Route choice decisions of e-bike users: GPS tracking data from the Noord-Brabant Region of The Netherlands

Thesis

Floor Luub

Graduation committee:
Chairman: Prof.dr. T.A. Arentze, Eindhoven University of Technology.
1st supervisor: Dr. G.Z. Dane, Eindhoven University of Technology.
2e supervisor: Dr. T. Feng, Eindhoven University of Technology.

Construction Management and Engineering
23-08-2018

Preface

This thesis is the final element in the fulfilment of my master's degree Construction Management and Engineering at the Eindhoven University of Technology. While writing this thesis hasn't always been easy, I did learn that I enjoy doing research even the, most dreaded by other students, statistical part. I could not have finished this thesis without the support of several people and would like to take the opportunity to thank them.

First, I want to thank Theo Artenze for his guidance during the final stages of this research, asking some critical questions about the methodology and making sure I understood the choice models applied.

In addition, I would like to thank my supervisors Tao Feng and Gamze Dane, I enjoyed our two-weekly meetings during which we discussed the issues I encountered, and you guided me into the best direction, or, we collaboratively tried to figure out why Nlogit was giving errors as output. Tao, without you there would be no dataset. Gamze, your door was always open when I was stuck with my thesis or on a personal level.

Last but not least, my husband Sjoerd and my parents for their unconditional support and faith in the outcome.

Floor Luub,

Uden, August 2018

Table of contents

Summary
In the Netherlands the increase in car travel not only puts a strain on the environment but the increasing congestions also have a negative effect on the accessibility of urban regions. The promotion of alternative transportation modes is a way to deal with these problems. Cycling is considered a sustainable alternative to motorized traffic for short trips. There are several projects to stimulate a behavioural change away from car travel towards bicycle usage. The Dutch government aims at an integrated approach and states that a cycling infrastructure of sufficient quality is necessary for such projects to succeed. Over the past years a new group of cyclists has emerged called the e-bikers, making use of a vastly growing number of e-bikes. This type of bicycle is equipped with a small electric motor that assists while pedalling, this makes it easier to travel longer distances, increases the average trip radius by an increased speed, making the rider feel less tired. This makes an e-bike a suitable mode of transportation for medium distance trips that otherwise are made by car or public transportation. The combination of an increasing share of e-bikes with a well-designed cycling network, could increase the number of trips by bicycle instead of car even further, reducing the burden of car travel on the accessibility and environment of a region. While the research in travel behaviour of cyclists has grown manifold over the past years, the rise of e-bike users as a specific group seems to have been overlooked.

The limited amount of research into e-bike usage and lack of knowledge whether this group differs in route choice decisions from traditional cyclists, left it unclear whether e-bike users have different needs for infrastructure. Route choice models are often used to analyse and gain understanding of usage of the existing cycling infrastructure because they can predict travellers' behaviour in certain situations. This study adds to the understanding of route choice decisions of e-bike users by developing a route choice model using GPS data. For comparison a route choice model for bike trips is estimated in order to determine if the route choice behaviour during e-bike trips differs.

Route choice models assume that individuals choose a route to travel from the places they are, the origin, to the places they want to be, their destination. Often there are multiple routes possible between the origin and destination. Therefore, when trying to predict a choice not only the actual routes chosen but also the routes that are not chosen, the alternatives, should be considered. A route is comprised of a chain of links, where each link contains several attributes, such as length, slope, scenery, etc. Most route choice models are based on the utility theory, which assumes that travellers try to maximize the utility and find the optimal combination of attribute values according to their preferences, when choosing amongst alternatives. The alternatives can be routes with completely different links but in actual infrastructure networks they often have some overlapping links, for instance a bridge or a tunnel. The sharing of a link or multiple links between the alternatives may cause correlation when estimating the route choice model.

A Multinomial Logit (MNL) model, which does not account for correlation caused by overlapping links, is estimated as a starting point. The MNL model is expanded by a path size factor that corrects the utility function when overlap occurs, creating a Path Size Logit (PSL) model. Next, to allow the repeated choice within the dataset and consider taste heterogeneity within the sample, a Mixed Logit (ML) is estimated. The final model is a combination of the ML and PSL.

Several relevant attributes that play a role in route choice behaviour during bike trips, and possibly during e-bike trips, have been identified by means of a literature review. These characteristics are classified into four categories: characteristics of the route, the trip, the

traveller and other circumstances. The dataset used to estimate the route choice models consists of GPS traces collected during March 2014 in the Noord Brabant region of the Netherlands. The GPS traces were transformed into trips by using the Trace Annotator software and matched to network data acquired from OpenStreetMap. Alternatives were generated by means of the K-shortest path method. Up to 5 alternatives were generated for each trip and added, along with the actual route to the dataset. The dataset used for the estimation of the models contains the variables 'gender' and 'age' as characteristics of the traveller, 'travel time' and 'distance' as characteristics of the route, and 'weekday', 'peak hours', 'daylight', 'PC work' and 'PC home' as characteristics of the trip, where 'PC work' and 'PC home' indicate if the end location of the trip corresponds with the home or work location of the respondent. Only 'travel time' and 'distance' varied over the alternatives. The other variables entered the model as interaction variables with distance or travel time. Since travel time and distance are very highly correlated, these attributes are separated into two different models. The models including distance have a higher pseudo-rho squared (ranging from 0.11 to 0.48) compared to the models including travel time (ranging from 0.02 to 0.46). The results of the distance models are included and discussed in the report, the results for the travel time models can be found in the appendices. The original dataset included 104,741 entries. After cleaning up the final data set contained 80,700 entries (17,626 trips made by 742 respondents). 732 respondents reported trips with an e-bike and 522 respondents reported trips with a bike. The sum of bikes and e-bikes exceeds the total number of respondents, which is caused by respondents owning both a bike and an e-bike. Comparing the estimated models for the different transportation modes (bike and e-bike) shows that the significant positive estimate for distance is similar for both modes of transportation. This indicates that the probability of a route being chosen increases when distance increases, this contradicts the base assumption of route choice modelling, that people prefer the shortest route. The positive estimate for distance may be caused by the alternative generation algorithm (k-shortest path) in combination with the few route specific variables included, which were identified in the literature study as factors with a significant effect on route choice. Part of the project from which the data was collected is that participants are rewarded for every kilometre they travel by bike or e-bike. This also explains the higher preference for a longer distance routes. The 'Path Size' variable estimate is positive and significant when included in the model, as is expected. The estimate for the standard deviation of distance is positive and significant for both bike and e-bike, meaning that for both bike trips and e-bike trips taste variation for distance exists in the sample. The fact that in the ML models the spread of the random variable distance is significant and the increase in model fit shows that the MNL and the PSL model do not account for taste variation and correlation between repeated choices. Of the socio-demographic variables, 'gender' has no significant effect but 'age' does have a significant effect on utility of a route. The age categories that are of influence are different for the two transportation modes. Several different variables, such as 'weekday' and 'peak hours', have an effect on route choice when respondents make e-bike trips but don't affect route choice decisions when travelling by bike. The other way around, 'daylight' has an effect on the route choice decision of bike trips but not on e-bike trips.

