-1.890 (-20.1)	-1.980 (-15.4)	-1.810 (-12.8)
Train->Bus	-2.090 (-22.0)	-2.220 (-17.0)	-1.950 (-13.6)
Train->Train	-1.180 (-19.7)	-1.120 (-14.0)	-1.290 (-14.0)
Number of estimated parameters:	16	16	16
Number of observations:	5,641	2,952	2,689
Null log-likelihood:	-20,172	-10,724	-9,442
Final log-likelihood:	-11,526	-5,935	-5,555
Likelihood ratio test:	17,291	9,578	7,776
Adjusted rho-square:	0.428	0.445	0.410

The table shows that all parameters have the expected sign and that the new variables improve the model fit. As expected the parameter estimate for the headway exceeding 6 minutes and for the regional and IC-train IVT exceeding 20 km is lower than for the lower variables.

Table 6-26 presents the parameter estimates scaled to bus IVT. The table shows that for the short headways each minute counts much more than each of the extra minutes for the long headways.

Table 6-26: Estimated parameter coefficients scaled to Bus IVT (=1.0) for All, Work and Leisure Trips for PS Logit model with *Transfer Characteristic, split headway* and *Regional + IC-train in-vehicle time.*

	Trip purpose		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
Up to 6 min	2.1	1.7	2.9
Above 6 min	0.3	0.3	0.4
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.6	0.8	0.5
Metro	0.6	0.6	0.5
Regional + IC-train			
Up to 20km	1.3	1.3	1.4
Above 20km	0.6	0.6	0.6
S-train	0.7	0.8	0.6
Access/Egress	2.8	2.5	3.2
Transfers			
Waiting Time	0.6	0.5	0.8
Walking Time	1.1	0.9	1.5
No. Transfer			
Bus->Bus	19.4	17.1	23.5
Bus->Train	14.3	12.7	17.2
Train->Bus	15.8	14.2	18.6
Train->Train	8.9	7.2	12.3

The new variables cause most parameter estimates to increase slightly compared to bus IVT. The order of the parameters is the same as in the previous model. The travellers perceive each minute of extra travel time for regional and IC-train very similar to how they perceive the metro IVT.

PS Logit model including transfer characteristics with trip distance band

The model including transfer characteristic is also estimated for the trip distance band with cut at 20 km (estimation results in Appendix 3). In these models, especially the regional train IVT relative to bus IVT is high for travellers on short trips compared to travellers on long trips. For short trips the travellers compare 1 regional train minute to 2.3 minutes in a bus whereas the long distance travellers compare 1 minute in a regional train to 0.9 minutes in a bus.

PS logit model including socio-economic variables

Two models with the model specification as above have been estimated in order to test whether gender or occupation of the traveller have an impact of the travellers preferences.

Appendix 3 presents the estimation results for the gender specification and shows that the preferences for men and women are very similar. The women avoid all aspects of transfers more than men (compared to bus IVT) but only to a small extent and also the women have a smaller dispreference for headway. The preferences for IVT are very similar for the two genders.

In Appendix 3 is also presented estimation results for traveller occupation which shows rather similar tendencies for the occupations; student, unemployed and employed. The group of unemployed has smaller dispreference for headway and higher dispreference for transfers and access/egress time.

Since these two estimation results do not show any greater impact of socio-economic variables on route choice parameters these are not used for the following model specifications.

6.4.4 Mixed Path Size Logit models

To investigate whether taste heterogeneity is found within the travellers in the data set, different setups of mixed logit models with random parameters are estimated. The models are estimated on the full data set since the differences between the two groups might be explained by other factors.

As explained the model specification from formula (6-41) is used and initially one model estimation is carried out for each of the model parameters to test whether the parameter is in fact random and follows a distribution all other things equal. Two distributions are tested: the log-normal and the normal. According to the literature (see section 6.1) the log-normal is used for the time parameters since it is expected that a distribution for the time parameters should be truncated at zero not imposing any travellers to have a positive attraction for more time spent travelling.

In the initial tests the following parameters show a tendency toward being mixed:

- Headway above 6 min.
- Bus IVT.
- Regional + IC-train IVT up to 20 km.
- S-train IVT.
- Access/Egress Time.
- Transfer Waiting Time.
- Transfer Bus -> Bus.
- Transfer Bus -> Train.
- Transfer Train -> Bus.
- Transfer Train -> Train.

This means that the data do not show heterogeneity in the perception of the parameters for Local Train, Metro in-vehicle Time, regional and IC-train IVT above 20 km, headway above 6 min and Transfer Walking Time since the standard deviation of the parameters were not significantly different from zero. The parameters are fixed (not distributed) in the Mixed Logit model.

The mixed logit models are estimated with 500 MLHS (Modified Latin Hypercube Sampling) draws as defined by Hess et al. (2006).

When estimating β and σ the mean and standard deviation for the parameters is calculated from the corresponding Normal distribution of the corresponding means and standard deviations from the Log-Normal distributions.

The mean for the Bus IVT is calculated as:

$$IVT_{Bus,mean} = e^{\beta + \frac{1}{2}\sigma^2} \tag{6-5}$$

And the standard deviation as:

$$IVT_{Bus,st.dev.} = e^{\beta + \frac{1}{2}\sigma^2} \sqrt{e^{\sigma^2} - 1} \tag{6-6}$$

Similar formulas can be written for the other parameters.

In order to scale the estimated parameters to the distributed bus IVT parameter each of the distributions are simulated by using 1,000,000 draws from the specified distribution. The mean and the 90 percentage confidence interval are simulated for each parameter.

The parameter estimates from the Mixed Path Size logit estimation is shown in Table 6-27.

Table 6-27: Estimated parameter coefficients and (robust t-test), mean and standard deviation for the underlying Normal distributions, and mean and 90%confidenceinterval for All Trips for Mixed PS Logit model.

ALL TRIPS Parameter	Estimated coefficient	t-test	Normal distr. Mean	St.dev	Simulation – Ratio to Bus IVT Mean	90% conf. int.
Headway						
Up to 6 min	-0.247	(-7.90)			**1.8**	[1.2-2.5]
Above 6 min, μ	-3.050	(-35.3)	-0.087		**0.6**	[0.1-1.4]
Above 6 min, σ	1.100	(10.1)		0.096		
In-vehicle time						
IVT Bus, μ	-1.940	(-73.6)	-0.151		**1.0**	[1.0-1.0]
IVT Bus, σ	0.304	(10.9)		0.096		
IVT Local Train	-0.123	(-6.75)			**0.9**	[0.6-1.3]
IVT Metro	-0.089	(-10.3)			**0.6**	[0.4-0.9]
IVT Regional + IC-train						
Up to 20 km, μ	-1.690	(-24.7)	-0.205		**1.5**	[0.6-2.6]
Up to 20 km, σ	0.460	(4.43)		0.138		
Above 20 km	-0.092	(-9.97)			**0.7**	[0.4-0.9]
IVT S-train	-0.110	(-23.4)			**0.8**	[0.5-1.1]
TT Access/Egress	-0.401	(-74.4)			**2.9**	[1.9-4.1]
Path Size Factor						
PSC	-0.703	(-17.9)				
Transfer						
Transfer Walking Time	-0.159	(-8.22)			**1.2**	[0.7-1.6]
Transfer Waiting Time	-0.084	(-37.0)			**0.6**	[0.4-0.9]
No. Transfers						
Bus -> Bus	-2.680	(-38.8)			**19.5**	[12.6-27.5]
Bus -> Train, μ	-0.635	(13.7)	-2.231		**16.3**	[5.7-30.4]
Bus -> Train, σ	0.579	(5.88)		1.600		
Train -> Bus, μ	-0.770	(17.7)	-2.700		**19.7**	[5.9-38.5]
Train -> Bus, σ	0.668	(6.68)		2.047		
Train -> Train	-1.230	(-19.5)			**9.0**	[5.8-12.6]
Number of Hess-Train draws:	500					
Number of estimated parameters:	21					
Number of observations:	5,641					
Null log-likelihood:	-20,172					
Final log-likelihood:	-11,418					
Likelihood ratio test:	17,508					
Adjusted rho-square:	0.433					

The table shows that for the all data model not all the assumed parameters are significant so not all the parameters tested for shows heterogeneity. The parameters which are in fact perceived differently by the travellers are:

- Headway above 6 min.
- Bus IVT.

- Regional and IC-train IVT up to 20 km.
- Transfer Bus -> Train.
- Transfer Train -> Bus.

The last parameters have standard deviations not significantly different from zero. The model fit improves considerably by allowing for heterogeneity. The estimates for the means are close to the estimates for the Path Size logit model.

Table 6-28: Estimated parameter coefficients and (robust t-test), mean and standard deviation for the underlying Normal distributions, and mean and 90%confidenceinterval for Work Trips for Mixed PS Logit model.

WORK TRIPS					Simulation - Ratio to Bus IVT	
	Estimated		Normal distr.			
Parameter	coefficient	t-test	Mean	St.dev	Mean	90% conf. int.
Headway						
Up to 6 min	-0.224	(-5.12)			**1.3**	[1.0-1.7]
Above 6 min, μ	-2.950	(-26.2)	-0.104		**0.6**	[0.1-1.4]
Above 6 min, σ	1.170	(8.36)		0.125		
In-vehicle time						
IVT Bus, μ	-1.770	(-55.8)	-0.174		**1.0**	[1.0-1.0]
IVT Bus, σ	0.193	(5.11)		0.107		
IVT Local Train	-0.160	(-5.24)			**1.0**	[0.7-1.2]
IVT Metro	-0.113	(-9.08)			**0.7**	[0.5-0.8]
IVT Regional + IC-train						
Up to 20 km, μ	-1.530	(-20.4)	-0.231		**1.4**	[0.8-2.1]
Up to 20 km, σ	0.356	(2.42)		0.149		
Above 20 km	-0.111	(-9.33)			**0.7**	[0.5-0.8]
IVT S-train	-0.139	(-22.0)			**0.8**	[0.6-1.0]
TT Access/Egress	-0.438	(-52.8)			**2.6**	[2.0-3.3]
Path Size Factor						
PSC	-0.695	(-12.2)				
Transfer						
Transfer Walking Time	-0.152	(-6.17)			**0.9**	[0.7-1.1]
Transfer Waiting Time	-0.083	(-19.7)			**0.5**	[0.4-0.6]
No. Transfers						
Bus -> Bus, μ	1.080	(-4.17)	-3.152		**18.8**	[10.1-29.5]
Bus -> Bus, σ	-0.369	(12.7)		2.047		
Bus -> Train, μ	-0.714	(4.65)	-2.376		**14.2**	[5.7-25.3]
Bus -> Train, σ	0.550	(15.5)		1.676		
Train -> Bus, μ	-0.858	(5.45)	-2.887		**17.3**	[5.9-32.4]
Train -> Bus, σ	0.636	(-14.0)		2.144		
Train -> Train	-1.200	(-19.7)			**7.2**	[5.5-9.0]
Number of Hess-Train draws:	500					
Number of estimated parameters:	22					
Number of observations:	2,952					
Null log-likelihood:	-10,724					
Final log-likelihood:	-5,863					
Likelihood ratio test:	9,722					
Adjusted rho-square:	0.451					

Table 6-29: Estimated parameter coefficients and (robust t-test), mean and standard deviation for the underlying Normal distributions, and mean and 90%confidenceinterval for Leisure Trips for Mixed PS Logit model.

LEISURE TRIPS					**Simulation – Ratio to Bus IVT**	
	Estimated		**Normal distr.**			
Parameter	**coefficient**	**t-test**	**Mean**	**St.dev**	**Mean**	**90% conf. int.**
Headway						
Up to 6 min	-0.283	(-6.16)			**2.8**	[1.3-4.6]
Above 6 min, μ	-3.080	(-24.0)	-0.070		**0.7**	[0.1-1.5]
Above 6 min, σ	0.917	(4.69)		0.065		
In-vehicle time						
IVT Bus, μ	-2.180	(-42.6)	-0.127		**1.0**	[1.0-1.0]
IVT Bus, σ	0.482	(9.40)		0.086		
IVT Local Train	-0.085	(-3.36)			**0.8**	[0.4-1.4]
IVT Metro	-0.063	(-5.23)			**0.6**	[0.3-1.0]
IVT Regional + IC-train						
Up to 20 km, μ	-1.850	(-13.3)	-0.196			
Up to 20 km, σ	0.663	(3.62)		0.148	**1.9**	[0.5-4.0]
Above 20 km	-0.067	(-4.07)			**0.7**	[0.3-1.1]
IVT S-train	-0.075	(-9.80)			**0.7**	[0.4-1.2]
TT Access/Egress	-0.372	(-50.9)			**3.7**	[1.8-6.1]
Path Size Factor						
PSC	-0.723	(-12.8)				
Transfer						
Transfer Walking Time	-0.179	(-5.44)			**1.8**	[0.9-2.9]
Transfer Waiting Time	-0.086	(-30.5)			**0.9**	[0.4-1.4]
No. Transfers						
Bus -> Bus	-2.630	(-24.7)			**26.2**	[12.5-43.2]
Bus -> Train, μ	-0.585	(7.21)	-2.192		**21.8**	[5.7-44.0]
Bus -> Train, σ	0.632	(3.77)		1.623		
Train -> Bus, μ	-0.702	(9.29)	-2.667		**26.5**	[5.7-55.8]
Train -> Bus, σ	0.747	(4.12)		2.138		
Train -> Train	-1.350	(-13.6)			**13.4**	[6.4-22.2]
Number of Hess-Train draws:	500					
Number of estimated parameters:	21					
Number of observations:	2,689					
Null log-likelihood:	-9,442					
Final log-likelihood:	-5,506					
Likelihood ratio test:	7,872					
Adjusted rho-square:	0.415					

6.5 Discussion and conclusion of the results

In the following section the results of the route choice parameter model estimation will be commented upon and discussed. It is important to take notice that the parameters estimated are estimated on a selection of travellers only including public transport travellers and the preferences found for these travellers will most likely show a more positive attitude to public transport than preferences for travellers in general.

6.5.1 Parameter estimates

Regional Train In-Vehicle Time

In the initial specifications of the route choice models the results show a perception of regional and IC-train which was the same or more negatively perceived than the bus IVT. The negative preference is highest for the leisure travellers (compared to the work travellers) and higher (in absolute values) for short trips than for long trips. Intuitively one would assume that the travellers would prefer regional train time over bus time since the trains are most often superior to the bus. The trains have the “rail factor” which is an attraction to the train which cannot be described in terms of time use etc. The trains are more comfortable to be seated in, the driving pattern (acceleration and deceleration) is often more comfortable in a train compared to a bus, etc. refer to Chapter 2 for more details on this.

The regional train is used more often by work travellers than leisure travellers and more often for long trips than for short trips. When testing for the importance of the trip length it is shown that the travellers at the long trips perceive the regional train IVT very close to how they perceive the bus IVT whereas the travellers on shorter trips have a very high negatively perception of the regional train IVT.

The importance of the frequency of the transport modes was tested to see if the low frequencies of the regional trains (see Chapter 2) caused this reluctance against the trains. The inclusion of the frequency parameter did improve the perception of the regional and IC-train IVT preference to be 0.8 minutes of bus IVT (see Table 6-21). In all models, however, regional and IC-trains are perceived to be the worst train IVT.

The reason for this is most likely found in the physical structures of the rail networks. The structure shows that the regional trains are most suitable for long trips whereas S-train and metro are most convenient for the shorter trips. This is for example emphasised by the train station structure. At the train stations served by several train service types often the trains serve different platforms. Often the S-trains and the metro platforms are placed most central and the access to the regional and IC-train platforms is harder. Also the physical dimensions of the train could play a role. The S-trains and metros are low vehicles with the floor at ground height whereas the regional and IC-trains are a combination of low vehicles and high vehicles with a stair to enter the train wagons. As discussed earlier the frequency could also play a role. The S-trains and metros have a higher frequency so if the traveller misses his planned departure, the next departure would most likely be in a few minutes. For the regional trains the frequencies are lower and a missed train will cause greater delays for the whole route. This is tested for with the frequency parameter and the results support the theory.

As seen in the number of trip legs table (Table 6-16) the regional and IC-trains are most often used on the longer trips and this is reflected in the traveller's perceptions of the trains. Many travellers are perhaps not aware of the possibility to use the regional trains as an alternative to for example S-train on the shorter trips.

Regional and IC-train IVT split at 20 km

When splitting the IVT for regional and IC-train at 20 kilometres so the IVT up to 20 km is defined as one variable and the IVT exceeding 20 km is defined as a second variable the parameter estimates for the two variables are very different for the two. The travellers perceive the IVT up to 20 km very negatively whereas the regional and IC-train travel time exceeding 20 minutes is not very negatively perceived.

Figure 6-7 shows the traveller's perception of the in-vehicle time in regional and IC-train with the split at 20 km. The graph shows a break at the 20 km trip leg point after which the in-vehicle time still is important for the traveller to minimise but the cost of an extra minutes is not as expensive for the traveller as an extra minute at the shorter trips legs.