The increasing share of e-bikes and the increasingly younger age group of e-bike users means that, even if in several occasions the same person can be both e-bike and bike user varying over different trips, merging e-bike with bike users should be supported by research. The results of this study indicate that there are indeed differences in route choice behaviour between these two transport modes and that policymakers should consider these differences.

Dutch summary

De toenemende noodzaak om te reizen en het toenemende aantal kilometers afgelegd met gemotoriseerd verkeer zorgt voor een groeiend aantal files en het toenemen van de gemiddelde reistijd. In Nederland veroorzaakt de toename van autoverkeer niet alleen een grotere belasting voor het milieu, maar hebben het aantal files ook een negatief effect op de bereikbaarheid van stedelijke gebieden. Het promoten van alternatieve vervoerswijzen is een manier om deze problemen te verminderen. Voor korte ritten is fietsen een duurzaam alternatief voor gemotoriseerd verkeer. Er zijn verschillende projecten om een gedragsverandering van auto- naar fietsgebruik te stimuleren. De Nederlandse overheid is gericht op een geïntegreerde aanpak en stelt dat een kwalitatief goede fiets infrastructuur nodig is om stimuleringsprojecten te laten slagen. Gedurende de afgelopen jaren is er een nieuwe groep fietsers opgekomen; de e-bike gebruikers. Dit type fiets is uitgerust met een kleine elektrische motor die ondersteuning biedt tijdens het trappen. Dit maakt het gemakkelijker om langere afstanden af te leggen, verhoogt de gemiddelde reisafstand door een hogere snelheid en zorgt ervoor dat de berijder minder vermoeid op zijn bestemming komt. Dit zorgt ervoor dat een e-bike een geschikt vervoersmiddel is voor ritten met een middellange afstand die anders met de auto of openbaar vervoer zouden worden gemaakt. De combinatie van het groeiende aandeel e-bikes en een goed ontworpen fietsnetwerk zou het aantal verplaatsingen per fiets, in plaats van per auto, nog verder kunnen verhogen. Dit zou de belasting van gemotoriseerd verkeer op de omgeving en de toegankelijkheid van een regio kunnen verbeteren. Hoewel het onderzoek naar het reisgedrag van fietsers de laatste jaren is toegenomen lijkt de opkomst van de e-bikes als specifieke groep onderbelicht gebleven.

Het beperkte aantal onderzoeken naar e-bike gebruik en het gebrek aan inzicht of deze groep verschilt van gewone fietsers zorgt ervoor dat niet duidelijk is of e-bike gebruikers andere eisen en behoeften hebben voor het fietsnetwerk. Om het gebruik van infrastructuur te analyseren en te begrijpen worden vaak routekeuze modellen gebruikt. Dit onderzoek draagt bij aan de kennis over routekeuze gedrag van e-bike gebruikers door het ontwikkelen van een routekeuze model met behulp van gps-data. Om de vergelijking te maken, is tevens een routekeuze model ontwikkeld voor traditionele fietsers.

Een routekeuze model gaat ervan uit dat personen een route kiezen om te reizen tussen de hun vertrekpunt en hun bestemming. Een route bestaat uit een reeks segmenten die verschillende kernmerken hebben, zoals lengte, helling, landschap, etc. Vaak zijn er meerdere routes tussen het vertrekpunt en de bestemming mogelijk. Daarom moet er bij het bepalen van een routekeuze model niet alleen rekening worden gehouden met de gekozen route, maar ook met de alternatieve routes die niet zijn gekozen. De alternatieve en gekozen routes kunnen volledig verschillen van elkaar maar kunnen ook één of meerdere segmenten overlap hebben, zoals een brug of een tunnel. De meeste routekeuze modellen zijn gebaseerd op de utiliteitstheorie die veronderstelt dat reizigers de utiliteit proberen te maximaliseren en een afweging maken tussen de verschillende route kenmerken om de optimale combinatie gebaseerd hun persoonlijke voorkeur te vinden. Als de alternatieve routes overeenkomende segmenten hebben zorgt dit voor correlatie tussen de alternatieven tijdens het schatten van het routekeuze model. Als startpunt voor het schatten van een routekeuze model wordt een Multinomial Logit model (MNL) toegepast, deze houdt echter geen rekening met de correlatie veroorzaakt door overlap. Door een ‘Path Size’ variabele toe te voegen aan het MNL-model wordt de utiliteit gecorrigeerd als er overlap plaats vindt, zo ontstaat er een Path Size Logit model (PSL). Doordat in een maand meerdere reizen door personen gerapporteerd worden

ontstaan er herhaalde keuzes in de dataset, daarnaast kan de voorkeur voor bepaalde route kenmerken verschillen per persoon. Om hiermee rekening te houden wordt een Mixed Logit model (ML) toegepast. Het eindmodel is een combinatie van het ML en PSL-model (ML+PSL).
Gedurende de literatuurstudie zijn verschillende kenmerken vastgesteld die van invloed kunnen zijn op route keuzegedrag bij het reizen met fiets of e-bike. Deze kenmerken zijn ingedeeld in vier categorieën: kenmerken van de route, de reis, de reiziger en overige. De dataset die gebruikt is in dit onderzoek bestaat uit gps-coördinaten die verzameld zijn gedurende maart 2014 in de provincie Noord-Brabant, Nederland. Voor de transformatie van gps-data naar reisdata is de TraceAnnotator software gebruikt. De alternatieven zijn gecreëerd door middel van een 'k-shortest path' logaritme, die tot 5 alternatieven per route genereert. De gekozen route en de gegenereerde alternatieven samen vormen de dataset. De dataset die gebruikt is voor het schatten van de modellen bevatten de variabelen: 'geslacht' (gender) en 'leeftijd' (age) als kenmerken van de reiziger, 'reistijd' (travel time) en 'afstand' (distance) als kenmerken van de route en 'werkdag' (weekday), 'spits' (peak hours), 'daglicht' (daylight), 'PC werk'(PC work) en 'PC thuis' (PC home) als kenmerken van de reis. 'PC thuis' en 'PC werk' geven aan of de eindlocatie van de reis overeenkomt met de thuis of werk locatie. Alleen de reistijd en afstand verschillen per route voor een reis, daarom worden de overige variabelen aan de modellen toegevoegd als interactie variabelen, met reistijd of afstand. Omdat reistijd en afstand te veel met elkaar correleren worden deze opgesplitst in twee verschilde sets modellen. Hierbij hebben de modellen met de variabele afstand een hogere pseudo-rho squared (van 0,11 tot 0,48) dan de modellen met reistijd (van 0,02 tot 0,46). De resultaten van de modellen met de variabel 'afstand' zijn opgenomen in het verslag, de resultaten van de modellen met de variabele 'reistijd' staan in de bijlagen.
De oorspronkelijke dataset bevatte 104.741 rijen aan gegevens. Na het schoonmaken van de data bleven er 80.700 rijen aan gegevens over, dit waren 17.626 reizen gemaakt door 742 personen. 732 personen hadden één of meerdere reizen met een e-bike gemaakt en 522 personen hadden één of meerdere reizen met een gewone fiets gemaakt. Het totaalaantal reizen met e-bikes en gewone fietsen is meer dan het totaalaantal personen. Dit betekent dat een aantal personen zowel een gewone fiets als een e-bike bezit en gebruikt.
Een vergelijking van de resultaten van de modellen voor de fiets en de e-bike toont dat 'afstand' een positief significant effect heeft op de keuze voor een route voor zowel de fiets als de e-bike. Dit betekent dat de kans dat een route gekozen wordt toeneemt naarmate de route langer is. Dit gaat tegen de basisveronderstelling van routekeuze modelering in dat mensen het liefst een zo kort mogelijke route reizen. Een verklaring voor de positieve coëfficiënt van 'afstand' is een combinatie van de manier waarop de alternatieven zijn gegenereerd ('k-shortest path') en het gebrek aan kenmerken van de route in het model. Diverse kenmerken van de route zijn volgens de literatuurstudie van invloed op de route keuze, maar zijn niet in dit onderzoek opgenomen door gebrek aan data hierover. Het project waaruit de data komt beloont deelnemers voor iedere kilometer die ze afleggen met de fiets of e-bike. Dit verklaart ook de voorkeur voor routes met langere afstanden. De significant positieve coëfficiënt van de 'Path Size' variabele is zoals verwacht voor alle modellen. De coëfficiënt voor de standard deviatie van de 'afstand' variabele is positief en significant voor zowel fiets als e-bikereizen. Dit betekent dat er voorkeursverschillen bestaan in de populatie voor 'afstand'. Daarnaast is de modelfit van de ML en de ML+PSL-modellen aanzienlijk beter. Dit toont aan dat het MNL en PSL-model niet voldoende zijn om de herhaalde keuze en voorkeursverschillen in de populatie te beschrijven. Van de kenmerken van de reiziger heeft 'geslacht' geen significant effect, maar 'leeftijd' wel. De leeftijdscategorieën die een effect

hebben op de route keuze verschilt tussen fiets en e-bike. De variabelen ‘werkdag’ en ‘spits’ hebben wel effect op de route keuze gedurende e-bikereizen, maar niet gedurende fietsreizen. Andersom geldt, dat ‘daglicht’ wel een effect heeft gedurende fietsreizen, maar niet tijdens reizen met een e-bike.

Het toenemende aantal e-bike gebruikers en de verjonging van deze gebruikersgroep betekent dat, hoewel dezelfde persoon zowel een e-bike gebruiker als een fietsgebruiker kan zijn het samenvoegen van deze tweegebruikersgroepen allen mag als dit wordt ondersteunt door onderzoek. De resultaten van dit onderzoek tonen dus aan dat er wel degelijk verschillen zijn in route keuzegedrag tussen de twee vervoerswijzen waarmee beleidsmakers rekening zouden moeten houden.

Abstract

In the Netherlands the increase in car travel not only puts a strain on the environment but the increase in congestions also has a negative effect on the accessibility of urban regions. The promotion of alternative transportation modes is a way to deal with these problems. Cycling is considered a sustainable alternative to motorized traffic for short trips. Over the past years a new group of cyclists has emerged called the e-bikers, making use of a vastly growing number of e-bikes. While the research in route choice behaviour of cyclists has grown manifold over the past years, the rise of e-bike users as a specific group seems to have been overlooked. This study adds to the understanding of route choice decisions of e-bike users by developing a route choice model using GPS data. For comparison a route choice model for bike users is estimated. The GPS data of 742 self-selected individuals corresponding to 17626 trips in the Noord Brabant region in the Netherlands collected during March 2014 was used to estimate a Multinomial Logit (MNL), Path Size Logit (PSL), Mixed Logit (ML) and Mixed Logit with Path Size Logit combination (ML+PSL), for trips made by bike and by e-bike. For both transportation modes the ML+PSL best described the data. The estimate for distance is significant and positive for both transportation modes. This indicates that the probability of a route being chosen increases when distance increases. This contradicts the base assumption of route choice modelling, and that people prefer the shortest route. A comparison between the two transportation modes indicates that several different variables, such as ‘weekday’ and ‘peak hours’, do have an effect on route choice when respondent make e-bike trips but don’t affect the route choice decisions when travelling by bike. The other way around, daylight has an effect on the route choice decision of bike trips but not on e-bike trip. The results of this study indicate that while the same person can be both an e-bike and bike user, depending on the trip, their route choice behaviour differs over the transport modes.

List of Abbreviations

RP	Revealed Preference
SP	Stated Preference
MNL	Multinomial Logit
PSL	Path Size Logit
ML	Mixed Logit
PS	Path Size
GPS	Global Positioning System
RC	Random Coefficients
GNL	Generalized Nested Logit
CNL	Cross-Nested Logit
CF	Commonality Factor
NL	Nested Logit
MNP	Multinomial Probit
IIA	Independence of irrelevant alternatives

List of Tables

1.0 Introduction

1.1 Problem definition

Travel demand keeps increasing in the Netherlands. In 2016 the motorized vehicle traffic kilometres increased with 3.1% compared to the year before (Rijkswaterstaat, 2017), which in turn leads to an increase in traffic congestion and average travel time. While on balance cars have become cleaner, the increase of motorized traffic caused an increase in the contribution of cars to air pollution (Milieudefensie, 2015). In the Netherlands the increase in car travel not only puts a strain on the environment but the increasing congestions also have a negative effect on the accessibility of urban regions. The promotion of alternative transportation mode is a way to deal with these problems. Cycling is considered a sustainable alternative to motorized traffic for short trips. The advantages of cycling are numerous, the physical activity has a positive effect on both physical and mental health (Heinen, van Wee & Maat, 2010; Rijksoverheid, 2016). Furthermore, when trips made by motorized vehicles are replaced with cycling trips, harmful emissions can be reduced, leading to a lower environmental impact. The substitution of cars by bikes also leads to less traffic congestion and improves the accessibility of regions. There are several projects to stimulate a behavioural change towards bicycle usage. In such projects participants collect points per cycling kilometre or there is a financial incentive (Tertoolen, Ruijs, Vree & Stelling, 2016). The ‘Beter Benutten’ (‘Optimising Use’) programme of the Dutch government aims at an integrated approach and states that a cycling infrastructure of sufficient quality is necessary for such projects to succeed (Rijksoverheid, 2016). To create and maintain sufficient quality of the cycling infrastructure the Dutch government created the ‘Ontwerpwijzer fietsverkeer’ (Design manual for bicycle traffic).