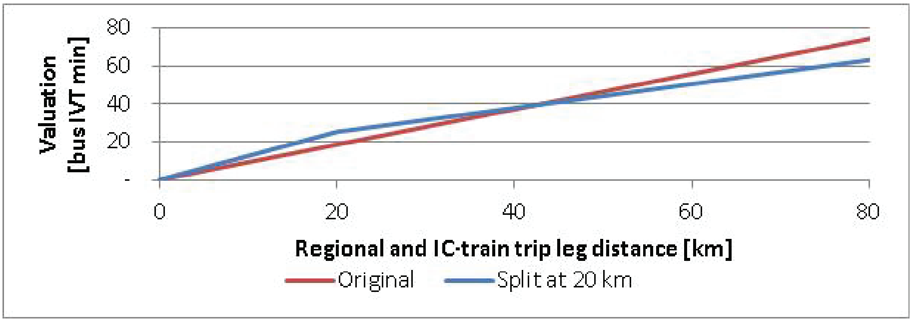

Figure 6-7: Preference for regional and IC-train in-vehicle time depending on trip leg distance scaled to bus IVT (=1).

Metro in-vehicle time

In a number of the model estimation results for leisure travellers the coefficient for the in-vehicle time in the metro is not significantly different from zero. The metro in Copenhagen has a high frequency (every 2 minutes during rush hours) and a high occupancy of travellers.

The results show that the metro IVT is significant for working travellers but not for leisure travellers. For work travellers the metro IVT is preferred (lowest negatively perceived) over all other public transport mode IVT's but it is perceived negatively and the work travellers try to minimize the number of minutes spend in the metro. They would rather arrive sooner than driving around in the metro.

The leisure travellers however seem to be indifferent of the travel time in the metro. They do not perceive negatively extra time spent in the metro compared to arriving at the destination. In general the leisure travellers perceive all the negative parameters less negatively than the work travellers and as the work travellers they also prefer the time in the metro over travel time in the other modes. The preference for metro might be explained by the characteristics of the metro. The metro is the newest rail line in the Greater Copenhagen Area and is served by trains without a driver. Children can sit in the front seats and play that they are driving. For a long period the metro attracted people who took a ride on the metro just for the fun. This might describe the indifference or only slightly negative perceived metro in-vehicle time.

In the final models when introducing the transfer split the metro IVT is also significant for the leisure travellers. This means that some of the preference for the metro lies in the transfer attributes and not in the travel time. The travellers prefer the metro over the other transport modes but they would try to minimise the travel time.

Metro IVT for Distance Bands

When segmenting the model data on distance bands the results show that the metro in-vehicle time for the longest trips is not significant for the “All trips” data set for the longest trips (exceeding 10 km and 20 km). Table 6-16 showed that the highest number of trips using metro as one of the trip legs is found within the shortest trips. When not segmenting for trip distance bands the metro IVT’s was significant.

<u>S-train in-vehicle time</u>

S-train is also highly preferred by the travellers. The S-train is used for a very high number of trips and since they serve the CBD of Copenhagen and the suburbs both commuters and leisure travellers have a high preference for these trains. Commuters are more attracted to S-train than leisure travellers and one of the reasons for this is that the S-trains are convenient to travellers because of the areas they serve (the suburbs where the commuters live and the city centre where people work). The S-train is preferred second to the metro but when adding the frequency parameter to the models it is revealed that especially higher frequencies (5-10 minutes) of the S-trains cause the high preferences for these train types. With the frequency parameter the S-train the travellers have lower preferences for S-train IVT than for local train IVT.

<u>Local train in-vehicle time</u>

The local trains are preferred over regional and IC-train, over bus and also over S-train when the frequency parameter is added to the model. The local trains often run with 30 minute frequencies and serve the smaller cities in the suburbs of Copenhagen. The trains typically have a high regularity and do not suffer from delays as much as other train types in the area. More than 99% of the local trains were punctual in 2012. The results show that the local train is very attractive to the travellers and that only the relatively low frequencies compromise the attractiveness of the local trains.

<u>Access/egress travel time</u>

One of the initial MNL models estimated in this chapter did not contain the access/egress travel time but only the public transport mode in-vehicle times. Adding the access/egress time to the model improved the model fit considerably. The access/egress time is estimated to be perceived 2-3 times worse than bus IVT. This means that a traveller would rather travel 3 minutes more by bus to avoid 1 minute of walking to the first stop or from the last stop. The estimate is constantly high throughout the various model specifications and is not very much affected by adding additional describing variables. The adding of the headway variable causes the estimate to increase compared to bus IVT. This implies that the zone connector time is even more important when taking headway into consideration. For short headways the timing of the arrival to the first stop is not as important as when the headway is high, because if the traveller misses a departure there is another one arriving shortly afterwards. This gives an inversely proportional relationship between the two variables and this effect is shown in the parameter estimates.

In Chapter 3 the importance of travel speed at the access/egress links was discussed and also several route choice models were specified and estimated using the various travel speed definitions (constant speed, constant increase in speed, the logistic curve approaching the tendency of the observed data). These models show that the lower the travel time defined at the access/egress links (the higher speed) the more negative the travellers perceive these travel times. The traveller tries to minimise the travel time but if an access/egress connector of 8 km has almost the same travel time as a connector of 1.5 km, this relationship would bias the model estimates of the travel times.

In the literature the connector travel speed is often defined to be constant without taking into consideration the fact that the not effective part of travelling by for example bicycle (parking etc.) minimises equivalent to the total travel time the longer the trip leg is. The investigations of the access/egress travel time in this thesis show that this is an important aspect to take into consideration when preparing the observed route choice data and specifying the models. The connector travel time is very important and leaving it out gives very low model fits so it is important to consider carefully how to define the travel speed and thereby the travel time for the travellers if not directly observed (by GPS data etc.).

In this thesis and with the data at hand, it is recommended to use a combination of a fitted logistic formula and a constant increase in the travel speed to describe the travel speed of access and egress legs up to 8 kilometres.

For further improvements it could be interesting to segment the travel speeds at the access and egress side of the trip or even better at the home-end and activity-end of the trip. The tables of mode shares from the TU Survey in Chapter 3 showed that the shares of access and egress modes are very different from each other and therefore also the travel speed at the connectors at the two ends are potentially different from each other.

Transfer waiting and walking time

In the initial analysis it is shown that the transfer attributes measured in time are significant in the route choice model. It is shown that the transfer waiting and walking time do not change (relative to bus IVT) when the headway is added to the model (the headway is important for the transport mode IVT's – see discussion below). When the transfer penalty is split for bus, train and direction the transfer walking time increases relative to the bus and the waiting time related to bus stays constant. The opposite was expected since the waiting time actually takes place at the stop/station the travellers departs the next vehicle at and the waiting time therefore could be depending on where the waiting time is spent.

The reason for the higher walking time is described by the preference for the mode types. A traveller is willing to walk longer to a train than to a bus and this is represented by the higher walking time when adding transfer split to the model.

Transfer penalty

When including the transfer penalty in the utility formulation the estimation results show as expected that all travellers perceive the number of transfers negatively and hereby prefer not to transfer or to keep the number of transfers to a minimum all other things equal. The importance of the transfer waiting and walking time is decreased when including the transfer penalty which shows that the parameters are to some degree correlated. They all refer to the act of transferring and the penalty is also consisting of parts of the waiting and walking times. The transfer waiting and walking time do however not describe fully the inconvenience of transferring since the transfer itself is also inconvenient for the traveller. When transferring the traveller has to leave the vehicle seated in, he has to bring his luggage etc., but also when transferring the total travel time on the route travelling is more vulnerable to delays. If the first vehicle is delayed the traveller risks not to reach the second vehicle in time and this uncertainty issue cannot be measured in waiting and walking time but is more likely a penalty as described by the transfer penalty.

Going even a bit further the travellers could be assumed to actually prefer to have a few minutes of waiting time when transferring which makes the route more robust to delays (also reported in the focus group interviews in Appendix 1). This could be looked further into in future research.

Transfer penalty split

When the transfer characteristics (in terms of transport mode transferred from and to) are included in the utility function it is seen that there is indeed a different perception of the different transfer types. The travellers prefer transfers within or to the train network over transfers within or to the bus network.

The train to train transfers are often at the level and the facilities at the transfer location are often good: covered for weather, perhaps a shop, good information about the departures etc. Second most preferred is the bus to the train. The transfer is out of level (from road to rail) but since the traveller will wait for the train at the train station the traveller will also here benefit from the above mentioned characteristics. Thirdly the transfer from train to bus is preferred and bus to bus is the least preferred. For the two last transfers the actual waiting time takes place at the bus stop which might have poorer characteristics than the train station. If the transfer is from a train to a bus the bus stop is most likely close to the station and the traveller is perhaps able to benefit from some of the services at the station: perhaps a shelter, a shop etc. The transfer from train to bus is often a transfer from a higher frequency to a lower frequency public transport service and the traveller can take this into consideration when choosing his route. If the transfer is between two buses and they are both low frequent delays can affect this transfer to a high degree.

Transfer penalty for distance bands

For the segmentation of the model data in distance bands is seen a tendency that the travellers at long trips have a higher transfer penalty (equivalent to bus IVT) compared with the travellers at shorter trips. Depending on the location of origin and destination shorter trips would often involve none or few transfers and longer trips would involve more transfers and therefore it could be counterintuitive that the travellers on long trips perceive the transfers more negatively. However this can be explained by travellers preferring to stay in the seat if working, reading, sleeping etc. and would rather travel longer than save a few minutes (16 min for "All trips" >10 km, 11 min for "All trips" <10 km) (which might be a small percentage of the total travel time).

In general the public transport route choice sets for longer and shorter trips are different in the average number of transfers in the route alternatives. For the longer routes more alternative routes with a higher number of transfers can be identified. If the travellers on the longer trips prefer a longer travel time over a high number of transfers the results will show a more negative preference for the transfers compared to the travellers at shorter trips.

The Path Size Factor

The Path Size models estimated in this chapter show that the adding of the Path Size factor to the MNL model do indeed improve the fit of the models. Also the behavioural interpretation of the models improves since the high overlap between alternative routes in metropolitan multimodal transport networks have to be taken into consideration.

The model estimation results in the all data models including the PSC factor show that the parameter estimate for the Path Size factor is negative. The PSC factor enters the utility function with a negative sign as given by the formula (6-14). A PSC factor value of zero means that the alternative is unique and the higher the absolute value the higher overlap the route alternative shares with other route alternatives in the choice set.

Hereby a high negative parameter coefficient for the PSC factors suggests that the travellers in the multimodal transportation network prefer to travel via routes with a high overlap. In private transportation road networks the PSC factors are perceived negatively since the travellers do not consider a route to be unique if the route is highly overlapping with other routes. The results from the public transport route choice model estimation in this thesis however show that the travellers in public transport multimodal network prefer non-unique (and thereby overlapping) routes.

When travelling along a route with a high overlap with other routes in the travellers' choice set the traveller will obtain a benefit and this benefit has shown to be very important. When the route has a high overlap with other routes the traveller will have access to a high number of alternatives along the route. If the traveller has been observed to choose a specific train departure from a specific train station, this station might have several other schedules departing close to the chosen schedule and serving the same origin and destination train stations. If this is the case the chosen route will have a high overlap with the other routes in the travellers choice set since the schedules close to the one chosen are very likely also present in the travellers choice set. If the traveller has transferred from a train to a bus along his route and the transfer location is served by a high number of buses also serving the traveller ‘s destination stop the route will have a high overlap, and thereby a high (absolute) PSC factor, with other routes in the choice set. In both examples the high overlap with a high number of alternative schedules possible to use by the traveller also leads to a high robustness in the route of the traveller. When offered several lines serving the desired stops or the areas close to the route will be considerably more robust to delays. If no alternatives are present, delays at one public transport line can affect the total travel time considerably if the traveller does not reach his connection in time and there is a very long waiting time for the next line in the correct direction to depart. When the traveller decides on a route and especially if the route involves transfers, the traveller might feel more confident to reach his destination in time if he has the possibility of making small changes in his original choice in case of delays. In this way he will still be able to conduct the trip within a minimum extra amount of travel time compared to a route with only very few alternatives and served by low frequent public transport services.

In the multimodal transportation route choice context the Path Size Factor can therefore be seen as a measure of the robustness of the trip the traveller is conducting. A high overlap is preferred since the robustness is higher and the total travel time is not as sensitive to delays as it will be if the route has a lower Path Size factor and thereby fewer alternatives for the traveller.

The models in this thesis do not consider the possibility of the traveller changing his route along the way. The models presume that the traveller will always choose his route before the beginning of the trip and stay on this route. This will often occur when the travellers have limited knowledge of the network and do not have access to information along the route. For many travellers the simultaneous route choice (choosing between all routes in the choice set before starting the trip) is not necessarily the correct description of the real world. However since the Path Size Factor to some degree explain the preference for robust routes the importance of this is to some degree softened.

As previously mentioned the negative sign for the PS factor also occurred in Hoogendoorn-Lanser (2005) who explained it by the assumption that travellers prefer train routes with high overlap because high overlap means that the service is highly frequent.

It could be very interesting to investigate if the impact of the Path Size Factor changes depending on the data the model is estimated at. It might be possible that a high Path Size Factor is perceived more attractive in networks often imposed by delays than in very reliable public transport networks.

Path Size factor for distance

When segmenting the public transport route choice data on distance bands a slightly different coefficient estimate of the Path Size factor scaled to bus IVT is seen. The tendency is that the long trips have a high preference for overlapping trips and the shorter trips have a lower preference. The estimate is negative for both so all data segmentations prefer the higher overlap over the lower. When segmenting on trip purpose it is revealed that the largest difference in the perception of overlap is found within the work travellers (commuters, education and business purposes). The preference for overlap is twice as high for long trips compared to short trips when segmenting the trips at a 20km threshold.

Several explanations can be found for this. Longer trips will often have a high overlap with other routes since only a few alternatives exist for the longest trip legs and the routes differentiate from each other on the feeder modes at each end. On shorter trips more different routes are available for the traveller and if the total route distance is short a different feeder mode in one end would mean a smaller total overlap between the route alternatives.

The results emphasise the importance of the choice set generation technique. If the technique does not allow for much overlap in the alternatives and a sorting procedure to only use partly unique routes are used the effect of the overlapping routes will not be captured and this could have a high impact on the model estimates and the model fits.

Highest headway parameter

The significance of adding the headway for the public transport mode with the highest headway/lowest frequency along the route to the model was tested. The results showed that for all trip purposes the inclusion of the headway resulted in a small increase in the annoyance of the regional, IC- and S-train IVT's (or a small decrease in the annoyance of bus IVT). This result implies that the transport mode on a specific route with the highest headway is often a bus. The travellers often use buses in the end of the trips which have a low frequency and this affects the model when adding the headway parameter. The scaled parameters show that all train IVT parameter coefficients increase 3, 7 and 14% for regional + IC-, S-train and metro train for the "All trips" model. The high increase in the parameter estimate for the metro when adding the headway to the model supports the theory that one of the reasons why travellers have a high preference for the metro is the high frequency (2 minutes during peak hours).

When including the headway of the vehicle with the lowest frequency in the utility function is seen that the travellers perceive the headway negatively. The headway is specified as half the time interval between two departures of the public transport mode serving the same two public transport stops/stations. As expected, the results show that the higher the headway the more inconvenient it is for the traveller.

Adding the headway to the model specification affects especially the preference for the Path Size overlap. The inclusion of headway causes the importance of the PSC coefficient to decrease which shows that there is a connection between the Path Size factor and the headway. At the same time the negative preference for the IVT's for all transport modes except bus and local train increase compared to the model estimates without the headway included and the model fits increase.

As described the PSC factor could be thought of as a measure of the robustness of the routes. The significance of the robustness of the routes is not equally as important when also the headway for the lowest frequency public transport mode along the route is included. The lowest frequency is the most vulnerable part of a multimodal trip and thereby the robustness of the route can to some degree be measured using the headway. Adding the headway to the model therefore means that part of the uncertainty in the models placed within the PSC factor is now measured using a description in minutes and the importance of the PSC factor decreases to some extent.

Headway split at 6 minutes

The test of splitting the headway variable into two variables for the headway up to 6 minutes and the headway exceeding 6 minutes shows that the travellers perceive each extra minute for the low headways higher than for the long headways.

The headway is shown in Figure 6-8 in its original form (no split) and with the split at 6 minutes.

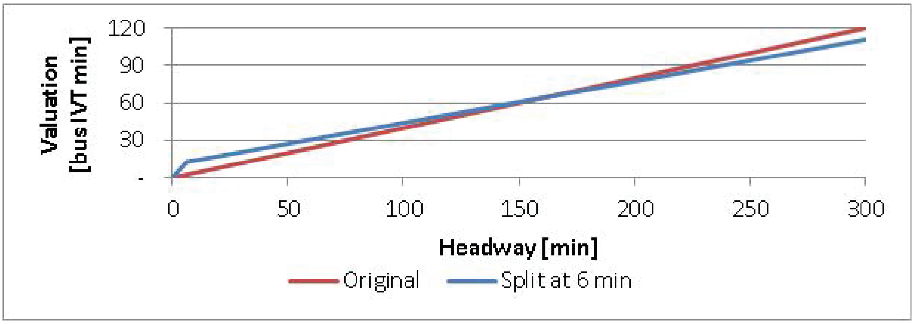

Figure 6-8: Preference for headway depending on trip leg distance scaled to bus IVT.

The segmentation of the headway variable shows a high improvement in the model fit but this is not the exact results expected. When the headway is low each extra minute is thought not to be as important because it might not be important whether you have to wait two or four minutes for the next bus/train to arrive. The longer headways are a higher annoyance to the travellers but each extra minute does not count as much as the extra minutes at low frequencies.

Trip purpose

To test for the importance of trip purposes on the route choice preferences the data sample is split in two purposes: *work related trips* and *leisure and other trips*. The models for the work travellers have the best model fits, also better than the "All trips" model, while the leisure trip model has a lower fit. The results show that travellers in the two trip purpose categories do consider the route choice parameters differently compared to bus IVT. The metro is not significant for the leisure travellers and the leisure travellers have higher preferences for in-vehicle time in all the train types. Compared to the work travellers, the leisure travellers have a higher reluctance against transferring and the attributes connected to transfers. The transfer penalty reluctance is almost 50% higher for leisure travellers (compared to bus IVT) and the transfer waiting and walking is perceived 30-50% worse. This means that the leisure travellers prefer to use fewer transport modes and accept a higher travel time in the modes compared to the work travellers who accept a smaller time benefit to make a transfer.

The definition of trip purposes and the changes in the model estimations show that there is indeed heterogeneity between the two groups of travellers. Some preferences are very similar to the two traveller groups but others (as explained) are different. The trip purpose segmentation adds complexity to the model and the modeller has to consider carefully whether to use an "All trips" model or the trip purpose split.

<u>Distance bands</u>

As described above the distance travelled influences the perception of the Path Size factors. Travellers on long trips have a higher preference to the Path Size factor (negatively) and therefore prefer routes with high overlap more than travellers on short trips.

For the "All trips" model especially the IVT for regional and IC-train and local trains are different in the two distance bands (relative to bus IVT). The travellers are more willing to use the two train modes on longer trips. For all data and for each of the two trip purposes the travellers prefer regional and IC-train over bus for longer trips and prefer bus over regional and IC-train for shorter trips

Also the transfer attributes of transfer waiting and transfer penalty show differences in the preferences at short and long trips. The travellers on long trips have a more negative preference against the number of transfers than the shorter trips. For leisure purposes the transfer waiting time is however perceived much worse at short trips compared to the long trips which means that the long trip travellers would be more willing to wait at the first transfer to save a second transfer than the travellers on short trips. The leisure purpose trips over 10 km is only 1/3 of the leisure trips sample but the long work trips are more than half the work trip sample. The result for the long leisure trips is not as robust as for the other data sets. This might explain some of the very large difference between the estimate for the long and short leisure trips.

In the final model the regional and IC-train IVT parameter is split in two at 20 km train trip legs and added to the model without distance band split. In this way it is possible to describe the very large differences found in these train types in a model not splitting for the parameters which do not change. The difference between the parameters is very high and these results show that the splitting of parameters is a more convenient method for describing the preferences varying with the trip distance.

6.5.2 Parameters compared to the literature

In the following the estimated parameters are compared to other scaled parameter found in literature and presented in Section 6.1.7.

Transport mode in-vehicle times

All models presented in the literature show that travellers prefer in-vehicle time in the train modes over bus in-vehicle time equivalent to the findings in this study. Nielsen and Frederiksen (2006) estimated a parameter for long IC-train travel times (>60 minutes) and found this to be exceeding the bus IVT which also in the case in this thesis.

Access/egress time

The literature shows very different valuations of access and egress time to and from public transport networks. The parameters are not always comparable since some (as this thesis) presumes access/egress time to describe only the private modes including walking used to travel to and from the public transport network and others assume that also for example buses used to the train station are access modes.

The presented literature estimate access egress time equivalent to bus minutes from 1.1 (Nielsen and Johansen, 2012), 1.6 (access time in private modes, Bovy and Hoogendoorn-Lanser, 2005), and 2.3 (access time, van der Waard, 1988). Only the value in van der Waard (1988) is as high as the 2-3 minutes found in this thesis, but the size of the value is comparable to the other literature on the subject.

Transfer waiting and walking time

The transfer waiting and walking time vary much in the different literature on public route choice model estimation. Eluru et al. (2012) estimated the waiting time to be 0.28 bus minutes. Nielsen and Frederiksen (2006) and Nielsen and Johansen (2012) assumed values of 1.1 for both values (2.3 for leisure in Nielsen and Frederiksen, 2006). Bovy and Hoogendoorn (2005) presented the highest values of approximately 2.0 for both values. In this thesis the values were estimated to be 0.6-1.1 for all purposes (also higher for leisure travellers) which are in line with the literature.

Transfer penalty

As mentioned the estimated values for transfer penalty are high in this thesis. Scaled to bus IVT the value is 14.5 for all transfer types and they vary from 9-19 for the transfers split on transfer type (train to train cheapest, bus-bus most expensive). The literature shows values of:

- 4.0 (Nielsen and Johansen, 2012).
- 3.8 (scaled to metro time, Raveau et al., 2011).
- 5.9 (van der Waard, 1988).
- 5.1 to 11.4 (high and low frequency transfer, Bovy and Hoogendoorn-Lanser, 2005).
- 15.1 (commuters, Nielsen and Frederiksen, 2006).
- 18.9 (all purposes, Vrtic and Axhausen, 2002).

These estimates cover a large range of the transfer penalty estimates in this thesis. The all transfer type variable estimated in this thesis is on the high end compared to other estimates but do not exceed all estimates found in literature.

Headway

Only a few models estimated headway and the parameter estimates vary from 0.3 (Vrtic and Axhausen, 2002), (half of the headway, Abrantes and Wardman, 2011), to 0.5 (Nielsen and Johansen, 2012 and Nielsen and Frederiksen, 2006), to 0.4. Before splitting the headway the parameter coefficients equal 0.3-0.5 which is very much in line with the literature.

Path Size factor

Not much literature shows estimations of Path Size logit coefficients for public transport models. In this thesis the coefficient is found to be negative and since the Path Size factor enters the utility function with a negative sign the sign of the coefficients represent the travellers' preference for high overlap. This conclusion equals the conclusions of Hoogendoorn-Lanser (2005).

From this review is seen that the parameter estimates of the model are very intuitive and comparable to other findings in the literature. Not all parameters can be compared since they are not tested in other models but those represented in other models seem reasonable.

6.5.3 Parameters not included in the models

The model estimations in the chapter tested a high number of different model parameters whereof many showed to be significant. The choice of parameters is very dependent on the data for the observed route and parameters for data not collected are of course very difficult to include in the models without making many assumptions. Also the network used for generation of route choice sets imposes limits to the parameters involved in the estimation of multimodal route choice models. In the literature is found a number of parameters which could be interesting also to include in the model estimations if the data allowed for this.

Costs

The costs of travelling on the public transport network are not part of the specified models. In the Greater Copenhagen Area the price for travelling is determined by the beginning and end of the public transport trip in a zone network structure and the fare for a trip is therefore not changing over alternatives. The only difference could be if a traveller decided to walk for a long distance to avoid paying for one or several zones. See the presentation of the fare structure in the Greater Copenhagen Area in Chapter 2.

The travellers do state the cost of their trip in the TU data and the fact whether they have a public transport season ticket or not. Still there are too many tricky issues concerning the costs in the zone structure why it was perceived as too cumbersome to go into details with defining a method to account for prices in the public network and therefore the costs are not part of the model. As mentioned van der Waard (1988) did not investigate the cost for the same reasons.

Since this is the case it is only possible to calculate trade-offs between different mode IVT, transfer penalty, etc. and not possible to calculate value-of-times. The value-of-times are often derived in other projects and could be used for comparison with the conditions in this model configuration.

Penalty if no seat

Nielsen and Frederiksen (2006) showed that the penalty if no seat was also a significant factor in the route choice. This penalty is a measure of the congestion in the public transport network. If the traveller is not able to get seated he has to stand up or might even be rejected and has to wait for the next departure serving the line. As discussed above some travellers are attracted to public transport modes where they can use the travel time to work, read etc. which is only possible if seated on the trip. Therefore the inclusion of such a parameter could improve the models.

Data on whether the traveller is seated is not collected in the TU Survey and was not part of the additional route choice questions developed in this thesis. Such a question could however be included in the survey and is worth to keep attention on. The network used for choice set generation do not use capacity restraints on the vehicles but with information of the travellers (correct OD matrices as input) such a factor would be possible to draw from the calculations. This would however be very time consuming and the benefits of using this extra parameter would have to be considered carefully.

Network topology attributes

Raveau et al. (2011) introduced network topological attributes in their route choice model estimations. They tested amongst others the importance of physical attributes of the train stations transferred at such as: whether or not the platforms used are served by escalators, whether or not a given transfer involves an ascending level change and collected the data for this via a field survey.

In the multimodal network of the Greater Copenhagen Area many transfers do involve transfers at stations both with and without escalators and often the traveller would have to change level to transfer between public transport services. These physical attributes have not been collected from the network of the Greater Copenhagen Area and since the multimodal network consists of a very high number of transfer possibilities the collection of the data would be very comprehensive, time-consuming and expensive. When collecting this data only at metro stations as in Raveau et al. (2011) only a limited number of different physical setups will be found but when including the whole multimodal network the many different modes involved and the many different constructions of stations, stops and transfer possibilities would be very huge. Such a data collection would set high demands to the people collecting the data, since the collection would have to be done consistently for the whole network. Having this data however would most likely improve the route choice models and this would definitely be an issue for future research.

Other factors

For the multimodal trips, information on the parking facilities could also be an important issue. With information on parking also the choice of station when travelling by car could be described more thoroughly. The information could be describing how the parking facilities are situated at the train station and whether or not the station has bicycle parking and car parking. This could also improve the models but the data has to be thoroughly collected for all stations and with comparable measures to include the facilities at all train stations at the same level. Some bus stops also have parking facilities but the investigation of parking facilities near to bus stops would most likely be too extensive to conduct with a similar method for the whole area.

The importance of the reliability of the transport modes could also be interesting to investigate. A measure for each mode type could be added to the models, but this would most likely be a too aggregate measure to improve the models. More information on the reliability for specific lines and at specific points in time would probably provide more explanatory power to the models.

6.5.4 Model characteristics

In this PhD thesis the Path Size Logit model is used and a variety of variables are tested in the different specifications of the model. The Path Size logit is chosen over the MNL model which is not suitable for route choice model specifications because of the Independence from Irrelevant Alternative (IIA) property. In actual transportation networks many routes will be overlapping and the model has to take this into account.

The Path Size models take the overlapping of alternatives into consideration. In car route choice models the Path Size factor is negative describing the situation that highly overlapping alternatives are not perceived as real alternatives in car networks. The results in this thesis, however, show that in the public transport network the Path Size factor is considered to be a positive characteristic and the travellers prefer routes with a high overlap over other routes in their route choice set with a low overlap.

The results from the PSL models are used to form the foundation for the extended Mixed Path Size Logit model. In the mixed model the travellers are allowed to have different perceptions of the variables and the heterogeneity of the traveller's preferences is described by lognormal distributions. The Mixed PS Logit models show that the traveller's preferences are indeed heterogeneous and that they follow the assumed distribution. The model fits are improved compared to the simpler PSL models.

Also others model types to model route choice exist. The models in this thesis assume that the traveller chooses between all whole routes before starting the trip and does not change the choice while travelling. This implies that the traveller has low knowledge of the present situation in the transport network and cannot access information en route or that the traveller chooses the same route repeatedly out of habit. Often travellers do have some knowledge of the network situation and have access to information during the trip. Dynamic modelling could solve these issues enabling the traveller to change route along the way if a change in the network situation occurs which make a new route the most optimal to the traveller. Dynamic modelling would set additional demands to the input data because also the present traffic situation in the network (delays, cancellations etc.) would be important for the description of the actual chosen route. Such data was not accessible for the study in this PhD project and is subject for future research.

Another alternative to Path Size Logit and Mixed Path Size Logit models is rule based modelling. When setting up a set of fuzzy logic rules the choices of routes can be described. The fuzzy logic approach is very time demanding for public transport route choice and the fuzzy genetic approach would be a theme for future research.

6.6 Summary and conclusions

This chapter has dealt with the issues of the multimodal public transport model estimations. The important and relevant literature on route choice modelling research and especially route choice for public transport and multimodal transport networks was presented.

The chapter goes through the findings of comparing the generated choice sets (from chapter 5) to the observed route choice data (from chapter 3). Important attributes of the two data sets are highlighted and compared to each other. In the presentation is shown that the observed routes are often within the shortest routes both in time and distance and that the travellers aim to reduce the number of transfers.

After this the parameters tested and investigated in the route choice model estimations are presented. These are:

- Dummies
  - Mode specific dummies.
  - Service type specific dummies.
- Time measures
  - In-vehicle time for mode types.
  - In-vehicle time for public transport services.
  - Access/egress time.
  - Walking time (transfer).
  - Waiting time (transfer).
  - Hidden waiting time.
- Trip purposes
  - Mode and trip purpose specific dummies.
  - In-vehicle time for mode and trip purpose.
- Trip distance
- Transfers
  - Number of transfers.
  - Transfers between public transport mode types.
  - Transfer location.
- Headway
  - Total headway.
  - Split at 6 min.
- Overlap
  - Path Size.

In the model estimation section, a large number of model specifications are set up. The first model specifications are multinomial logit models not taking the overlap of routes into consideration. When extending to Path Size models accounting for the overlap between routes an improvement in the model fits is shown and the model are more theoretically founded.

The results section above presents the model specifications and the estimation results and scaled values for a selected range of the models. The chapter shows that with the more elaborated model specifications the model fit improves and the parameter estimates are scaled to the bus IVT become more intuitively correct.

The model estimations show that the best PSL model setup includes the parameters and trip purposes presented in the table below where the parameters are scaled to the bus IVT (=1.0).

Final estimation results from PSL model specification – same as Table 6-26.

	Trip purpose		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
Up to 6 min	2.1	1.7	2.9
Above 6 min	0.3	0.3	0.4
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.6	0.8	0.5
Metro	0.6	0.6	0.5
Regional + IC-train			
Up to 20km	1.3	1.3	1.4
Above 20km	0.6	0.6	0.6
S-train	0.7	0.8	0.6
Access/Egress	2.8	2.5	3.2
Transfers			
Waiting Time	0.6	0.5	0.8
Walking Time	1.1	0.9	1.5
No. Transfer			
Bus->Bus	19.4	17.1	23.5
Bus->Train	14.3	12.7	17.2
Train->Bus	15.8	14.2	18.6
Train->Train	8.9	7.2	12.3

The in-vehicle times for all public and private transport modes are considered negatively and the travellers seek to minimise the total in-vehicle time. Also the time related to transfers (waiting and walking) is perceived negatively. The route choice preferences of the travellers show high reluctance against transferring and the estimates show the preferred order of the transfer types to be (best one first):

- Train to train.
- Train to bus.
- Bus to train.
- Bus to bus.

The travellers perceive regional and IC-train IVT for short trips to have a higher cost than the other public transport mode in-vehicles times. The bus is the second worse and the remaining train modes are considered slightly differently by travellers in the different traveller groups.

Access and egress time describes the time used for walking and biking to the first public transport stop and from the last public transport stop. The parameter estimate scaled to bus IVT shows that the travellers prefers to travel in a public transport mode over the access/egress time and that they would rather travel a few minutes longer in a vehicular transport mode compared to having a longer access/egress trip. The value was expected to be smaller (compared to literature) and higher value can be caused be the definition of the travel speed on the connectors or by the methods of generating route choice alternatives.

The mixed path size logit model shows that the travellers do indeed perceive the following parameters differently and that the perception of the parameters are said to follow a log-normal distribution.

- Headway above 6 min.
- Bus IVT.
- Regional and IC-train IVT up to 20 km.
- Transfer waiting time.
- Transfer Bus -> Bus.
- Transfer Bus -> Train.
- Transfer Train -> Bus.

The model fit improves considerably by allowing for heterogeneity. The estimates for the means are close to the estimates for the Path Size logit model but the relatively high standard deviations show that the parameters are much better reproduced when allowed to follow a distribution.

The last section of this chapter discusses the findings from the model estimations and touches upon the difficulties of using Path Sizes logit and mixed Path Size logit models for estimating multimodal route choice models.

The parameter estimates from this PhD study are compared to parameters previously found in public transport route choice modelling literature. The parameters which are also represented in other literature are close to other parameter estimates and this shows that the parameter estimates are reasonable.

7 CONCLUSIONS AND RECOMMENDATIONS

The study presented in this PhD thesis has touched upon many different research subjects within the multimodal route choice modelling process such as:

- Establishing the framework of multimodal transport networks used throughout the thesis and presenting the network of the real life transport network used in this thesis.
- Investigating survey methods and designing a survey method to collect public route choice data in an ongoing data collection from a large and representative sample.
- Analysing the data observations from the public transport vs. private transport mode choice at the mode chain level perspective.
- Generating route choice sets in a large scale network and establishing a measure for quality checking the generated choice sets.
- Specifying route choice models for Path Size Logit and Mixed Path Size Logit models and validating the estimation results.