Over the past years a new group of cyclists has emerged called the e-bikers, making use of a vastly growing number of e-bikes. This type of bicycle is equipped with a small electric motor that assists while pedalling, this makes it easier to travel longer distances, increases the average trip radius by an increased speed, and making the rider feel less tired. Furthermore, it lowers the effort cyclists have to deliver to overcome natural obstacles such as elevations and wind (Wachotsch, Kolodziej, Specht, Kohlmeyer, & Petrikowski, 2014). Over the past years the market share of e-bikes has increased up to almost a third of the new bike sales in the Netherlands (BOVAG-RAI vereniging, 2016). Also, an increasingly younger age group of e-bike buyers show that the market is expanding and a further increase in the number of e-bike users is expected (Rabobank, 2017). The advantages of reduced emissions and positive health effects that traditional bicycles offer, also apply to the e-bike. And with the pedal assistance that reduces the needed effort, it is easier to travel longer distances, making e-bikes a suitable mode of transportation for medium distance trips that otherwise are made by car or public transport. Increasing the e-bike mode share by stimulating people to use e-bikes for medium distance travel can further reduce the burden of car travel on the accessibility and environment of a region. The combination of an increasing share of e-bikes with a well-designed cycling network, could increase the number of trips made by bicycle instead of car even further (Goudappel Coffeng, 2011). While the research in route choice behaviour of cyclists has grown manifold over the past years, the rise of e-bike users as a specific group seems to have been overlooked. The Dutch government acknowledges the potential of the e-bike in further increasing the bicycle mode share, but it is not included in the design manual. This is due to the fact that the design manual was last updated in 2006, the mayor rise in market share of e-bikes happened after that (Stichting BOVAG-RAI, 2016), and there is a limited amount of research into the desired infrastructure by e-bikes.

1.2 Research objective and questions

The limited amount of research into e-bike usage and lack of knowledge whether this group differs in route choice decisions from traditional cyclists, left it unclear whether e-bike users have different needs for infrastructure. In order to increase the e-bike share through a well-designed network, it is necessary to clarify which characteristics influence the route choice of e-bike users and whether there is a need for a new infrastructure. Route choice models are often used to analyse and gain understanding of usage of the existing cycling infrastructure, because they can predict travellers' behaviour in certain situations (Bovy & Stern, 1990). The objective of this research is to add to the understanding of the desired infrastructure of e-bikes by developing a route choice model using GPS data.
Comparison made to route choice decisions of traditional cyclists can will determine is travel behaviour differs between bike trips and e-bike trips. The following question will give direction to this research.

What determinants can be identified for route choice decisions of e-bike users?

The question can be broken down into the following sub questions.

1. Which characteristics play a role in route choice for traditional cyclists?
2. Which relevant variables can be identified in route choice for e-bike users?
3. What does a large GPS dataset contribute to the field of route choice models?
4. What is the effect of the identified variables on route choice for bike and e-bike users?
5. To what extent do the effect of route choice determinants differ for e-bike users and traditional cyclists?

1.3 Research design

Route choice models assume that individuals choose a route to travel from the places they are, known as origin, to the places they want to be, their destination. A route is comprised of a chain of links, where each link contains several attributes, such as length, slope, scenery, etc. Often there are multiple routes possible between the origin and destination. Therefore, when trying to predict a choice not only the actual routes chosen but also the routes that are not chosen, the alternatives, should be considered. The alternatives can be routes with completely different links but in actual infrastructure networks they often have some overlapping links, for instance a bridge or a tunnel. The choice set is the list of all possible routes that an individual can choose between their origin and destination. Route choice models can identify the factors, and their effects, that influence the choice of route and can support policymakers with design decisions for the improvement of the network.

Over the past years many studies have been conducted in route choice behaviour of cyclist. Examples of such studies include Landis, Vattikuti and Brannick (1997), Frejinger (2008), Sener, Eluru and Bhat (2009), Broach, Dill and Gliebe (2012) and Menghini, Carrasco, Schüssler and Axhausen (2009). These studies on route choice behaviour of cyclists can be categorized into two categories. First, those that use *stated* preference data and second, those that use *revealed* preference data. For collecting stated preference data respondents are asked to state what their choice would be in a hypothetical situation while revealed preference data are the actual choices made by respondents in the actual world (Train, 2009). Data collection is often easier and less expensive using stated preference methods. It also allows for testing a hypothetical situation, which is not possible with revealed preference methods. However, it is difficult for respondents to visualize the available choices and the actual choices in real-world situations can differ from the stated preferences in surveys (Broach et al., 2012). A study by Aultman-Hall (1996) used revealed preference data by asking respondents to recall chosen routes and draw these on a map. This method had the disadvantage that the recalled routes could differ from the actual routes chosen and it was time consuming which led to relatively small datasets. More recently however, Global Positioning Data (GPS) has been used in bicycle route choice studies (Menghini et al., 2009; Hood, Sall & Charlton, 2011; Usyukov, 2013). This data is automatically collected through GPS devices given to respondents. The advantage is that it reveals the actual route chosen with accuracy of a couple of metres. This relatively low-cost method of data collection made the use of large datasets possible. Collected GPS data however does need to undergo several steps of data annotation before it can be used directly in route choice models. The GPS dataset used in previous studies is relatively small compared to the GPS datasets that have become available due to the growing number of smartphones and location-based applications. This raises the question what the consequences of such large datasets for route choice modelling are. Therefor this sub-question is added to the list of sub questions that are answered in this research.

1.5 Reading guide

Chapter 2 of this thesis consists a literature review to answer the first, second and third sub-questions defined in paragraph 1.2. A review of existing literature on route choice decisions of cyclists and, when available, of e-bikes will answer the first and second sub question. Followed by an overview of data collection methods for route choice modelling, this will provide an answer to the third sub-question. Chapter 2 ends with an overview of the choice set generation methods and available models for route choice decisions, this provides a theoretical justification of the method applied in this thesis. Chapter 3 describes the methodology. First data collection method and second the model. The results of the application of the model to the data are presented in Chapter 4. This chapter starts with the descriptive analysis. The next part consists of the statistical analysis, with the results separated for bike and e-bike. The last section of this chapter consists of the discussion in which the results for the two transportation modes are compared and discussed, this will provide an answer to the fourth and fifth sub-question. Chapter 5 contains the conclusions and recommendations of this thesis, and some discussions for future research.

2.0 Literature review

The purpose of this research is to add to the understanding of the desired infrastructure of e-bikes, this is done by developing a route choice model and comparing the results to those of traditional bicycles. Therefore, this chapter includes the current state of the field regarding route choice of both traditional bicycles and e-bikes. The first part of this chapter focuses on the influential factors found by previous bicycle route choice studies, for both traditional bicycles and e-bikes were available. Thereafter, the methods of data collection, including GPS data and the different models available are discussed.

2.1 Factors that influence route choice behaviour

Bovy & Stern (1990) describe four categories for factors influencing route choice behaviour, namely characteristics of the route, the trip, the traveller and other circumstances.

2.1.1 Characteristics of the Route

The characteristics of the route are subdivided into three main classes, attributes of the route and its associated roads, attributes of the traffic encountered along the route, and attributes of the road environment.