This chapter starts with a summary of the thesis chapters. Section 7.2 follows with a longer list of conclusions and section 7.3 highlights the main conclusions from the thesis. Finally, recommendations for future work are given.

7.1 Summary

Throughout the thesis the research methods and the results have been discussed and concluded upon and this chapter will give a short summary of the thesis and recommendations for further research.

7.1.1 Framework

The framework for the study was set up. The attributes of trips, legs, transport modes, transfers etc. were defined and the process of conducting a trip was explained. The Greater Copenhagen Area used for the study was presented and elements important for the study were explained.

7.1.2 Data collection approach

Literature on data collection methods for (public transport) route choice was investigated and showed only very little effort in estimating route choices for public transport travellers. Recommendations for public transport route choice data collection methods were formulated and the thesis developed a questionnaire to collect public route choice data which were detailed enough to reconstruct the observed route in a GIS network. The questions collected detailed information on access modes, stations, lines, departure and arrival times, transfers, and egress modes.

The results of a full scale test of the data collection method conducted in the PhD project were presented followed by the results from the implementation in the national travel survey. More than 6,000 observations were collected and processed in this study. Methods of map-matching the route choice observations to a GIS network were presented and visualisation results illustrated. Finally, the access and egress parts of the trips were analysed.

7.1.3 Public transport route choice data

From the collected data the characteristics of the trips and the travellers in private and public transport modes were analysed and important aspects were highlighted. The detailed information about the travellers and their trips gave a thorough insight into the choices of the travellers in the private and public transport networks of the Greater Copenhagen Area.

The public transport mode chains were analysed and presented in this chapter.

7.1.4 Generation and quality assessment of route choice sets

The thesis proposed a doubly stochastic approach for generating alternative routes that are relevant to the travellers. The method took consideration the perceived costs of the network elements and the heterogeneity in the preferences of the travellers. Comparing the generated choice sets with the observed route choices the coverage gave a measure of the behavioural plausibility of the route choice generation technique.

The generated route choice sets were presented for a few trips and the choice sets were validated.

7.1.5 Estimation of public transport route choice models

The route choice sets were analysed and the variables used in the various setups of the model specifications were presented. The data allowed for a large variety of route choice variables and the most interesting were identified and hypothesis for the significance of the variables were introduced.

In this thesis Path Size Logit models were estimated, the model specification process was explained and the route choice variable assumptions were tested. A variety of model specifications were presented and developed through the chapter. The important preferences of individual travellers within trip purposes were tested for and so is the importance of the trip distance.

The most important findings from the model specifications were discussed and the assumptions of the important variables were concluded upon.

7.2 Conclusions on public transport route choice

In this section the most important conclusions of this PhD thesis are presented and the most relevant findings from the data collection method development, the route choice generation process and the estimation of route choice parameters for public transport travellers are emphasised.

7.2.1 Data collection method

Collecting observations on route choice for public transport passengers is not straightforward. GPS devices cannot be used (alone) because of the lack of detail in the collected data. To describe correctly public transport route choice information on access/egress mode, transport mode, line use, etc. is needed and information on trip purposes and social characteristics of the travellers also contributes to the route choice analyses. This information is not collected by GPS devices and the assumptions used for identification of mode choice from GPS data do not provide sufficiently precise data for route choice estimation purposes. Some smart card data collection systems contain detailed information of the public transport modes and lines but have a lack of information on the access/egress parts of the trips which is fundamental for the multimodal route choice description.

This thesis developed a questionnaire to collect route choice data for public transport travellers in a detailed but effective and not too time consuming way. The questionnaire fulfilled the requirements to set up the survey such as the detail level of the route descriptions, transport modes used, simplicity of the choices, possibility of reproducing the routes, a large number of respondents, and continuous data collection.

The test survey of the questionnaire at the Technical University of Denmark showed that it was in fact possible to collect the route observations and fulfil the requirements explained above. The integration of the route choice questions in the Danish Travel Survey showed that the additional time use because of the additional questions was not too high and the collected observations were of high quality.

The map-matching procedures refined for use in this study showed that it is possible to map-match the collected data to a GIS network which enlarges the possible uses of the route choice observations considerably. For this study the map-matching created the possibility of directly comparing the observed routes to the generated route choice sets and use a combination of the data sources for route choice model estimations.

7.2.2 Transport mode choices

The observed trip observations were analysed and factors important to the choice between private and public transportation and the choice between unimodal and multimodal trips were revealed. The private/public transport (mode choice) analysis showed that the distance between origin and destination, trip purpose, gender, age group, car ownership, household income, and distance to nearest train station were all important for the general transport mode choice. The examined factors were all characteristics of the traveller or the trip, and some of these can be difficult to modify in order to change the mode choice of the traveller. However, all these factors are important to be aware of when planning transportation services and when informing about these services.

The second analysis concluded that for the share of multimodal trips the trip distance, trip purpose, ownership of public transport tickets and of the bicycle, and distance to nearest station were important. The significant factors are all implicit to the choice of public transport. For the politicians to use these findings it is important to look at the possibilities of encouraging people to buy season tickets, or to purchase a bicycle. Also improving conditions for bicycle users, for example by improving bicycle parking at bus stops and train stations, is important for the decision makers to be aware of. Since commuters most often conduct multimodal trips the decision makers could focus on improving the conditions for the multimodal trips in peak hours where most commuter trips are performed to attract even more commuters to multimodal trips.

Analysis of transfer preferences from the TU survey data showed that many travellers try to avoid transfers. In a specific example up to six minutes of total travel time could be saved by transferring between two modes in the train and metro system. Most travellers actually chose not to take this transfer and instead stayed in the first train they boarded.

7.2.3 Route choice sets

Route choice sets were generated for the travellers observed in the collected data. The route choices were generated by the use of a model which is probit-based, in order to account for similarities across alternatives, and doubly stochastic, in order to account for heterogeneity in both the perceived value of the time components and the perception of the link impedance. Several formulations of the utility functions were tested and the variances of the error components and the error terms were varied to test for the best solution. The tests showed that the single stochastic generation functions were outperformed, and both the variation of the VOT-terms and the variation of the error term contributed to reach a good coverage. Higher variance produced more unique routes, but if the variance was too high the same alternatives were continuously generated, with consequent low efficiency in the production of alternatives.

7.2.4 Estimation of route choice parameters

Several model specifications were estimated testing for different hypotheses of parameter preferences. The model specifications were also tested for differences in preferences for travellers with different trip purposes and differences for different trip distances. The results were presented as estimated values and as parameter estimates scaled to bus IVT. The fare payment was not part of the model because of the fixed fare system and therefore it was not possible to estimate value-of-time parameters for the models. The scaled parameters, however, are a valuable measure for the preferences and for their relationship in-between and were used as a foundation for the discussions of the preferences.

The table below presents the results from the final Mixed Path Size Logit “All trips” model with the simulated mean scaled to the bus IVT and the simulated 90 percent confidence interval.

Mixed PSL values from Table 6-27.

Parameter	Simulated mean, scaled to Bus IVT	90% confidence interval
Headway		
Up to 6 min	1.8	[1.2-2.5]
Above 6 min	0.6	[0.1-1.4]
In-vehicle time		
IVT Bus	1.0	[1.0-1.0]
IVT Local Train	0.9	[0.6-1.3]
IVT Metro	0.6	[0.4-0.9]
IVT Regional + IC-train		
Up to 20 km	1.5	[0.6-2.6]
Above 20 km	0.7	[0.4-0.9]
IVT S-train	0.8	[0.5-1.1]
TT Access/Egress	2.9	[1.9-4.1]
Transfer		
Transfer Walking Time	1.2	[0.7-1.6]
Transfer Waiting Time	0.6	[0.4-0.9]
No. Transfers		
Bus -> Bus	19.5	[12.6-27.5]
Bus -> Train	16.3	[5.7-30.4]
Train -> Bus	19.7	[5.9-38.5]
Train -> Train	9.0	[5.8-12.6]

In-vehicle times

The final model showed that travellers did prefer regional and IC-train IVT over bus IVT and preferred the metro no matter the other parameters added to the model. The S-train was preferred second to the metro but when adding the frequency parameter to the models it was revealed that especially higher frequencies (5-10 minutes) of the S-trains caused the high preferences for these train types. The results showed that the local train is very attractive to the travellers and that only the relatively low frequencies compromised the attractiveness of the local trains. Finally one scaled access/egress minute was equal to 2-3 minutes in a bus and therefore the travellers sought to reduce the access/egress time compared to the transport mode in-vehicle time.

Transfer related attributes

The transfer attributes measured in time were significant for all route choice models estimated. A traveller was willing to walk longer to a train than to a bus and this was represented by the higher walking time when adding transfer splits to the model. The transfer penalty was perceived negatively in all models because the travellers preferred to avoid transfers all other things equal.

Including the transfer characteristics (in terms of transport mode transferred from and to) in the utility function showed that there was indeed a different perception for the transfer types. The travellers preferred transfers within or to the train network over transfers within or to the bus network which can be described by the characteristics of the stops and the attractiveness of the transport modes transferred to.

The Path Size Factor

The adding of the Path Size factor to the MNL model improved the fit of the models. Also the behavioural interpretation of the models improved when taking into consideration the high overlap between alternative routes in metropolitan multimodal transport networks.

In the multimodal transportation route choice context the Path Size Factor can be considered as a measure of the robustness of the trip the traveller is conducting. The results showed that a high overlap was preferred since the robustness was higher and the total travel time was not as sensitive to delays as it would be if the route has a lower Path Size factor and thereby fewer alternatives for the traveller.

The results emphasised the importance of the choice set generation technique. If the technique does not allow for much overlap in the alternatives and a sorting procedure to only use partly unique routes are used the effect of the overlapping routes will not be captured and this could have a high impact on the model estimates and the model fits.

Highest Headway Parameter

The headway parameter was significantly different from zero and negative in all model specifications. Inclusion of the headway resulted in a small increase in the annoyance of the regional, IC- and S-train IVT's (or a small decrease in the annoyance of bus IVT). Inclusion of the headway parameter in the model specification affected especially the preference for the Path Size overlap. The inclusion of the headway caused the importance of the PSC coefficient to decrease since part of the uncertainty in the models placed within the PSC factor was instead measured using a description in minutes and the importance of the PSC factor decreased to some extent.

The test of splitting the headway variable into two variables for the headway up to 6 minutes and the headway exceeding 6 minutes showed that the travellers value each extra minute for the low headways higher than for the long headways.

Trip purpose

To test for the importance of trip purposes on the route choice preferences the data sample was split in two purposes: *work related trips* and *leisure and other trips*. The models for work travellers had the best model fits, also better than the "All trips" model, while the "Leisure trips" model had a lower fit. The results showed that travellers in the two trip purpose categories do consider the route choice parameters differently compared to bus IVT. The leisure travellers have higher preferences for the in-vehicle time for all train types. Compared to the work travellers, the leisure travellers have a higher reluctance against transferring and the attributes connected to transfers. The transfer penalty reluctance was almost 50% higher for leisure travellers (compared to bus IVT) and the transfer waiting and walking times were perceived 30-50% worse.

The definition of trip purposes and the changes in the model estimations showed that there was indeed heterogeneity between the two groups of travellers. Some preferences were very similar for the two traveller groups while others were different. The trip purpose segmentation adds complexity to the model and the modeller has to consider carefully whether to use an “All trips” model or the trip purpose split.

Distance bands

The distance travelled influenced the perception of the Path Size factor. Travellers on long trips have a higher preference to the Path Size factor (negatively) and therefore prefer routes with high overlap more than travellers on short trips.

For the “All trips” model especially the IVT for regional and IC-train and local trains were different in the two distance bands (relative to bus IVT). The travellers were more willing to use the two train mode types on longer trips. For all data and for each of the two trip purposes the travellers preferred regional and IC-train over bus for longer trips and preferred bus over regional and IC-train for shorter trips. Also the attributes of transfer waiting and transfer penalty showed differences in the preference at short and long trips. The travellers on long trips had a more negative preference against the number of transfers than on the shorter trips. For leisure purposes the transfer waiting time was however perceived much worse at short trips compared to the long trips which means that the long trip travellers were more willing to wait longer at the first transfer to save a second transfer than the travellers on short trips. The results for the long leisure trips were not as robust as for the other data sets and this might explain some of the very large difference between the estimates for the long and short leisure trips.

In the final model the regional and IC-train IVT parameter was split in two at 20 km train trip legs and added to the model without distance band split. In this way it was possible to describe the very large differences found for these train types in a model not splitting for the parameters which did not change.

7.2.5 Path Size Logit and Mixed Path Size Logit Models

The Path Size Logit models were used for the initial analyses of the public transport route choice in this thesis. The models show to estimate the travellers’ preferences very well with a high model fit in the models. The Path Size improved the behavioural interpretation of the models since the overlapping between alternatives have to be taken into consideration.

The extension of the model framework to a Mixed Path Size Logit model increased the explanatory factor of the models. Also within the defined trip purposes the travellers had very different perceptions of the route attributes and this was explained by the high model fits of the models describing the relevant factors as following a lognormal distribution.

7.3 Main conclusions and contributions

The PhD study was able to develop a method to collect route choice for passengers in a multimodal public transport network. The PhD study tested the route choice questions as part of a travel diary survey in a full scale test and concluded that the additional questions did not extend the duration of the answering process or the drop-out rate significantly. By asking for information about; start and end location, departure and arrival times, stops and stations travelled via, and bus and train lines used the collected data were however detailed enough to reconstruct the observed route in a GIS network using methods developed in and in connection to this study. The route choice questions are now a fully integrated part of the Danish Travel Survey.

The method of generating route choice sets by use of a schedule-based stochastic transit assignment model showed that the choice sets with the highest coverage was generated using the formulation with variance for both the costs and for the error terms in the doubly stochastic function. The scale parameters of 1 gave the best coverage with the coverage defined as the number of trips with a certain percentage overlap between observed and generated route. The overlaps were calculated at both link and stop level. The technique of generating path choice sets in a large-scale network proved to be very useful and to give comprehensive input to the following route choice estimations.

The PhD study showed that the route choice formulations of the Path Size Logit and Mixed Path Size Logit models gave model estimates with intuitively correct signs and sizes. A large number of parameters were defined from the route observations and choice sets and the parameters important for the route choice of the passengers in the multimodal public transport network proved to be; IVTs for each specific mode, attributes related to transfers (both time and location), and the headway. The Path Size component ensures that the overlapping of route alternatives is considered. The estimation of the Path Size factor models showed that in public transport networks routes which have a high overlap with other routes were more attractive to the travellers since the high overlap embedded a higher robustness to the route. The importance of the trip purpose was emphasised by comparing the trip purpose categories and the importance of the trip distance was revealed when defining length and time thresholds for the parameter attributes. Along with the model estimations in this thesis was provided a thorough description of the important route choice parameters and a detailed discussion of the significance of the findings. The conclusions can be used for future development of public passenger route choice models and the route choice parameter values found for the Greater Copenhagen Area can be used to improve the model descriptions of the actual route choices in the network.

7.4 Recommendations and future work

Data collection

The collection of route choice for public transport passengers proved to be an effective method of collecting the most important attributes to reconstruct the observed route choices. The method of this thesis can be improved by adding more questions and soften some of the simplifications in the method. The bus stop used for boarding and alighting bus modes is a piece of data which was not collected in the TU survey. This was to minimize the complexity of the survey and the respondent burden but the data on bus stops would be very useful. The respondents were only allowed to answer in five minute intervals when they leave home but data on one-minute intervals might make the map-matching more precise.

Combining the travel dairy questionnaire with data from GPS devices would improve the data source considerable. The issues addressed above could be solved and the issues of travellers stating the wrong travel time and trip length would be solved as well. This would however be an expensive solution and would add to the respondent burden. The TU survey data are used for many other purposes and this should be taken into consideration when considering making changes in the survey.

Another possibility would be to combine the public transport smart card data with the TU survey data. This would give more detailed information on the lines and stop used but not about the access/egress modes which are important to the multimodal routes.

Time attributes

As expected the travellers preferred train over bus and the most preferred transport mode is the metro. The thesis showed that the frequency is the greatest reason for the differences between the various train modes. The metro is therefore preferred mainly because of the high frequency and the travellers are more reluctant to the regional and IC-trains because of the low frequency. This is a very important point for decision makers to bear in mind when suggesting changes to the public transport network.

Transfer attributes

Travellers preferred transfers to a train mode over transfers to a bus mode. This was explained by the higher comfort of train stations and possibility of entertainment (shops etc.) while waiting. Decision makers have to be aware of the attractiveness of the trains and maybe improve the conditions for the bus waiting areas. Also the waiting time was an issue and the results showed the importance of designing the public transport system to minimize the inconveniences of transferring such as the transfer waiting time. The waiting time has to be low in order to minimise the total travel time but also the travellers wish to have a buffer of a few minutes to be sure that they are able to reach the connection in time.