2.1.1.1 Attributes of the route and road

The base assumption for all route choice models is: people prefer the shortest route in order to shorten the time spent on the travel. This applies for cyclists as well. The length of a route can be measured by distance as well as travel time. Several studies found that distance and/or travel time were indeed important determinants in route choice (Bradley & Bovy, 1984; Broach et al., 2012; Sener et al., 2009; Usyukov, 2013). In addition to travel time and distance, cyclists take also into account the reduction of the physical effort. The use of an e-bike instead of a traditional bike reduces the physical effort necessary (Cherry & Cervero, 2007; Dill & Rose, 2012; MacArthur, Dill, & Person, 2014). However Allemann & Raubal (2015) found that minimizing distance of the route still was the most important factor, but *"slightly less important than for bikers".*

The presence and type of cycling facility is a factor often included in research. There are three categories of cycling facility that are of influence (Hunt & Abraham, 2007): 'Mixed traffic', no separation of cyclists and other traffic, 'bike lane', designated cycling lane on a shared roadway and 'bike path', a separate facility exclusively for cyclists and other non-motorized traffic. Bradley & Bovy (1984) found cycling facility to be only of minor influence, however Shafizadeh & Niemeier (1997) found that "*some respondents would rather bicycle longer distances on a separate bicycle path than shorter distances on the street with some vehicular traffic"*. More recent studies into the trade-off between shortest route and bicycle facility have also found that cyclists are willing to increase trip length and travel time for the presence of a bicycle facility (Hunt & Abraham, 2007; Misra & Watkins, 2017). This is linked with the continuity of the cycling facilities, which is often measured as the percentage of the route where a bicycle facility is available. Cyclists prefer routes with continuous cycling facilities over routes that are interrupted (Stinson & Bhat, 2003) and are willing to bicycle 12 minutes more if the bicycle facility is continuous (Sener et al., 2009).

2.1.1.2 Attributes of the traffic

Previous studies often include attributes of obstacles, such as number of stop signs, traffic lights and left turns (Ben-Akiva & Bierlaire, 1999). However, conflicting results on the effect of these obstacles are found. Several studies found the number of stop signs, traffic lights and left turns to have a significant but limited influence (Stinson & Bhat, 2003; Ton, Cats, Duives, & Hoogendoorn, 2017). Sener, Eluru & Bhat (2009) found that obstacles strongly influence the likelihood of using a route. Broach, Dill & Gliebe (2012) acknowledge that cyclist generally avoid stop signs and traffic lights but also state that these signals can also facilitate a left turn against traffic. Allemann & Raubal (2015) found that the number of traffic lights per ridden kilometre of the routes chosen by e-bikers was higher compared to the routes of cyclists (Broach et al., 2012).
Cyclists seem to be sensitive to a high volume and speed of mixed traffic, this is however correlated with other attributes. The presence of a separate bike lane seems to mitigate the negative effect of other traffic. And when the traffic volumes are high the effect of traffic lights improve. This could be explained by the increase in perceived safety. The perceived safety is however influenced by a number of other attributes, such as the actual safety, level of experience of the rider and gender (Huisman & Hengeveld, 2014; Hunt & Abraham, 2007). Buehler & Pucher (2012) found that the actual safety, measured as the fatality rate per 10.000 bicyclists, is an important determinant of cycling levels. Whether it has a direct effect on route choice is unclear.

2.1.1.3 Attributes of the environment

Previous literature contradicts each other on whether the slope of a route is of influence on the route choice. Several SP studies state that a moderate slope, both uphill and downhill, are preferred over flat terrain (Sener et al., 2009; Stinson & Bhat, 2003). The results of a GPS data study by Menghini, Carrasco, Schüssler, & Axhausen (2009) agree that slope of the route had hardly any impact on route choice. However the location of their study was “hilly” Zürich, Switzerland, and they state that “*this would need to be tested again in a city where the hill side could be detoured around”.* Broach et al. (2012) found, in a RP study using GPS data from Portland, Oregon US, that slopes with a gradient of 2% or more had a large negative effect on the likelihood of a route being chosen. The difference in findings may be caused by the method of data collection and if trip purpose is included in the research or not. Recreational cyclist may like the exercise that comes with cycling uphill. And, in a stated preference study respondents may say they prefer some hills in their route, but when riding, they avoid them. An e-bike assists a rider when riding uphill, this would suggest that slope has little influence on route choice for e-bike users. This is indeed the case when riding uphill, however when riding downhill e-bike users found avoiding steep road segments the second most important factor when choosing their route. This may be caused by the higher average weight of an e-bike and a larger need of control (Allemann & Raubal, 2015).
Scenery or land use is not often included in research into cyclist route choice behaviour, because this attribute may be subjective and is difficult to measure directly. The few stated preference studies that did include it found a majority of their respondents prefer routes with aesthetically pleasing scenery (van Overdijk, 2016; Winters, Davidson, Kao, & Teschke, 2011). However a revealed preference study by Milakis & Athanasopoulos (2014) contradicts this and explain the difference in results that their research studied a *“metropolitan cycling network that is to serve utilitarian rather than recreational.”*

2.1.2 Characteristics of the Trip

Route choice behaviour is significantly influenced by the purpose of the trip, but this is highly correlated with all other attributes. Commuting cyclists tend to have different route preferences than recreational cyclists due to their different value of time (Ben-Akiva & Bierlaire, 1999). Often trip purpose is used to make a distinction of cyclist into separate groups due to their large variety of preferences and several studies only estimated route choice models for one of these groups (Stinson & Bhat, 2003; Aultman-Hall, 1996). When time of day of the trip is included in route choice studies, it is only used for the descriptive analysis and not as a factor influencing the choice behaviour. Time of day may have an effect on the perceived safety of a route (Axhausen, Schönfelder, Wolf, Oliveira, & Samaga, 2003). While this may be correlated such as street lighting along the route, at night cyclists may trade in the shortest path for a better lit route, several studies recommend to include it into the route choice model (Misra & Watkins, 2017; Reddy et al., 2010).

2.1.3 Characteristics of the traveller

Route choice models predict a decision, therefore it is important to include characteristics of the one that makes the decision, in this case the traveller. Gender and age are the most important attributes to include in a route choice study, often differences in route preferences were found between men and women, and amongst different age categories. For example Dill & Gliebe (2008) found that the average speed of women is lower and females tend to avoid hills more than men. With an increase of age, the number of cycling trips and distance decline, however this effect is mitigated with e-bikes (Fietsberaad, 2013).

2.1.4 Other Circumstances

The weather is an important factor in the decision whether to cycle or not, 'good' weather leads to an increase in the amount of cyclists (Dill & Gliebe, 2008; Romanillos, Zaltz Austwick, Ettema, & De Kruijf, 2016a). Whether cyclists change their route depending on the weather is not clear. Rain may cause the road to become wet and slippery and thus affect the safety or windy conditions may increase, or decrease, the effort necessary to deliver when cycling (Weber, Scaramuzza, & Schmitt, 2014).

Research on route choice behaviour of cyclist is extensive and so are the factors that have been identified as having an influence on bicycle route choice behaviour in previous studies. The main factors are discussed in the previous part of this chapter an overview of these factors and their reference studies is shown in Table 1.

Table 1- Overview factors influencing route choice behaviour of cyclists

	Factor	*References*
Trip	Trip purpose	(Aultman-Hall, 1996; Ben-Akiva & Bierlaire, 1999; Stinson & Bhat, 2003)
	Time of day and/or daylight and/or peak hours	(Dill & Gliebe, 2008; Li, Muresan, & Fu, 2017; Ton, Cats, et al., 2017; Winters et al., 2011)
Route *Road*	Cycling facility	(Aultman-Hall, 1996; Bradley & Bovy, 1984; Casello & Usyukov, 2014; Hood et al., 2011; Hunt & Abraham, 2007; Landis et al., 1997; Li et al., 2017; Menghini et al., 2009; Misra & Watkins, 2017; Sener et al., 2009; Shafizadeh & Niemeier, 1997; Stinson & Bhat, 2003)
	Travel time and/or distance	(Allemann & Raubal, 2015; Bradley & Bovy, 1984; Broach et al., 2012; Hunt & Abraham, 2007; Li, Muresan, & Fu, 2017; Menghini et al., 2009; Misra & Watkins, 2017; Sener et al., 2009; Ton, Cats, et al., 2017; Usyukov, 2013)
Route *Traffic*	Obstacles (number of left turns, Stop signs and/or traffic lights)	(Allemann & Raubal, 2015; Ben-Akiva & Bierlaire, 1999; Bierlaire, Chen, & Newman, 2013; Broach et al., 2012; Menghini et al., 2009; Sener et al., 2009; Stinson & Bhat, 2003; Ton, Cats, et al., 2017)
	Volume	(Bradley & Bovy, 1984; Hunt & Abraham, 2007; Landis et al., 1997; Li et al., 2017)
	Safety (perceived and/ or actual)	(Broach et al., 2012; Buehler & Pucher, 2012; Casello & Usyukov, 2014; Huisman & Hengeveld, 2014; Hunt & Abraham, 2007)
	Street lights	(Misra & Watkins, 2017)
Route *Environment*	Slope (uphill and/ or downhill)	(Broach et al., 2012; Casello & Usyukov, 2014; Dill & Gliebe, 2008; Hood et al., 2011; Hunt & Abraham, 2007; Landis et al., 1997; Menghini et al., 2009; Sener et al., 2009; Stinson & Bhat, 2003; Zimmermann et al., 2017)
	scenery	(Landis et al., 1997; Milakis & Athanasopoulos, 2014; van Overdijk, 2016; Winters et al., 2011)
Traveller	Gender	(Aultman-Hall, 1996; Dill & Gliebe, 2008; Heinen, Maat, & van Wee, 2013; Stinson & Bhat, 2003)
	Age	(Aultman-Hall, 1996; Hunt & Abraham, 2007; Stinson & Bhat, 2003)
Other circumstances	Weather	(Dill & Gliebe, 2008; Romanillos, Zaltz Austwick, Ettema, & De Kruijf, 2016b; Stinson & Bhat, 2003; Weber et al., 2014)

2.2 Data collection methods

The methods of data collection in route choice for cyclist vary greatly amongst the different studies. The data used is either stated preference data or revealed preference data.

2.2.1 Stated Preference data.

For collecting stated preference data respondents are asked what their choice would be in a hypothetical situation. Stated preference data in route choice behaviour is either rank, rate or choice data. For the collection of rank data respondents are asked to rank the given alternatives or just rank the importance of the attributes. Rank data provides a preference order of the alternatives or attributes but since it is ordinal scaled data, no quantity of how much one alternative is preferred over the other, rank data cannot be used to estimate a discrete choice model and predict a decision in alternatives that are not included in to study. Rate data is similar to rank data, but respondents are asked to quantify the difference between their first, second and third choice by scoring them on a pre-determined point scale. This results in interval scaled data which first has to be transformed in order to estimate a choice model (Hensher, 1994). For the collection of choice data respondents are asked to choose one of the alternatives. In this case the data is binary and can be directly used in the estimation of route choice models. For SP data collection a survey is set up which can contain both actual and hypothetical scenarios. The alternative scenarios are described by their attributes and it is therefore necessary to know beforehand what attributes may influence the decision and are measured at what level. The description of the attributes needs to be as specific as possible. For example, different respondents might have a different perception of 'short' and 'long', therefore an attribute such as distance should best be described with quantitative levels, such as metres or kilometres. It is possible to include qualitative variables which cannot be quantified, in such cases a respondents' perception of an attribute is measured. However, the perception can be influenced by other variables that are not included in the study, this should be considered when interpreting the results. The total number of possible alternatives, is every possible combination of attribute values. This might represent real life situations in which a respondent chooses from all possible alternatives, but even with a small number of variables the choice set will become too large to include in a survey. Reducing the number of alternatives can be done per respondent, a respondent chooses from a randomly sampled number of alternatives but aggregating all respondents' choice sets still provides the full list of alternatives, this increases the number of respondents necessary to estimate the model. Reducing the number of alternatives can also be done by including only the feasible alternatives or by using unlabelled alternatives, which are only described by the value of their attributes and where the combination of values can describe multiple scenarios. A drawback of stated preference data is that it difficult for respondents to make a clear distinction between alternatives and visualize them when the choice set is very large. Another disadvantage of SP data collection is that the process of stating might influence the data, a respondent states one thing but does another. Personal constraints, such as available money or mobility problems, might not be taken into consideration, these can be included in the model as attributes, but not including them could lead to over evaluation of other attributes.

2.2.2 Revealed Preference data

Revealed preference data is collected by observing the actual choices made by respondents in the real world instead of stating their preferences in hypothetical situations, as is with SP data the case (Train, 2009). The use of RP data mitigates the over valuation of attributes that occurs with SP data collection. Traditionally, RP data was collected by asking respondents to trace their routes previously taken on a map. This led to a dataset containing actual route data but also came with some disadvantages. Recalled routes could differ from the actual routes chosen and, alternative routes and their characteristics that are considered during the choice are not known. The process was not only time consuming for the respondent which had to recall all routes taken but also for the researcher, who had to manually enter each route into GIS. The very large dataset necessary to estimate the choice model made this method time consuming and costly.

The emergence of GPS technology made large scale RP data collection possible at a low cost. "Global positioning system (GPS) technology was originally developed in the 1970's for military purposes and even though it has been available for civil purposes since the 1983 it was not until the 1990's that it became widespread in its integration into consumer devices" (Romanillos et al., 2016a). Initially GPS technology was limited to navigation systems in vehicles but as the GPS technology improved, devices got smaller, cheaper, more accurate and GPS became a standard feature in smartphones. The last decade the volume of GPS data rose substantially due to the growing number of smartphones and location based applications, which collect the data automatically and thus requires little effort from the respondent (van de Coevering, De Kruijf, & Bussche, 2014). One of the earliest studies into the behaviour of cyclists using GPS data was conducted by Harvey & Krizek (2007). They handed out specialised GPS devices to 51 respondents living in Minneapolis and collected data over a period of 3 weeks, resulting in a dataset of 938 trips. They conclude that “it seems that larger studies would be quite feasible using similar technology”. A study of cyclist behaviour in Zurich by Menghini et al. (2009) proved this right when they created a route choice model for cyclists using a GPS dataset that was created by giving GPS devices to 2435 respondents over a period of 6 days. From this dataset 11.000 trips were identified and used to estimate the route choice model.

Not only has the growing volume of GPS usage made it an attractive data source for researchers but also the high level of detail that it provides is influential. The current state of the technology makes it possible to collect data at an individual level every few seconds and with an accuracy of a couple of metres. Where before planning, design and urban theory focussed on the effect of large scale changes such as new highway building or neighbourhood redevelopment, researchers are now able to assess movement of individuals or entities such as cars and use of local infrastructure, at the level of bike lanes or bicycle parking (Batty, 2012).