Other attributes

Other attributes about the public transport network could be interesting to add to the model formulations to test for the significance on the route choice decision. Parking availability, parking facilities, network topology attributes and other descriptions of the train stations might be important to the choice of the train station. Seating possibility and regularity might be important to the choice of transport mode. These issues have to be considered if renewing the transport survey or collecting route choice data in another survey. Some data can be collected independently on the survey (train station attributes) but then it is important to have a collection procedure which is exactly the same so the collected data can be compared.

Scaled parameters

Only parameter values and values scaled to bus in-vehicle time were estimated since no cost elements were part of the models. If the estimation results were to be used for a route choice assignment actual value-of-time measured would have to be calculated. This can be done by scaling all parameters to a known value-of-time for bus in-vehicle-time. For this method it is essential that the bus IVT value-of-time is well founded and validated since it will have an enormous effect on the value-of-times for all route choice attributes.

Path size factor

It could be very interesting to investigate if the impact of the Path Size Factor changes depending on the data the model is estimated on. It might be possible that a high Path Size Factor is perceived more attractive in networks often imposed by delays than in very reliable public transport networks. To test this data can be collected during a period where the transport modes operate with high punctuality and a period where the transport mode operations are exposed to unforeseen delays. Most likely the travellers' knowledge of planned delays would affect the results so the setup of the test has to be considered thoroughly.

The route choice generation technique is very important to take into consideration. The Path Size analysis showed that the overlapping of alternatives is an attribute highly preferred by the public transport passengers. This emphasises the importance of the choice set generation process since when partly overlapping alternatives are sorted out they cannot provide this effect. In car networks the route alternatives with small variations are not thought of as fully comparable routes but in public transport networks the routes with higher overlap are the most preferable to use since the robustness is higher if the traveller is offered options to change choices along the route. The route choice sets have to be created with carefulness to be able to describe this important factor for public transport route choice systems.

Further work with the study in this thesis

The Mixed Path Size Logit models should be investigated more thoroughly. The model estimation is very time consuming and the completion of a variety of model specifications have not been possible to conduct in this thesis. In future work this would be a very important issue to address since the mixing of the parameters is definitely present and adds to the importance of the model estimations.

It could be very interesting to convert the estimated parameters to value input to the route choice model in the Greater Copenhagen Area to investigate whether the output generates the routes investigated. The most optimal method would be to use another set of route choice observations to test for the reproduction of the correct routes and validate the parameters estimates to avoid endogeneity and not biasing the results by estimating the model and testing it with the same data. This is definitely a very interesting process and will most likely be conducted when new data sources are possible to obtain.

8 REFERENCES

Abdel-Aty, M.A., Kitamura, R. & Jovanis, P. (1995a). Exploring Route Choice Behavior using Geographic Information System-based Alternative Routes and Hypothetical Travel Time Information Input. *Transportation Research Record: Journal of the Transportation Research Board*, 1493, pp. 74–80.

Abdel-Aty. M.A., Kitamura, R., Jovanis, P.P., Reddy, P. & Vaughn, K.M. (1995b). New Approach to Route Choice Data Collection: Multiphase, Computer-aided Telephone Interview Panel Surveys Using Geographic Information Systems Data base. *Transportation Research Record: Journal of the Transportation Research Board*, 1493, pp. 159-169.

Abdelghany, K.F.S. (2001). *Stochastic Dynamic Traffic Assignment for Intermodal Transportation Networks with Consistent Information Supply Strategies*. PhD Thesis. University of Texas, Austin.

Abdelghany, K.F.S. & Mahmassani, H. S. (1999). Multi-Objective Shortest Path Algorithm for Large Scale Intermodal Networks. *INFORMS, Philadelphia, Fall*.

Abdelghany, K.F.S. & Mahmassani, H.S. (2001). Dynamic Trip Assignment-Simulation Model for Intermodal Transportation Networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1771, pp. 52-60.

Abrantes, P.A.L. & Wardman, M.R. (2011). Meta-analysis of UK Values of Travel Time: An Update. *Transportation Research Part A: Policy and Practice*, 45 (1), pp. 1-17.

Anderson, M.K.A. (2010a). Characteristics of Trips and Travellers in Private and Public Transportation in the Danish Travel Survey data. In *Selected proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Anderson, M.K.A. (2010b). Development and Assessment of a Data Collection Method for Route Choice in Public Transport. In *Selected proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Anderson, M.K.A. & Rasmussen, T.K. (2010). Matching Observed Public Route Choice Data to a GIS Network. In *Selected proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark.

Ankomah, P.K., Crompton, J.L. & Baker, D.A. (1995). A Study of Pleasure Travelers' Cognitive Distance Assessments, *Journal of Travel Research*, vol. 34, no. 2, pp. 12-18.

Azevedo, J.A., Santos Costa, M.E.O., Silvestre Madeira, J.J.E.R., & Vieira Martins, E.Q. (1993). An Algorithm for the Ranking of Shortest Paths. *European Journal of Operational Research*, 69(1), pp. 97–106.

Axhausen, K. W., Haupt, T., Fell, B. & Heidl, U. (2001). Searching for the Rail Bonus: Results from a Panel SP/RP Study. *European Journal of Transport and Infrastructure Research*, 1(4), pp. 353-369.

Barnes, G. (2005). The Importance of Trip Destination in Determining Transit Share. *Journal of Public Transportation*, volume 8, no.2, pp. 1-15.

Barry, J.J., Newhouser, R., Rahbee, A. & Sayeda, S. (2002). Origin and Destination Estimation in New York City with Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board,* 1817, pp. 183-187.

Barry, J.J., Freimer, R. & Slavin, H. (2009). Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. *Transportation Research Record: Journal of the Transportation Research Board,* 2112, pp. 53-61.

Bekhor, S., Ben-Akiva, M.E. & Ramming, S. (2002). Adaptation of Logit Kernel to Route Choice Situation. *Transportation Research Record: Journal of the Transportation Research Board,* 1805, pp. 78-85.

Bekhor, S., Ben-Akiva, M.E. & Ramming, M.S. (2006). Evaluation of Choice Set Generation Algorithms for Route Choice Models. *Annals of Operations Research*, 144 (1), pp.235–247.

Bekhor, S., Toledo, T., & Prashker, J.N. (2008). Effects of Choice Set Size and Route Choice Models on Path-based Traffic Assignment. *Transportmetrica*, 4 (2), pp. 117-133.

Ben Akiva, M.E., Bergman, M.J., Daly, A.J. & Ramaswamy, R. (1984). Modelling Inter Urban Route Choice Behaviour. In *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. VNU Science Press, pp. 299-330.

Ben-Akiva, M.E. & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In *Handbook of Transportation Science*, pp. 5-33.

Ben-Akiva, M.E. & Bolduc, D. (1996). *Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure*. Working Paper, MIT, Cambridge, USA.

Ben-Akiva, M.E., Bolduc, D. & Daly, A., (1993). Estimation of Travel Choice Models with Randomly Distributed Values of Time. *Transportation Research Record: Journal of the Transportation Research Board*, 1413, pp. 88-97.

Ben-Akiva, M.E. & Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press.

Benjamins, M., Lindveld, K. & Van Nes, R. (2002). Multimodal Travel Choice Modelling: a Supernetwork Approach. *Proceedings of the 81th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Bielli, M., Boulmakoul, A. & Mouncif, H. (2006). Object Modeling and Path Computation for Multimodal Travel Systems. *European Journal of Operational Research*, 175, pp. 1705-1730.

Bierlaire, M. (2009). *Estimation of Discrete Choice Models with BIOGEME*. Version 1.8. Ecole Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory, Switzerland.

Bierlaire, M. & Frejinger, E. (2009). Route Choice Modeling with Network-free Data. *Transportation Research Part C*, 16, pp. 187-198.

Bliemer, M.C.J. & Bovy, P.H.L. (2008). Impact of Route Choice Set on Route Choice Probabilities. *Transportation Research Record: Journal of the Transportation Research Board,* 2076, pp. 10-19.

Bovy, P.H.L. (1996). Stochastic Traffic Assignment Technique Enhancements for Congested Networks. *Transportation Modelling for Tomorrow*. Delft University Press. The Netherlands.

Bovy, P.H.L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews*, 29 (1), pp. 43-68.

Bovy, P.H.L., Bekhor, S. & Prato, C.G. (2009). The Factor of Revisited Path Size: Alternative Derivation. *Transportation Research Record: Journal of the Transportation Research Board*, 2076, pp. 132-140.

Bovy, P.H.L. & Bradley, M.A. (1985). Route Choice Analyzed with Stated-preference Approaches. *Transportation Research Record: Journal of the Transportation Research Board*, 1307, pp. 11-20.

Bovy, P.H.L., & Fiorenzo-Catalano, S. (2007). Stochastic Route Choice Set Generation: Behavioral and Probabilistic Foundations. *Transportmetrica*, *3* (3), pp. 173–189.

Bovy, P.H.L. & Hoogendoorn-Lanser, S. (2005). Modelling Route Choice Behaviour in Multi-modal Transport Networks. *Transportation*, 32 (4), pp. 341-368.

Bricka, S. & Wolf, J. (2008). *Chicago Regional Household Travel Inventory Draft GPS – Final Report prepared for Chicago Metropolitan Agency for Planning*, NuStats, Austin.

Cascetta, E. (2001). *Transportation Systems Engineering: Theory and Methods*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Cascetta, E., Nuzzolo, A., Russo F., & Vitetta, A. (1996). A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and some Calibration Results for Interurban Networks. In *International Symposium on Transportation and Traffic Theory*, pp. 697-711.

Cascetta, E. & Papola, A. (2001). Random Utility Models with Implicit Availability/Perception of Choice Alternatives for the Simulation of Travel Demand. *Transportation Research Part C: Emerging Technologies*, 9 (4), pp. 249-263.

Cascetta, E. & Papola, A. (2003). A Joint Mode-Transit Service Choice Model Incorporating the Effect of Regional Transport Service Timetables. *Transportation Research Part B: Methodological*, 37 (7), pp. 595–614.

Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons Learned from the New York City Case Study. *Transportation Research Part A: Policy and Practice*, 44 (10), pp. 830-840.

Christensen, L. (2000). *Transportvaner og Kollektiv Trafikforsyning*, Report from DMU, no. 320, http://www2.dmu.dk/1_viden/2_Publikationer/3_fagrapporter/rapporter/fr320.pdf 2010-06-28, (in Danish).

Christensen, L. (2001). *Bystruktur og Transportadfærd – Hvad siger Transportvaneundersøgelsen?*. Report from DMU, no. 382, http://www2.dmu.dk/1_viden/2_Publikationer/3_fagrapporter/rapporter/FR382.pdf (in Danish), 2010-06-28.

Christensen, L. (2013). The Role of Web Interviews as Part of a National Travel Survey. In: *Transport Survey Methods: Best Practice for Decision Making*. Emerald Group Publishing Limited, 2013. pp. 115-153.

Christensen, L. & Jensen, T. (2008). Potentiale for Overflytning af Korte Bilture til Cykel og Gang. In *Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark (in Danish).

Christiansen, H. (2009). Vedr. Modernisering af Transportvaneundersøgelsen. In *Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark (in Danish).

Christiansen, H. & Haunstrup B. (2011). *The Danish National Survey - Declaration of Variables, TU 2006-10, version 2.* DTU Transport, Kgs. Lyngby, Denmark.

Chu, C. (1989). A Paired Combinatorial Logit Model for Travel Demand Analysis. *Proceedings of Fifth World Conference on Transportation Research,* 4**,** pp. 295–309.

Chung, E. H., & Shalaby, A. (2005). A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. *Transportation Planning and Technology*, 28 (5), pp. 381-401.

Clifton, K. & Muhs, C. (2012). Capturing and Representing Multimodal Trips in Travel Surveys: Review of the Practice. *Transportation Research Record: Journal of the Transportation Research Board,* 2285, pp. 74-83.

Daganzo, C.F. & Sheffi, Y. (1977). On Stochastic Models of Traffic Assignment. *Transportation Science*, 11, pp. 253-274.

Daly, A. & Gunn, H. (2002). Simultaneous Analysis of Choice and Transfer Price Data. *Proceedings of the European Transport Conference*.

Danish Ministry of Transport (2009). *Mobility that Creates Value.* The Strategic Foundation of the Danish Ministry of Transport, Danish Ministry of Transport, http://www.trm.dk/graphics/Synkron-Library/trafikministeriet/Publikationer/2009/Mobilitet%20der%20skaber%20v%E6rdi%20%20-%20netversion.pdf (in Danish), 2013-09-10.

De la Barra, T., Perez, B. & Anez, J. (1993). Multi-dimensional Path Search and Assignment. Proceedings of the *21st PTRC Summer Meeting*.

DTU (2007). *DTU's Årsrapport 2007*, http://wwwx.dtu.dk/upload/administrationen%20-%20101/om%20dtu/årsrapport%202007.pdf (in Danish), 2013-09-10.

Eluru, N., Chakour, V. & El-Geneidy, A.M. (2012). Travel Mode Choice and Transit Route Choice Behaviour in Montreal - Insights from McGill University Members Commute Patterns. *Public Transport*, 4, pp. 129-149.

European Commission (1995). *The Citizen's Network – Fulfilling the Potential of Public Passenger Transport in Europe.* Green Paper, Commission of the European Communities, Brussels, 1995.

European Commission (2001). *White paper European Transport Police for 2010: Time to Decide*. European Communities, http://ec.europa.eu/transport/strategies/doc/2001_white_paper/lb_texte_complet_en.pdf 2013-09-10.

Fingerplanen (2007): http://www.naturstyrelsen.dk/NR/rdonlyres/168AEF1C-EE66-4FE9-95D3-92B5D4452BFD/0/9788772797793.pdf 2013-09-10.

Fiorenzo-Catalano, S., van Nes, R., Bovy, P.H.L. (2004). Choice Set Generation for Multi-modal Travel Analysis. *European Journal of Transport and Infrastructure Research*, 4 (2), pp. 195-209.

Florian, M. (2004). Finding Shortest Time-dependent Paths in Schedule-based Transit Networks: A Label Setting Algorithm. In Wilson, N.H.M. and Nuzzolo, A. (Eds.), *Schedule-Based Dynamic Transit Modeling: Theory and Applications*, Kluwer Academic Publishers, pp. 43-52.

Fosgerau, M., Frejinger, E. & Karlstrom, A. (2013). A Link Based Network Route Choice Model with Unrestricted Choice Set. *Transportation Research Part B: Methodological*, 56, pp. 70-80.

Fosgerau, M., Hjorth, K., Lyk-Jensen, S.V. & Marott, J. (2007). *The Danish Value of Time Study – Final Report.* Report 5, Danish Transport Research Institute, http://www.dtu.dk/upload/institutter/dtu%20transport/pdf_dtf/rapporter/the%20danish%20value%20of%20time%20study_250208.pdf 2010-06-29.

Frejinger, E. (2007). Random Sampling of Alternatives in a Route Choice Context. In *Proceedings of the European Transport Conference*.

Frejinger, E., Bierlaire, M. & Ben-Akiva, M. (2009). Sampling of Alternatives for Route Choice Modeling. *Transportation Research Part B: Methodological*, 43 (10), pp. 984-994.

Frick, A., Bächtiger, M.T. & Reips, U.D. (2001). Financial Incentives, Personal Information and Drop Out Rate in Online Studies. Chapter 12 in Reaps, U.D. & Bosnjak, M. (Eds.), *Dimensions of Internet Science*, Pabst Science Publishers, pp. 209-220.

Friedrich, M., Hofsäss, I. & Wekeck, S. (2001). Timetable-based Transit Assignment using Branch and Bound Techniques. *Transportation Research Record: Journal of the Transportation Research Board*, 1752, pp. 100-107.

Gong, H., Chen, C., Bialostozky, E. & Lawson, C. (2012). A GPS/GIS Method for Travel Mode Detection in New York City. *Computers, Environment and Urban Systems*, 36, pp. 131-139.

Government of India (2006). *National Urban Transport Policy*. Ministry of Urban Development, Government of India, New Delhi, http://www.urbanindia.nic.in/ 2011-11-04

Halldórsdóttir, K. (2010). *Choice of Access Mode to Passenger Trains*. M.Sc. Thesis, DTU Transport, Technical University of Denmark, Kgs. Lyngby, Denmark.

Han, B., Algers, S. & Engelson, L., (2001). Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by using the Mixed Logit Model. *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Hensher, D. A. (2001). The Sensitivity of the Valuation of Travel Time Savings to the Specification of Unobserved Effects. *Transportation Research Part E: Logistics and Transportation Review*, 37 (2), pp. 129-142.

Hess, S., Train, K.E. & Polak, J.W. (2006). On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice. *Transportation Research Part B: Methodological*, 40 (2), pp. 147-163.

Holm, J. (2009). GPS Data som Grundlag for en National Rejsehastighedsdatabase. In *Proceedings for the Annual Transport Conference in Aalborg,* Aalborg University, Denmark (in Danish).

Hood, J., Sall, E. & Charlton, B. (2011). A GPS-based Bicycle Route Choice Model for San Francisco, California. *Transportation Letters*, 3 (1), pp. 63-75.

Hoogendoorn-Lanser, S. (2005). *Modelling Travel Behaviour in Multi-modal Networks*. PhD Thesis, TRAIL Research School: TRAIL Thesis Series T2005/4, Technical University of Delft, the Netherlands.

Hoogendoorn-Lanser, S., & Bovy, P. (2007). Modeling Overlap in Multimodal Route Choice by Including Trip Part-Specific Path Size Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2003, pp. 74–83.

Hoogendoorn-Lanser, S., Van Nes, R. & Bovy, P. (2005). Path Size Modeling in Multimodal Route Choice Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 1921, pp. 27-34.

Hoogendoorn-Lanser S., van Nes, R. & Bovy , P.H.L. (2006). A Rule-based Approach to Multi-modal Choice Set Generation. Paper presented at *the 11th International Conference on Travel Behaviour Research*, Kyoto.

Hoogendoorn-Lanser, S., van Nes, R. & Hoogendoorn, S.P. (2006). Modeling Transfers in Multimodal Trips: Explaining Correlations. *Transportation Research Record: Journal of the Transportation Research Board*, 1985, pp. 144-153.

Horn, M.E.T. (2003). An Extended Model and Procedural Framework for Planning Multi-modal Passenger Journeys. *Transportation Research Part B: Methodological*, 37 (7), pp. 641-660.

Itsubo, S. & Hato, E. (2006). A Study of the Effectiveness of a Household Travel Survey using GPS-equipped Cell Phones and a WEB Diary through a Comparative Study with a Paper Based Travel Survey. In *85th Annual Meeting of the Transportation Research Board, Washington, DC.*

Jan, O., Horowitz, A., & Peng, Z. (2000). Using GPS Data to Understand Variations in Path Choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1725, pp. 37-44.

Jensen, C. (2009). Danskernes Transport – Hvor Meget, Hvordan, Hvor og Hvornår? In *Proceedings for the Annual Transport Conference in Aalborg,* Aalborg University, Denmark (in Danish).

Jovicic, G. & Hansen, C.O. (2003). A Passenger Travel Demand Model for Copenhagen. *Transportation Research Part A: Policy and Practice*, 37 (4), pp. 333-349.

Koppelman, F.S. & Wen, C. (2000). The Paired Combinatorial Logit Model: Properties, Estimation and Application. *Transportation Research Part B: Methodological*, 34 (2), pp. 75-89.

Krygsman, S. & Dijst, M. (2001). Multimodal Trips in the Netherlands: Microlevel Individual Attributes and Residential Context. *Transportation Research Record: Journal of the Transportation Research Board*, 1753, pp. 11-19.

Krygsman, S., Dijst, M. & Arentze, T. (2004). Multimodal Public Transport: an Analysis of Travel Time Elements and the Interconnectivity Ratio. *Transport Policy*, 11 (3), pp. 265-275.

Larsen, M.K. (2008). Indsamling af Data for Rutevalg i Kollektiv Transport. *Selected proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark (in Danish).

Larsen, M.K., Nielsen, O.A., Prato, C.G. & Rasmussen, T.K. (2010). Generation and Quality Assessment of Route Choice Sets in Public Transport Networks by means of Data Analysis. In *Proceedings of the European Transport Conference*.

LINK (2010). *Recommendations and Strategies for Passenger Intermodality in Europe.*The European Forum on Intermodal Passenger Travel, Project funded by the European Commission, Directorate-General for Mobility and Transport (DG MOVE), 2010, http://www.linkforum.eu/index.phtml?ID1=932 2011-11-07.

Litman, T. (2008). Valuing Transit Service Quality Improvements. *Journal of Public Transportation*, 11 (2), pp. 43-63.

Lo, H.K., Yip, C.-W. & Wan, Q.K. (2004). Modeling competitive multi-modal transit services: A nested logit approach. *Transportation Research Part C: Emerging Technologies*, 12 (3), pp. 251-272.

Lorenc, T., Brunton, G., Oliver, S., Oliver, K. & Oakley, A. (2008). Attitudes to Walking and Cycling Among Children, Young People and Parents: A Systematic Review. *Journal of Epidemiology and Community Health*, 62 (10), pp. 852-857.

Lozano, A. & Storchi, G. (2001). Shortest Viable Path Algorithm in Multimodal Networks. *Transportation Research Part A: Policy and Practice,* 35 (3), pp. 225-241.

Luce, R. (1959). *Individual Choice: a Theoretical Analysis*. John Wiley & Sons, New York.

Mabit, S.L. & Nielsen, O.A. (2006). The Effect of Correlated Value of Travel Time Savings in Public Transport Assignment. In *Proceedings of the European Transport Conference*. Strasbourg.

Mahmassani, H.S., Joseph, T. & Jou, R.-C. (1993). Survey Approach for Study of Urban Commuter Choice Dynamics, *Transportation Research Record: Journal of the Transportation Research Board*, 1412, pp. 80-89.

Mandir, E., Pillat, J., Friedrich, M. & Schiller, C. (2010). Choice Set Generation and Model Identification for Route Choice Using GPSData from Smart Phones. In *Proceedings of the Conference on Innovations in Travel Modeling of the Transportation Research Board*, Tempe, Arizona.

Manski, C.T. (1973). *The Analysis of Qualitative Choice*. PhD Thesis. Department of Economics, MIT, Cambridge, Massachusetts.

Manski, C.T. (1977). The Structure of Random Utility Models. *Theory and Decision*, 8 (3), pp. 229-254.

Marschak, J. (1960). Binary Choice Constraints on Random Utility Indicators. In K. Arrow (Ed.), *Stanford Symposium on Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P. (Ed.), *Frontiers in Econometrics*, Academic Press, New York.

McFadden, D. (1978). Modeling the Choice of Residential Location. In: A. Karlqvist, L. Lundquist, F. Snickkars and J. Weibull (Eds.), *Spatial Interaction Theory and Residential Location*, pp. 75–96, Amsterdam, North Holland.

McFadden, D. & Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15 (5), pp. 447-470.

Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M., Waraich, R.A. & Axhausen, K.W. (2010). Large-scale Agent-based Travel Demand Optimization Applied to Switzerland, Including Mode Choice. Paper presented at the *12th World Conference on Transportation Research*, Lisbon.

Nielsen, O.A. (1998). A Large-scale Stochastic Multi-class Traffic Assignment Model for the Copenhagen Region. In *Proceedings of Triennial Symposium on Transportation Analyses*, Puerto Rico.

Nielsen, O.A. (2000). A Stochastic Transit Assignment Model Considering Differences in Passengers Utility Functions. *Transportation Research Part B: Methodological*, 34 (5), pp. 377-402.

Nielsen, O.A. (2004a). A Large Scale Stochastic Multi-class Schedule-based Transit Model with Random Coefficients. Chapter 4 in Wilson, N.H.M. and Nuzzolo, A. (Eds.), *Schedule-Based Dynamic Transit Modeling: Theory and Applications*, Kluwer Academic Publishers.

Nielsen, O.A. (2004b). Behavioural Responses to pricing Schemes: Description of the Danish AKTA Experiment. *Journal of Intelligent Transportation Systems*, 8 (4), pp. 233-251.

Nielsen, O.A., Daly, A. & Frederiksen, R.D. (2002). A Stochastic Route Choice Model for Car Travellers in the Copenhagen Region. *Networks and Spatial Economics*, 2 (4), pp. 327-346.

Nielsen, O.A., & Frederiksen, R.D. (2006). Optimisation of Timetable-based, Stochastic Transit Assignment Models based on MSA. *Annals of Operations Research*, 144 (1), pp. 263-285.

Nielsen, O.A., & Frederiksen, R.D. (2009). Large-scale Schedule-based Transit Assignment – Further Optimisation of the Solution Algorithms. Chapter 7 in Wilson, N.H.M. and Nuzzolo, A. (Eds.), *Schedule-Based Modeling of Transportation Networks*, Kluwer Academic Publisher.

Nielsen, O.A., Hansen, C.O. & Daly, A. (2001). A Large-scale Model System for the Copenhagen-Ringsted Railway Project. In Hensher, D. (Ed.), *Travel Behaviour Research: The Leading Edge*, Pergamon Press, pp. 603-626.

Nielsen, O.A. & Johansen, M.H. (2012). Køreplansbaseret Rutevalgsmodel og Matricer for Kollektiv Trafik i Landstrafikmodellen v1.0. In *Proceedings for the Annual Transport Conference in Aalborg*, Aalborg University, Denmark (in Danish).

Nuzzolo, A. & Crisalli, U. (2004). The Schedule-Based Approach in Dynamic Transit Modelling: A General Overview. In Wilson, N.H.M. and Nuzzolo, A. (Eds.), *Schedule-Based Dynamic Transit Modeling: Theory and Applications*, Kluwer Academic Publishers, pp. 1-24.

Nuzzolo, A., Russo, F. & Crisalli, U (1997). A Pseudo-dynamic Assignment to Extraurban Networks using a C-logit Route Choice Model. In *25th European Transport Forum, Proceedings of Seminar F, Transportation Planning Methods*, II, pp. 95-105.

Nuzzolo, A., Russo, F. & Crisalli, U. (2001). A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. *Transportation Science*, 35 (3), pp. 268-285.

NYMTC (2000). *Regional Travel - Household Interview Survey*. New York Metropolitan Transportation Council (NYMTC) and North Jersey Transportation Planning Authority (NJTPA). Prepared by NuStats International in association with Parsons Brinckerhoff Quade & Douglas, Inc.

Ohmori, N., Nakazato, M., Harata, N., Sasaki, K. & Nishii, K. (2006). Activity Diary Surveys using GPS Mobile Phones and PDA. In *85th Annual Meeting of the Transportation Research Board, Washington*, DC, pp. 22-26.

Ortúzar J.d.D. & Willumsen L.G. (2011). *Modelling Transport*. Fourth edition, Wiley, West Sussex, England.

Pelletier, M. P., Trépanier, M. & Morency, C. (2011). Smart Card Data Use in Public Transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19 (4), pp. 557-568.

Porter, S.R. & Whitcomb, M.E. (2003). The Impact of Lottery Incentives on Survey Response Rates. *Research in Higher Education*, 44 (4), pp. 389-407.

Prashker, J.N. & Bekhor S. (1998). Investigation of Stochastic Network Loading Procedures. *Transportation Research Record: Journal of the Transportation Research Board,* 1645, pp. 94-102.

Prashker, J.N. & Bekhor, S. (2004). Route Choice Models used in the Stochastic User Equilibrium Problem: A Review. *Transport reviews*, 24 (4), pp. 437-463.

Prato, C.G. (2005). *Latent Factors and Route Choice Behaviour.* PhD Thesis, Turin Polytechnic, Italy.

Prato, C.G. (2009). Route Choice Modeling: Past, Present and Future Research Directions. *Journal of Choice Modelling*, 2 (1), pp. 65-100.

Prato, C.G. (2012). Meta-analysis of Choice Set Generation Effects on Route Choice Model Estimates and Predictions. *Transport*, 27 (3), pp. 286-298.

Prato, C.G. & Bekhor, S. (2006). Applying Branch-and-bound Technique to Route Choice Set Generation. *Transportation Research Record: Journal of the Transportation Research Board*, 1985, pp. 19-28.

Prato C.G. & Bekhor S. (2007). Modeling Route Choice Behavior: How Relevant is the Composition of the Choice Set? *Transportation Research Record: Journal of the Transportation Research Board*, 2003, pp. 64-73.

Ramming, M.S. (2002). *Network Knowledge and Route Choice*. PhD Thesis. Massachusetts Institute of Technology, Cambridge, USA.

Rasmussen, T.K. (2010). *Rutevalg i Kollektiv Transport i Hovedstadsområdet – Generering og Kvalitetsanalyser af Valgsæt*. M.Sc. Thesis, DTU Transport, Technical University of Denmark, Kgs. Lyngby, Denmark (in Danish).

Rasmussen, T. K., Ingvardsson, J.B., Nielsen, O. A. & Halldórsdóttir, K. (2013). Enhancing the Use of Global Positioning Systems Data from Wearable GPS Devices as Travel Survey Method: A Case Study in the Copenhagen Region. In *2nd Symposium of the European Association for Research in Transportation*, Stockholm, Sweden.

Raveau, S. Muñoz, J.C. & de Grange, L. (2011). A Topological Route Choice Model for Metro. *Transportation Research Part A: Policy and Practice*, 45 (2), pp. 138-147.

Rejseplanen (2011). *Rejseplanens API Documentation*. Documentation for the Danish Travel Planner, available at http://labs.rejseplanen.dk/files/api/rest_documentation_latest.pdf 2013-09-10 (in Danish).

Rich, J.H., Nielsen, O.A. & Hansen, C.O. (2010). Overall Design of the Danish National Transport Model. Proceedings for the *Annual Transport Conference in Aalborg,* Aalborg University.

Rosenbloom, S. (2004). Mobility of the Elderly – Good News and Bad News. In *Conference Proceedings Transportation Research Board, 27, Transportation in an Aging Society – A Decade of Experience*, technical papers and reports from a conference November 7-9, 1999, Bethesda, Maryland, pp. 3-21.

Scherer, M. & Dziekan, K. (2012). Bus or Rail - an Approach to Explain the Psychological Rail Factor. *Journal of Public Transportation*, 15 (1), pp. 75-93.

Schönfelder, S., Axhausen, K., Antille, N. & Bierlaire, M. (2002). Exploring the Potentials of Automatically Collected GPS Data for Travel Behaviour Analysis – A Swedish data source. In Möltgen, J. and Wytzisk, A. (Eds.), *GI-Technologien für Verkehr und Logistik*, number 13 in *IfGIprints*. Institut für Geoinformatik, Universität Münster, Münster, pp. 155-179.

Schuessler, N. & Axhausen, K.W. (2009). Processing Raw Data from Global Positioning Systems without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board,* 2105, pp. 28-36.

Seaborn, C., Attanucci, J. & Wilson, N.H. (2009). Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board,* 2121, pp. 55-62.

Sheffi, Y. & Powell, W.B. (1982). An Algorithm for the Equilibrium Assignment Problem with Random Link Times. *Networks*, 12 (2), pp.191-207.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, USA.

Slavin, H., Rabinowicz, A., Brandon, J., Flammia, G. & Freimer, R. (2009). Using Automated Fare Collection Data, GIS, and Dynamic Schedule Queries to Improve Transit Data and Transit Assignment Model. Chapter 6 in Wilson, N.H.M. & Nuzzolo, A. (Eds.), *Schedule-Based Modeling of Transportation Networks,* Kluwer Academic Publisher, pp. 101-118.

Stopher, P.R. & Greaves, S.P. (2007). Household Travel Surveys: Where are We Going? *Transportation Research Part A: Policy and Practice*, 41 (5), pp. 367-381.

Sun, Y. & Xu, R. (2012). Rail Transit Travel Time Reliability and Estimation of Passenger Route Choice Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2275, pp. 58-67.

Thurstone, L. (1927). A Law of Comparative Judgment. *Psychological Review*, 34 (4), pp. 273-286.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Trépanier, M., Tranchant, N. & Chapleau, R. (2007). Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. *Journal of Intelligent Transportation Systems: Technology, Planning and Operations*, 11 (1), pp. 1-14.

Tsui, S.Y.A. & Shalaby, A. (2006). Enhanced System for Link and Mode Identification for GPS-Based Personal Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board,* 1972, pp. 38–45.

Umbach, P.D. (2004). Web surveys: Best practices. *New Directions for Institutional Research*, 121, pp. 23-38.

Utsunomiya, M., Attanucci, J. & Wilson, N. (2006). Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 1971, pp. 119-126.

Van der Waard, J. (1988). The Relative Importance of Public Transport Trip Time Attributes in Route Choice. In *Proceedings PTRC Summer Annual Meeting, 16th*, Bath, United Kingdom.

Van Nes, R. (2002). *Design of Multimodal Transport Networks – A Hierarchical Approach*. PhD Thesis, TRAIL Research School: TRAIL Thesis Series T2002/5, Technical University of Delft, the Netherlands.

Van Nes, R. & Hoogendoorn-Lanser, S. (2006). On the Use of Choice Sets for Estimation and Prediction in Route Choice. In *Proceedings for the CVS Congress*, Amsterdam, the Netherlands.

Vovsha, P. (1997). The Cross-Nested Logit Model: Application to Mode Choice in the Tel Aviv Metropolitan Area. *Transportation Research Record: Journal of the Transportation Research Board*, 1607, pp. 13-20.

Vovsha, P. & Bekhor, S. (1998). The Link-Nested Logit Model of Route Choice: Overcoming Route Overlapping Problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1645, pp. 133-142.

Vrtic, M. & Axhausen, K.W. (2002). Impact of Tilting Trains in Switzerland: Route Choice Model of Regional and Long-Distance Public Transport Trips. Presented at *82nd Annual Meeting of the Transportation Research Board*, Washington, D.C.

Vrtic, M., Schüssler, N., Erath, A., Axhausen, K., Frejinger, E., Bierlaire, M., Stojanovic, S., Rudel, R. & Maggi, R. (2006). *Including Travelling Costs in the Modelling of Mobility Behaviour*. Final report for SVI research program Mobility Pricing: Project B1, on behalf of the Swiss Federal Department of the Environment, Transport, Energy and Communications, IVT ETH Zurich, ROSO EPF Lausanne and USI Lugano.