The use of GPS data in scientific research also has some disadvantages. Where traditional data collection for scientific purposes must be authentic and validated, GPS data collected through smartphone and location based apps is often done by commercial companies that do not hold the same standards or they are not collected for specific research purposes. (Liu, Li, Li, & Wu, 2016). Furthermore, algorithm dynamics, changes made in the sampling and processing algorithms by commercial companies to improve their services, can result in biased or wrong conclusions, if researchers are not notified of these changes (Lazer, Kennedy, King, & Vespignani, 2014). Another downside to data collection by commercial companies is that they must comply with privacy rules and regulations, datasets therefore lack sociodemographic variables. But even without the sociodemographic data of users, privacy remains a major issue, 95% of the users can be identified by just four spatial-temporal position records. It is therefore necessary that while the data may be open and freely accessible, researchers are aware of the sensitive content of the data, replace userIDs with pseudo codes and minimize the risk of data leakage by secure storage (Liu et al., 2016; Romanillos et al., 2016a).

GPS data collection is not immune to the sample bias problem. This method of data collection only contains respondents with a smartphone that use (cycling) apps and voluntary upload their routes. This leads to a self-selective sample of cycling enthusiasts.

Using GPS data has several implications for the application of a route choice model. A large dataset obtained through GPS data collection contains random variations and noise that occurs during data collection, in combination with the fine-grained level of detail, can cause overfitting when estimating the model (Liu et al., 2016). Checking if overfitting occurs can be done by splitting the dataset into a set used for the estimation of the model (80% of all observations) and a test set (20% of all observations) used to validate the model (Ton, Duives, Cats, & Hoogendoorn, 2017; Zimmermann et al., 2017).

Raw GPS data is comprised of coordinates and needs to undergo several steps of data annotation before it can be used for the estimation of route choice models. The GPS coordinates that are collected need to be sorted by route and matched to link on the transportation network. By assuming a certain amount of wait time between two data points, trips are distinguished and points that are not located in the buffer of a link in the network

should be excluded as these can be considered as measurement errors (Menghini et al., 2009). This process can be automated by use of map-matching algorithms, however Harvey & Krizek (2007) found a manual inspection of the data necessary to exclude errors. When the size of the dataset increases this is not feasible anymore. Over the past decade the map-matching algorithms have improved making the manual check redundant. When selecting a map-matching algorithm their strengths and weaknesses must be considered. Some algorithms perform better in urban areas than others and the trade-off between quality and computation time should be made (Quddus, Ochieng, & Noland, 2007). Map matching the coordinates can be done to a transportation network which also includes values on other attributes, such as number stop signs, or type of facility. It is, similar to SP, necessary to know beforehand which attributes might influence the decision and are extracted from the transportation network data.
The disadvantage of only knowing the actual choice but not the alternatives from which it was selected also applies to GPS data, the next section describes how handle this disadvantage.

2.3 Choice set and models

When estimating a route choice model with RP data, it is necessary to know the chosen route but also the alternative routes that are not chosen. The group of all possible routes between a given origin and a destination from which the traveller will make his choice is called the Choice set (Bovy & Stern, 1990). Not all possible routes are routes that are considered, for example routes with a very large detour. Computing and including routes that a traveller would not consider in the choice set is time consuming and unnecessary. Algorithms used to generate the choice set should include all and only feasible choices. To evaluate the performance of an algorithm the following measures are used: The computational time, number of routes in the choice set and the coverage.

2.3.1 Choice set generation

A large group of alternative generation methods are based on the shortest path algorithm proposed by Dijkstra (1959). The most straight forward approach is the K-shortest path algorithm. This algorithm calculates the shortest path, then the second shortest path until the desired number, k, of shortest paths is reached. The shortest path approach has the problem that it assumes perfect knowledge of the network and the shortest path between origin and destination, this can be problematic for large and complex networks. Azevedo et al. (1993) use Dijkstra's algorithm computing for the shortest path, then remove all or some links on this shortest path and calculate the new shortest path in the network. This approach, known as link elimination, has the downside that it generates similar alternatives with little variation, resulting in a low quality choice set (Frejinger, 2008). la Barra, Pérez, & Añez (1993) propose an alternative approach called link penalty. Instead of eliminating links, the links in the shortest path are penalized by increasing the generalized cost on these links and then the shortest path is calculated again. This approach allows essential links in the network, such as bridges and railroad crossings, to be used in several alternatives. A downside to this approach is that the same path can be calculated several times and the algorithm performs very poorly in terms of the computational time and coverage (Bekhor, Ben-Akiva, & Ramming, 2006; C. Prato & Bekhor, 2007). In order to improve the quality of the choice set and prevent the high level of overlap amongst alternatives Zijpp & Fiorenzo Catalano (2005) apply the constrained k-shortest path algorithm. This approach uses the shortest path algorithm in combination with detour and overlap constraints. These constraints also slow down the generation process.

Making it less attractive in terms of computational time. Ben-Akiva, Bergman, Daly & Ramaswamy (1984) proposed the labelling approach that is based on the maximization of the route utility based on certain labels such as fastest, shortest or most scenic path. This can generate alternatives with high variability, but the same path can also be generated for different labels risking a lack of spatial variability. The coverage of this method can be relatively high, but this strongly depends on the number of labels included, more labels means higher coverage but also more time necessary to calculate the alternatives (Ramming, 2002). The branch and bound approach is not based on the shortest path algorithm but constructs a tree of all possible paths from origin to destination. With the inclusion of similarity and temporal constraints this algorithm generates alternatives with high variability (C. Prato & Bekhor, 2006). The coverage with this approach is very high, however the large number of possible alternatives, which is also responsible for the high coverage, makes this method computationally inefficient (Fiorenzo-Calatano, 2007).

There are also numerous approaches using stochastic methods for generating alternatives. Repeatedly are the link attribute cost and individual preferences drawn randomly from a probability function and the shortest path in the network is calculated. Ramming (2002) applied the simulation approach using the Monte Carlo technique to make 48 draws from a normal distribution. The advantage of this method is the large number of alternatives with a high coverage and a relatively low computation time. However, the variability and the number of unique paths depends on the standard deviation of the distribution. Little variance results in few unique generated paths (C. Prato & Bekhor, 2007). The doubly stochastic approach proposed by Bovy & Fiorenzo-Catalano (2007) uses variance in both attributes and parameters found in literature, to reflect the different knowledge of the network and the variation in perception and preference of the different attributes by individual travellers. They conclude that the algorithm is computationally very efficient, has a high coverage and outperforms singly stochastic approaches, but made no comparison to deterministic algorithms. A disadvantage of this approach is that the calibration of the probability function prior to the alternative generation, is vulnerable to errors due to difficulty in collecting from consideration sets from previous studies (Hood et al., 2011). A downside of using generation algorithms is that they might not include the chosen route in the choice set for this reason Cats et al. (2017) propose an empirical approach, only including the chosen routes in the choice set. However, this resulted in a very low model fit.