Walmsley, D.J. & Jenkins, J.M. (1992). Cognitive Distance: A Neglected lssue in Travel Behavior. *Journal of Travel Research,* 31 (1), pp. 24-29.

Wardrop, J.G. (1952). Some Theoretical Aspects of Road Traffic Research. *Proceedings, Institute of Civil Engineers, London*, PART II, 1, pp. 325-378.

Wen, C. & Koppelman F. (2001). The Generalized Nested Logit Model. *Transportation Research Part B: Methodological*, 35 (7), pp. 627-641.

Wilson, N.H., Zhao, J. & Rahbee, A. (2009). The Potential Impact of Automated Data Collection Systems on Urban Public Transport Planning. Chapter 5 in Wilson, N.H.M. & Nuzzolo, A. (Eds.), *Schedule-Based Modeling of Transportation Networks*, Kluwer Academic Publisher, pp. 75-99.

Wolf, J. (2004). Applications of New Technologies in Travel Surveys. Paper presented at the *7th International Conference on Travel Survey Meth*ods, Costa Rica.

Yai, T., Iwakura, S. & Morichi, S, (1997). Multinomial Probit with Structured Covariance for Route Choice Behavior. *Transportation Research Part B: Methodological*, 31 (3), pp. 195-207.

Zabic, M. (2011). *GNSS-based Road Charging Systems – Assessment of Vehicle Location Determination*. PhD Thesis, DTU Transport, Kgs. Lyngby, Denmark.

Zhao, J. (2004). *The Planning and Analysis Implications of Automatic Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modelling Examples*. M.Sc. Thesis, MIT, USA.

Zhou, Z., Chen, A. & Bekhor, S. (2012). C-logit Stochastic User Equilibrium Model - Formulations and Solution Algorithm. *Transportmetrica*, 8 (1), pp. 17-41.

Ziliaskopoulos, A. & Wardell, W. (2000). An Intermodal Optimum Path Algorithm for Multimodal Networks with Dynamic Arc Travel Times and Switching Delays. European Journal of Operational Research, 125 (3), pp. 486-502.

http://www.maps.google.com 2013-09-10.

http://mapsof.net 2013-09-10.

http://www.moviatrafik.dk 2013-09-10.

http://www.trm.dk/da/publikationer/2010/redegørelse+om+kollektiv+trafik/ 2013-09-10.

APPENDIXES

APPENDIX 1

Focus group interviews

The following are made on the basis of two focus group interviews carried out at the Technical University of Denmark by the PhD student Leise Neel Jansen in September 2003. The description builds on the Danish documentation of the focus group interviews.

In September 2003 two focus group interviews concerning route choice in public transportation were carried out at the Technical University of Denmark (DTU). In total 17 Danish students travelling to DTU using public transportation several times a week participated in the focus groups. The group members were in the age of 21-40 years (in average 22.9 years in focus group 1 and 25.4 in focus group 2). Six men and 11 women participated. Only two of the participants (both in focus group 1) had a car at their disposal. The average travel time per trip was 46.4 minutes for focus group 1 and 36.5 for focus group 2.

During the conversations the students were guided by predetermined questions making sure the group members touched upon the relevant topics. In the following the main findings from the interviews are presented.

Focus group 1

The participants all agreed that the www.rejseplanen.dk was the preferred option for investigating new routes from A to B. They also used advices from friends and own experiences. When deciding on the preferred route the participants stuck to the route and stayed updated with changes in the yearly revision of the timetables.

All participants in the group agreed that the *travel time* was the most important factor in choosing route and improvement of others factors should preferably not cause extra travel time. Another important factor was the *frequency*. Most group members agreed that they preferred routes with high frequency because a high frequency meant that if they missed a departure (if being too late or the departure was cancelled) the waiting time for the next departure was acceptable. One focus group member who used a route with a long travel time and a low departure frequency did not agree on this and stated that if he had just one suitable departure he would make sure to be at the stop in time for departure.

The participants had different opinions of *transfers* in the network. Some participants were very reluctant to transferring because it disturbed the trip and the risk of losing a connection was thought of as a stress factor. Other participants did not see the transfers as a problem and were not unwilling to transfer if they could achieve a shorter total travel time. One participant stated that the reluctance depended on the general regularity of the transport modes transferring from and to. Some participants preferred to transfer at locations where it was possible to see from the alighting stop to the next boarding stop in order to know whether the connection was within reach and whether the traveller should hurry up. Also transfer locations with shops etc. were preferred. These extra factors were however only important for the transfer location if all other things were equal.

Regarding *regularity* the participants preferred routes with seldom delays.

The participants did not consider *comfort* much when choosing route. If the travel time was the same they would prefer the route with highest comfort but they would not change to a slower route in order to improve comfort.

By default the participants always started their route from the *stop location* closest to the origin. When walking the participants were more flexible at summertime compared to wintertime because they associated the walking with risk of getting delayed especially when walking for a long distance. At the destination the location of the departure stop was not quite as important because the egress walk could not cause additional delays in the system. Many participants did not consider the option of biking to a stop further away because of the risk of theft. Participants were also not willing to choose a stop located further away in order to save money because most participants (6 of 8) had public transport season tickets.

Also the opportunity of *travelling along with others* was important for the route choice of the focus group participants.

The participants typically selected route according to the *time of day* (or home end/activity end). Going to DTU in the morning they often travelled by their perceived best route which for most was the route with shortest travel time. In the afternoon they were more willing to travel by a less optimal route in order to improve other factors such as the view or travelling along with others.

They all *decided from home* which route to use. It was not possible for all participants to change route underway but those who had the opportunity were willing to change route along the way if a delay made another route faster.

All participants had one route which they almost always used and did not change between routes. In order to change route they should be presented to a route with *shorter travel time* and some wanted *fewer transfers*. Also they were willing to change to a *cheaper* route but they did not think of this to be realistic because of the fare structure (se chapter 2).

In terms of *mode* the participants preferred fast buses and trains but only if all other things were equal. Some preferred bus over train because they found it easier to cheat with the fare payment in the bus. Some participants sometimes used bicycle instead of public transportation to DTU explained by the opportunity to be more independent, save money, get exercise and they believed that the bicycle was more trustworthy than the bus company.

Focus group 2

In this group the participants used the www.rejseplanen.dk to discover routes but they also listened to the experiences of others. Several participants had knowledge of the area prior to their study at DTU and chose to use buses they knew in advance.

Of factors important to their route choice the participants mentioned the *regularity* of the public transport modes. Some participants felt that they did not have any alternatives to the route used and would not change route despite the poor regularity of the mode(s) on the chosen route. Some chose a specific route in order to avoid transport modes with poor regularity (preferred train over bus); others try to avoid passing the train station in Lyngby because the high number of travellers in the peak periods and the transfer delay.

The *frequency* on the route was also considered important. According to the participants a high frequency has many advantages such as i) if the bus is full and the traveller cannot board the next vehicle will arrive shortly, ii) it is less stressful not to have to be at the stop for a specific departure, iii) many possible departures to chose from mean more freedom when the end of the activity (school) varies, and iv) if the route involves a transfer it is preferable that the vehicle transferred to has a high frequency so the traveller has more options if the vehicle on the first trip leg is delayed. For a route with a low frequency it was perceived even more important that the transport vehicle leaved on time and especially did not leave before time which some participants often experienced.

In the choice of *departure time* the participants were willing to use an earlier departure than the desired in order to obtain a better route in terms of comfort (less noise, seating possible, less transfers, etc.).

Regarding *comfort* the participants had both pros and cons for the various public transport modes. Some preferred train because of the more smooth movements without hard breaking and only few stops. Also in the trains it is easier to find seating and the boarding and alighting is easier. Others preferred the bus because they liked the opportunity to sit more privately in the bus (often two person seats compared to 4 person groups in the trains). They also found the buses safer because of the face to face contact to the bus driver. On the other hand the same persons found that the indoor environment in the buses was poor (always too warm). Most group members agreed that the comfort factor only affected their route choice if all other things were equal. Only on days with nice weather they would rather walk/bicycle for shorter distances (about 20 min of walking) than use bus because of the poor comfort. But two participants always used bicycle from the train station in Lyngby to DTU (approx. 3.5 km) because of the poor comfort in the buses serving the same route.

The *travel time* was also important for the members in this group. Compared to focus group 1 they were however more willing to accept a longer travel time in order to obtain other advantages such as fewer transfers and shorter access and egress walking distances. The importance of the travel time depended on the *time of day*. In the morning the travel time should be as short as possible, and in the afternoon the chosen route was not necessarily the shortest if other routes had other advantages. In the evening the travel time was ascribed little importance compared to for example safety and comfort. During daytime the extra travel time to obtain other advantages should not exceed more than 10-15 minutes.

The participants preferred not to *transfer*. Some did not have a real choice whether to transfer or not and some accepted a transfer to obtain a better route. They believed that the reluctance against transferring was very much dependent on the weather as they found transferring more acceptable on a nice summer day than on a wet autumn day. The participants associated the transfers with uncertainty because a delay often leads to loss of a connection. In order to avoid this they planned a time buffer when transferring which was thought of as time waste if the transport modes were not delayed. When having transfers as part of the route the waiting time at the stop should not be too long but also not too short. The participants had difficulties agreeing on the preferred time for transferring, some said 2-3 minutes other said 10-15 minutes. They wanted the transfer time to be as short as possible without risking the connection to be lost.

The participants almost always preferred to use the *stop closest to* and were unwilling to walk for longer distances in order to save travel time, money etc. They would often prefer to drive extra time in bus instead of walking. They would also not use bicycle to a stop further away in order to save a fare zone because of too low fare difference and the fact that most of them had a public transport season ticket already including the first zone (the home zone).

Typically the participants used the *same route* going to DTU everyday and seldom checked for new attractive routes. On the return trip they chose different routes depending on the time and whether they had errands or other activities along the way.

Sometimes the participants preferred other *modes* than public transport for example because of the independency in choice of departure time offered by the private modes such as bicycle and car and the free choice of scenery. The car was mentioned for its comfort and the bicycle for the cheapness and exercise factor.

Summing up on focus groups

Two focus group interviews among in total 17 DTU students were carried out in September 2003. The group discussions pinpointed to a number of factors being important for these travellers in their daily choice of route in public transportation going between home and study at DTU.

The important characteristics of the multimodal public transport network discussed were:

- **Travel time**
    - The total travel time is the most important factor in route choice.
    - A limited increase in travel time is accepted if transfers can be avoided or the waiting time can be reduced.

  - Travel time is by far the most important factor when choosing route in the morning. In the afternoon other factors also affects the route choice.
- **Mode characteristics**
  - High frequency of public transport modes important to assure another departure if the first is missed.
  - Good regularity is important.
- **Transfers**
  - Transfers are preferably avoided.
  - Waiting time at transfers should be short (but not too short).
  - At transfer locations good visibility from alighting stop to boarding stop is important.

The factors in the above are all interesting to investigate further. The discussions offer an idea of the factors important for the choice of route but are also the statements of these specific individuals. All focus group members travel for the same purpose, at the same destination and at approximately the same time of the day. Such focus group discussions are easily affected by how willing the group members are to discuss and the views of the most talking group members can easily seem to be the views of the whole group even if this is not the case. The statements of the group can be used as a foundation for the following investigations and analyses of route choice determinants in this PhD thesis.

APPENDIX 2

Mode chain combinations of multiple-leg and multimodal trips used by 20 or more travellers as a percentage of all, home-end and activity-end trips

Mode chain				Number of trips			Percentage of multimodal		
Mode 1	Mode 2	Mode 3	Mode 4	All	Home	Activity	All	Home	Activity
Bus	Bus	-	-	506	236	270	10.63	4.96	5.67
Bus	S-train	-	-	267	154	113	5.61	3.23	2.37
S-train	Bus	-	-	250	91	159	5.25	1.91	3.34
Bicycle	S-train	-	-	158	128	30	3.32	2.69	0.63
S-train	Bicycle	-	-	146	5	141	3.07	0.10	2.96
S-train	Metro	-	-	130	67	63	2.73	1.41	1.32
Metro	Bus	-	-	127	37	90	2.67	0.78	1.89
Bus	Metro	-	-	124	71	53	2.60	1.49	1.11
Metro	S-train	-	-	117	37	80	2.46	0.78	1.68
Bicycle	S-train	Bicycle	-	116	54	62	2.44	1.13	1.30
Bus	S-train	Bus	-	104	54	50	2.18	1.13	1.05
Other train	Bus	-	-	96	36	60	2.02	0.76	1.26
Bus	Other train	-	-	91	46	45	1.91	0.97	0.94
S-train	S-train	-	-	91	41	50	1.91	0.86	1.05
Bicycle	Other train	-	-	88	73	15	1.85	1.53	0.31
Other train	Bicycle	-	-	81	6	75	1.70	0.13	1.57
Bicycle	Bus	-	-	72	60	12	1.51	1.26	0.25
Bus	Bicycle	-	-	68	6	62	1.43	0.13	1.30
Bus	Other train	Bus	-	64	33	31	1.34	0.69	0.65
Car Pass	S-train	-	-	55	40	15	1.15	0.84	0.31
S-train	Car Pass	-	-	51	7	44	1.07	0.15	0.92
S-train	Car Driver	-	-	50	0	50	1.05	0.00	1.05
Car Driver	S-train	-	-	46	39	7	0.97	0.82	0.15
S-train	Other train	-	-	45	18	27	0.94	0.38	0.57
Other train	Car Pass	-	-	42	3	39	0.88	0.06	0.82
Other train	S-train	-	-	41	20	21	0.86	0.42	0.44
Bicycle	Other train	Bus	-	39	37	2	0.82	0.78	0.04
Bicycle	S-train	Bus	-	38	34	4	0.80	0.71	0.08
Bus	S-train	Bicycle	-	38	2	36	0.80	0.04	0.76
Car Driver	Other	Car Driver	-	36	7	29	0.76	0.15	0.61
Car Pass	Other train	-	-	36	24	12	0.76	0.50	0.25
Bicycle	Other train	Bicycle	-	35	17	18	0.73	0.36	0.38
Car Driver	Other train	-	-	34	27	7	0.71	0.57	0.15
Bicycle	S-train	Metro	-	33	28	5	0.69	0.59	0.10
Bus	Other train	Bicycle	-	31	2	29	0.65	0.04	0.61

Continues next page....

Mode chain				Number of trips			Percentage of multimodal		
Mode 1	**Mode 2**	**Mode 3**	**Mode 4**	**All**	**Home**	**Activity**	**All**	**Home**	**Activity**
Metro	Bicycle	-	-	30	2	28	0.63	0.04	0.59
Bicycle	Metro	--	-	29	23	6	0.61	0.48	0.13
Bus	S-train	S-train	Metro	29	20	9	0.61	0.42	0.19
Other train	Car Driver	-	-	29	1	28	0.61	0.02	0.59
Car Pass	Other train	Bus	-	28	12	16	0.59	0.25	0.34
Bus	Other train	Car Pass	-	28	13	15	0.59	0.27	0.31
Bus	Bus	Bus	-	27	13	14	0.57	0.27	0.29
Metro	S-train	Bus	-	24	7	17	0.50	0.15	0.36
Car Pass	Other train	Car Pass	-	23	7	16	0.48	0.15	0.34
S-train	S-train	Bus	-	21	11	10	0.44	0.23	0.21
S-train	S-train	Bicycle	-	20	3	17	0.42	0.06	0.36
Metro	S-train	Bicycle	-	20	0	20	0.42	0.00	0.42
			Totals	3,654	1,652	2,002	76.73	34.69	42.04

APPENDIX 3

Table 0-1: Number of public transport legs for public transport mode for trip purpose and distance band

	Mode	Total trip distance band [km] All	<10	10-25	>25
All Trips	Bus	3,403	2,143	924	336
	S-train	2,488	924	1,067	497
	Metro	1,086	694	261	131
	Regional + IC-Train	636	76	165	395
	Local Train	169	41	62	66
	Sum	7,782	3,878	2,479	1,425
Work Trips	Bus	1,718	932	570	216
	S-train	1,463	451	707	305
	Metro	589	336	173	80
	Regional + IC-train	414	40	113	261
	Local Train	94	17	40	37
	Sum	4,278	1,776	1,603	899
Leisure Trips	Bus	1,685	1,211	354	120
	S-train	1,025	473	360	192
	Metro	497	358	88	51
	Regional + IC-train	222	36	52	134
	Local Train	75	24	22	29
	Sum	3,504	2,102	876	526

Table 0-2: Estimated parameter coefficients and (robust t test) for All Trips in two divisions of distance bands

All Trips	**Trip distance band**		
Parameter	**<10 km**	**10-25km**	**>25km**
In vehicle Time			
Bus	-0.179 (-24.5)	-0.112 (-17.9)	-0.125 (-13.7)
Local Train	-0.154 (-4.05)	-0.084 (-6.37)	-0.068 (-4.27)
Metro	-0.045 (-3.62)	-0.026 (-1.88)	-0.002 (-0.12)
Regional + IC-train	-0.196 (-6.73)	-0.106 (-7.00)	-0.090 (-8.19)
S-train	-0.089 (-8.29)	-0.047 (-7.21)	-0.100 (-10.4)
Access/Egress	-0.386 (-36.9)	-0.299 (-29.4)	-0.290 (-19.9)
Path Size Factor			
PSC	-0.798 (-12.1)	-0.745 (-6.20)	-0.457 (-3.38)
Transfers			
Waiting Time	-0.166 (-13.1)	-0.066 (-11.5)	-0.033 (-3.81)
Walking Time	-0.165 (-8.34)	-0.099 (-4.57)	-0.140 (-5.12)
No. Transfer	-2.010 (-29.9)	-1.880 (-26.5)	-1.850 (-13.9)
Number of estimated parameters:	10	10	10
Number of observations:	3,185	1,635	821
Number of individuals:	3,185	1,635	821
Null log-likelihood:	-11,108	-5,960	-3,115
Init log-likelihood:	-11,108	-5,960	-3,115
Final log-likelihood:	-6,351	-3,428	-1,779
Likelihood ratio test:	9,513	5,064	2,671
Rho-square:	0.428	0.425	0.429
Adjusted rho-square:	0.427	0.423	0.426

Table 0-3: Estimated parameter coefficients and (robust t test) for Work Trips in two divisions of distance bands

Work trips	**Total trip distance [km]**				
Parameter	**<10 km**	**10-25km**	**>25km**	**<20 km**	**>20km**
In vehicle Time					
Bus	-0.209 (-17.8)	-0.132 (-15.2)	-0.156 (-11.6)	-0.179 (-22.6)	-0.131 (-12.7)
Local Train	-0.176 (-2.87)	-0.107 (-6.47)	-0.118 (-5.96)	-0.138 (-6.41)	-0.100 (-6.38)
Metro	-0.061 (-3.44)	-0.050 (-2.92)	-0.015 (-0.63)	-0.062 (-4.50)	-0.001 (-0.03)
Regional + IC-train	-0.223 (-4.98)	-0.124 (-6.00)	-0.115 (-7.30)	-0.168 (-6.92)	-0.107 (-7.95)
S-train	-0.111 (-6.24)	-0.070 (-7.90)	-0.132 (-9.72)	-0.095 (-10.2)	-0.101 (-9.31)
Access/Egress	-0.427 (-23.8)	-0.312 (-22.7)	-0.330 (-15.7)	-0.382 (-30.7)	-0.315 (-19.1)
Path Size Factor					
PSC	-0.773 (-8.00)	-0.862 (-5.15)	-0.448 (-2.57)	-0.857 (-10.0)	-0.419 (-2.56)
Transfers					
Waiting Time	-0.128 (-8.40)	-0.072 (-9.32)	-0.050 (-6.27)	-0.100 (-12.2)	-0.053 (-7.55)
Walking Time	-0.166 (-5.79)	-0.100 (-3.25)	-0.177 (-4.81)	-0.137 (-6.63)	-0.150 (-3.58)
No. Transfer	-2.120 (-19.3)	-1.870 (-20.2)	-1.800 (-9.93)	-2.030 (-26.4)	-1.880 (-13.4)
Number of estimated parameters:	10	10	10	10	10
Number of observations:	1,392	1,045	515	2,213	739
Number of individuals:	1,392	1,045	515	2,213	739
Null log-likelihood:	-4,951	-3,818	-1,963	-7,938	-2,791
Init log-likelihood:	-4,951	-3,818	-1,963	-7,938	-2,791
Final log-likelihood:	-2,766	-2,148	-1,090	-4,456	-1,582
Likelihood ratio test:	4,371	3,339	1,745	6,965	2,418
Rho-square:	0.441	0.437	0.444	0.439	0.433
Adjusted rho-square:	0.439	0.435	0.439	0.437	0.430

Table 0-4: Estimated parameter coefficients and (robust t test) for Leisure Trips in two divisions of distance bands

Leisure trips	Total trips distance [km]				
Parameter	**<10 km**	**10-25km**	**>25km**	**<20 km**	**>20km**
In vehicle Time					
Bus	-0.154 (-16.6)	-0.083 (-9.64)	-0.087 (-7.17)	-0.125 (-17.7)	-0.081 (-7.19)
Local Train	-0.134 (-2.73)	-0.060 (-3.34)	-0.021 (-1.00)	-0.089 (-3.19)	-0.041 (-2.52)
Metro	-0.032 (-1.87)	0.005 (0.22)	0.025 (1.15)	-0.011 (-0.79)	-0.001 (-0.05)
Regional + IC-train	-0.174 (-4.56)	-0.084 (-4.03)	-0.052 (-3.63)	-0.120 (-5.33)	-0.052 (-3.55)
S-train	-0.072 (-5.60)	-0.009 (-1.04)	-0.054 (-4.31)	-0.042 (-4.86)	-0.036 (-3.13)
Access/Egress	-0.356 (-28.1)	-0.286 (-18.8)	-0.239 (-12.3)	-0.335 (-32.5)	-0.244 (-13.8)
Path Size Factor					
PSC	-0.825 (-9.19)	-0.576 (-3.30)	-0.513 (-2.41)	-0.762 (-9.48)	-0.707 (-3.23)
Transfers					
Walking Time	-0.179 (-6.56)	-0.103 (-3.57)	-0.075 (-1.85)	-0.148 (-7.12)	-0.086 (-2.43)
No. Transfer	-1.920 (-22.2)	-1.950 (-16.7)	-1.980 (-10.2)	-1.980 (-27.6)	-1.900 (-12.0)
Waiting Time	-0.219 (-10.8)	-0.057 (-6.68)	-0.016 (-1.34)	-0.148 (-12.9)	-0.018 (-1.66)
Number of estimated parameters:	10	10	10	10	10
Number of observations:	1,793	590	306	2,294	395
Number of individuals:	1,793	590	306	2,294	395
Null log-likelihood:	-6,153	-2,142	-1,152	-7,965	-1,481
Init log-likelihood:	-6,153	-2,142	-1,152	-7,965	-1,481
Final log-likelihood:	-3,551	-1,261	-669	-4,678	-874
Likelihood ratio test:	5,203	1,762	967	6,576	1,214
Rho-square:	0.423	0.411	0.420	0.413	0.410
Adjusted rho-square:	0.421	0.407	0.411	0.412	0.403

Table 0-5: Estimated parameter coefficients and (robust t test) for All, Work and Leisure Trips for PS Logit model with *First Headway* parameter

	Trip Purpose		
Parameter	**All**	**Work**	**Leisure + Other**
Headway			
½ of First	-0.060 (-7.99)	-0.055 (-5.32)	-0.067 (-6.30)
In vehicle Time			
Bus	-0.138 (-30.1)	-0.161 (-24.3)	-0.109 (-17.6)
Local Train	-0.103 (-9.44)	-0.136 (-11.2)	-0.069 (-3.56)
Metro	-0.030 (-3.58)	-0.047 (-4.12)	-0.009 (-0.72)
Regional + IC-train	-0.116 (-13.2)	-0.136 (-11.6)	-0.093 (-6.83)
S-train	-0.078 (-14.9)	-0.101 (-13.8)	-0.045 (-6.35)
Access/Egress	-0.352 (-50.8)	-0.373 (-37.2)	-0.330 (-34.3)
Path Size Factor			
PSC	-0.686 (-12.4)	-0.698 (-8.87)	-0.672 (-8.73)
Transfers			
Waiting Time	-0.082 (-13.5)	-0.081 (-13.9)	-0.084 (-6.92)
Walking Time	-0.128 (-9.83)	-0.133 (-7.14)	-0.127 (-7.17)
No. Transfer	-2.070 (-42.5)	-2.050 (-30.2)	-2.130 (-29.1)
Number of estimated parameters:	11	11	11
Number of observations:	5,641	2,952	2,689
Number of individuals:	5,641	2,952	2,689
Null log-likelihood:	-20,172	-10,722	-9,442
Init log-likelihood:	-20,172	-10,722	-9,442
Final log-likelihood:	-11,632	-6,030	-5,572
Likelihood ratio test:	17,079	9,385	7,741
Rho-square:	0.423	0.438	0.410
Adjusted rho-square:	0.423	0.437	0.409

Table 0-6: Estimated parameter coefficients and (robust t-test) for All Trips for final PS Logit model for the cut-at-10km and cut-at-20km distance band

All Trips	**Total trip distance**			
Parameter	**<10 km**	**>10km**	**<20 km**	**>20km**
Headway				
Up to 6 min	-0.376 (-7.13)	-0.029 (-0.48)	-0.289 (-6.51)	-0.020 (-0.21)
Above 6 min	-0.080 (-9.44)	-0.021 (-2.64)	-0.066 (-9.80)	-0.010 (-0.89)
In vehicle Time				
Bus	-0.169 (-21.9)	-0.106 (-21.2)	-0.144 (-26.2)	-0.097 (-0.21)
Local Train	-0.178 (-4.41)	-0.062 (-7.73)	-0.138 (-7.96)	-0.044 (-0.89)
Metro	-0.101 (-6.96)	-0.066 (-5.53)	-0.087 (-8.02)	-0.060 (-14.4)
Regional + IC-train				
Up to 20 km	-0.006 (-7.42)	-0.075 (-7.19)	-0.040 (-9.55)	-0.086 (-4.01)
Above 20 km	-0.259 (-0.17)	-0.127 (-6.95)	-0.211 (-1.47)	-0.092 (-7.24)
S-train	-0.130 (-10.5)	-0.078 (-13.0)	-0.096 (-13.8)	-0.086 (-9.82)
Access/Egress	-0.413 (-36.8)	-0.324 (-35.5)	-0.385 (-45.1)	-0.307 (-4.10)
Path Size Factor				
PSC	-0.662 (-9.58)	-0.608 (-6.45)	-0.698 (-11.4)	-0.548 (-7.24)
Transfers				
Walking Time	-0.211 (-7.74)	-0.104 (-5.07)	-0.159 (-8.30)	-0.126 (-4.01)
Waiting Time	-0.163 (-12.6)	-0.050 (-8.59)	-0.116 (-16.5)	-0.037 (-3.92)
No. Transfer				
Bus -> Bus	-2.310 (-24.9)	-2.750 (-17.6)	-2.460 (-30.3)	-2.630 (-9.82)
Bus -> Train	-1.500 (-10.4)	-2.130 (-17.9)	-1.740 (-15.9)	-2.190 (-3.92)
Train -> Bus	-1.780 (-11.8)	-2.300 (-19.1)	-2.010 (-17.5)	-2.250 (-4.10)
Train -> Train	-1.260 (-11.6)	-1.080 (-15.4)	-1.130 (-16.1)	-1.140 (-7.88)
Number of estimated parameters:	16	16	16	16
Number of observations:	3,185	2,456	4,507	1,134
Null log-likelihood:	-11,108	-9,070	-15,908	-4,272
Final log-likelihood:	-6,156	-5,172	-8,911	-2,470
Likelihood ratio test:	9,903	7,796	13,994	3,603
Adjusted rho-square:	0.444	0.428	0.439	0.418

Table 0-7: Estimated parameter coefficients and (robust t-test) for Traveller gender for the final PS Logit model

	Traveller gender			
Parameter	**Men**		**Women**	
Headway				
Up to 6 min	-0.373	(-6.30)	-0.210	(-4.03)
Above 6 min	-0.055	(-5.76)	-0.037	(-4.56)
In vehicle Time				
Bus	-0.140	(-19.4)	-0.126	(-22.1)
Local Train	-0.071	(-4.83)	-0.081	(-6.37)
Metro	-0.074	(-5.23)	-0.077	(-6.56)
Regional + IC-train				
Up to 20 km	-0.183	(-7.16)	-0.157	(-7.22)
Above 20 km	-0.094	(-6.15)	-0.074	(-4.13)
S-train	-0.102	(-12.0)	-0.090	(-12.2)
Access/Egress	-0.380	(-32.5)	-0.357	(-39.1)
Path Size Factor				
PSC	-0.683	(-8.08)	-0.680	(-9.14)
Transfers				
Walking Time	-0.111	(-4.20)	-0.173	(-7.98)
Waiting Time	-0.075	(-8.79)	-0.084	(-9.75)
No. Transfer				
Bus -> Bus	-2.63	(-21.1)	-2.52	(-24.3)
Bus -> Train	-2.14	(-15.0)	-1.71	(-13.7)
Train -> Bus	-2.17	(-13.3)	-2.02	(-17.5)
Train -> Train	-1.19	(-13.1)	-1.17	(-14.5)
Number of estimated parameters:	16		16	
Number of observations:	2,237		3,395	
Null log-likelihood:	-8,031		-12,106	
Final log-likelihood:	-4,495		-6,999	
Likelihood ratio test:	7,072		10,213	
Adjusted rho-square:	0.438		0.421	

Table 0-8: Estimated parameter coefficients scaled to Bus IVT (=1.0) for Traveller gender for PS Logit model

	Traveller gender	
Parameter	**Men**	**Women**
Headway		
Up to 6 min	2.7	1.7
Above 6 min	0.4	0.3
In vehicle Time		
Bus	1.0	1.0
Local Train	0.5	0.6
Metro	0.5	0.6
Regional + IC-train		
Up to 20 km	1.3	1.2
Above 20 km	0.7	0.6
S-train	0.7	0.7
Access/Egress	2.7	2.8
Transfers		
Walking Time	0.8	1.4
Waiting Time	0.5	0.7
No. Transfer		
Bus -> Bus	18.8	20.0
Bus -> Train	15.3	13.6
Train -> Bus	15.5	16.0
Train -> Train	8.5	9.3

Table 0-9: Estimated parameter coefficients and (robust t-test) for Traveller occupation for the final PS Logit model

Parameter	Traveller occupation: Student		Unemployed		Employed	
Headway						
Up to 6 min	-0.299	(-3.92)	-0.066	(-0.70)	-0.327	(-6.26)
Above 6 min	-0.053	(-6.47)	-0.059	(-5.80)	-0.033	(-2.76)
In vehicle Time						
Bus	-0.117	(-15.4)	-0.102	(-10.2)	-0.156	(-24.4)
Local Train	-0.076	(-5.76)	-0.083	(-4.58)	-0.089	(-4.98)
Metro	-0.058	(-4.00)	-0.032	(-1.46)	-0.108	(-7.67)
Regional + IC-train						
Up to 20 km	-0.122	(-4.60)	-0.225	(-3.82)	-0.200	(-9.06)
Above 20 km	-0.062	(-3.66)	-0.067	(-1.80)	-0.103	(-6.17)
S-train	-0.079	(-7.68)	-0.060	(-5.26)	-0.121	(-15.7)
Access/Egress	-0.350	(-29.5)	-0.392	(-20.6)	-0.378	(-35.6)
Path Size Factor						
PSC	-0.591	(-5.93)	-0.594	(-4.65)	-0.759	(-9.43)
Transfers						
Walking Time	-0.112	(-3.66)	-0.255	(-6.49)	-0.137	(-5.88)
Waiting Time	-0.058	(-9.06)	-0.106	(-9.54)	-0.091	(-7.84)
No. Transfer						
Bus -> Bus	-2.47	(-20.2)	-2.37	(-16.0)	-2.77	(-19.1)
Bus -> Train	-1.7	(-10.4)	-1.48	(-6.76)	-2.14	(-15.5)
Train -> Bus	-1.66	(-10.4)	-1.92	(-9.01)	-2.41	(-16.7)
Train -> Train	-1.17	(-9.78)	-1.31	(-8.27)	-1.14	(-14.7)
Number of estimated parameters:	16		16		16	
Number of observations:	1,720		1,091		2,821	
Null log-likelihood:	-6,117		-3,757		-10,252	
Final log-likelihood:	-3,638		-2,108		-5,630	
Likelihood ratio test:	4,957		3,298		9,244	
Adjusted rho-square:	0.403		0.435		0.449	

Table 0-10: Estimated parameter coefficients scaled to Bus IVT (=1.0) for Traveller occupation for PS Logit model

	Traveller occupation		
Parameter	**Student**	**Unemployed**	**Employed**
Headway			
Up to 6 min	2.6	0.7	2.1
Above 6 min	0.5	0.6	0.2
In vehicle Time			
Bus	1.0	1.0	1.0
Local Train	0.7	0.8	0.6
Metro	0.5	0.3	0.7
Regional + IC-train			
Up to 20 km	1.0	2.2	1.3
Above 20 km	0.5	0.7	0.7
S-train	0.7	0.6	0.8
Access/Egress	3.0	3.8	2.4
Transfers			
Walking Time	1.0	2.5	0.9
Waiting Time	0.5	1.0	0.6
No. Transfer			
Bus -> Bus	21.1	23.2	17.8
Bus -> Train	14.5	14.5	13.7
Train -> Bus	14.2	18.8	15.4
Train -> Train	10.0	12.8	7.3

DTU Transport performs research and provides education on traffic and transport planning. It advises the Danish Ministry of Transport on infrastructure, economic appraisals, transport policy and road safety and collects data on the transport habits of the population. DTU Transport collaborates with companies on such topics as logistics, public transport and intelligent transport systems.

DTU Transport
Department of Transport
Technical University of Denmark

Bygningstorvet 116B
2800 Kgs. Lyngby
Tel. +45 45 25 65 00

www.transport.dtu.dk