The different approaches for the generation of alternatives all have their advantages and disadvantages. Determining which of these approaches to use depends on the goal of the research and available resources. After the process of generating the alternatives and combining the alternatives with the actual routes chosen, the choice set is ready to be used in the estimation of a route choice model.

2.3.2 Models

There is a wide variety of models used in route choice research. Most models in travel behaviour studies are based on the utility theory, which assumes that travellers try to maximize the utility and find the optimal combination of attribute values according to their preferences, when choosing amongst alternatives. In choice set C_n of individual n the utility U_{in} of alternative i is given by:

$$U_{in} = V_{in} + \varepsilon_{in}$$

The deterministic term V_{in} consists of individual characteristics and alternative attributes. The random error term ε_{in} is incorporated to account for uncertainty caused by unobserved individual characteristics, unobserved attributes or measurement errors. To estimate the probability that a certain alternative is chosen several different models have been proposed in literature. They can be grouped by the way they deal with the overlapping problem, which is created when alternative routes share a link or multiple links. The first and most basic group of models are those that do not account for overlap amongst alternatives, for example the Multinomial Logit (MNL) or Nested Logit (NL). Second is the group of models that use a tree structure and account for overlap through the random error component of the utility function, examples are Cross Nested Logit (CNL) and the Generalized Nested Logit (GNL). The third group of models allow overlap amongst alternatives by adding an extra attribute which accounts for overlapping paths, to the deterministic part of the utility function of the model. The often-used Path Size Logit (PSL) and the C-logit model are examples of this. The Fourth group contains the models that account for overlap by allowing covariance between the error terms of the alternatives, examples of such models are the Multinomial Probit model (MNP) and Mixed Logit model (ML). Table 2 shows an overview of the described models and some of their variations.

The MNL is based on a linear predictor function of the deterministic term, $V_{in} = \beta x_{in}$, were β is the coefficient to be estimated and x_{in} is the vector of attributes. Assuming that travellers will choose the alternative with the highest utility the probability that traveller n will choose alternative i from the choice set $C_n$is given by the following probability function:

$$P(i|C_n) = \frac{e^{V_{in}}}{\sum_{j\in C_n} e^{V_{jn}}}$$

The MNL model is restricted by the assumption of independence of irrelevant alternatives (IIA), which implies that adding another route to the choice set should not affect the relative odds between the two routes considered. This is not the case when a route overlaps with other routes and could lead to utility overestimation of alternative routes that share links (Fiorenzo-Calatano, 2007). The Nested Logit (NL) model somewhat relaxes the IIA by the creation of a tree structure consisting of nests based on overlapping alternatives. Within these nests correlation amongst the alternatives is allowed. A limitation is that no correlation between the different nests is allowed and each alternative can only belong to one of the nests. This forms a problem when dealing with large real networks in which a route shares links with many alternatives and cannot be assigned to one nest only. The Cross Nested Logit (CNL) has the same tree structure containing nests but allows overlap between the nests by adding a nesting parameter. The nesting parameter represents a degree of overlap between the nests. The Generalized Nested Logit model (GNL) is a variant of the CNL but also allows the nesting parameter to vary for each nest. While both models overcome the problem of correlation amongst nests they both create an extremely large and complex model structure when applied to real networks. The C-logit model, proposed by Cascetta, Russo, Viola, & Vitetta (2002), is an expansion of the basic MNL model with a commonality factor. The commonality factor is a degree of overlap of a route with all other routes in the choice set which penalizes the utility of a route when overlap occurs to prevent the overestimation. There are three different formulations for the commonality factor, but no guidance on which one to use in what case. The Path Size Logit model (PSL) proposed by Ben-Akiva & Bierlaire (1999) is very often used in route choice modelling due to its easiness to calculate and low computational effort. The model accounts for the overlap in routes in a similar way as the C-logit model. The attribute accounting for the overlap amongst the alternatives is called the

Path Size (PS) attribute and is directly added to the utility function of each route. Several variations for this model a proposed but comparison of the models show similar results. A disadvantage of this model is that the PS attribute can only account for part of the correlation (C. G. Prato, 2009). The Multinomial Probit (MNP) model assumes the error terms in the utility function are normally distributed and are allowed to be correlated with each other, and therefore does not suffer from the IIA assumption causing the overlap problem. However, the MNP model does not have a closed form and therefor requires simulation, this is computationally infeasible for large networks with many alternatives. The Mixed Logit (ML) model has characteristics of both the MNL model and the MNP model. Where in the MNP model assumptions are made over the covariance matrix of the error term the ML model has an extra error term added to the utility function that is the source of heteroscedasticity. This way the model contains a Normal and an Extreme Value distributed error term. The advantage of this model over the other models is that it allows alternatives to be correlated, it does not suffer from the overlap problem, and the simulation required is simpler than that of MNP, making it computational feasible for route choice decisions. A Random Coefficients (RC) model is a form of ML in where coefficients can be randomly distributed, this allows taste variation across travellers and correlations between by repeated choice.

Table 2 - Overview models

Model name	***Function***	*Proposed by*
MNL	$U_{in} = V_{in} + \varepsilon_{in}$ $P(i\|C_n) = \frac{e^{V_{in}}}{\sum_{j\in C_n} e^{V_{jn}}}$	
Cross-Nested Logit (CNL)	$U_{in} = V_{in} + \varepsilon_{in} + V_{C_{mn}} + \varepsilon_{C_{mn}} + ln\alpha_{im}$ $P(i\|C_n) = \sum_{m=1}^{M} P(C_{mn}\|C_n)P_n(i\|C_{mn})$ $P(i\|C_{mn}) = \frac{\alpha_{im}e^{V_{in}}}{\sum_{j\in C_{mn}} \alpha_{jm}e^{V_{jn}}}$ $P(C_{mn}\|C_n) = \frac{e^{V_{C_{mn}}+\mu_m I_{C_{mn}}}}{\sum_{l=1}^{M} e^{V_{C_{ln}}+\mu_m I_{C_{ln}}}}$ $I_{C_{mn}} = ln \sum_{j\in C_{mn}} (\alpha_{mj}e^{V_{jn}})^{1/\mu_m}$ $\alpha_{ai} = \frac{l_a}{L_i}\delta_{ai}$	Vovsha (1997)
Generalized Nested Logit (GNL)	$U_{in} = V_{in} + \varepsilon_{in} + V_{C_{mn}} + \varepsilon_{C_{mn}} + ln\alpha_{im}$ $P(i\|C_n) = \sum_{m=1}^{M} P(C_{mn}\|C_n)P_n(i\|C_{mn})$ $P(i\|C_{mn}) = \frac{\alpha_{im}e^{V_{in}}}{\sum_{j\in C_{mn}} \alpha_{jm}e^{V_{jn}}}$ $P(C_{mn}\|C_n) = \frac{e^{V_{C_{mn}}+\mu_m I_{C_{mn}}}}{\sum_{l=1}^{M} e^{V_{C_{ln}}+\mu_m I_{C_{ln}}}}$	(Wen & Koppelman, 2001)

	$\alpha_{ai} = \frac{l_a}{L_i}\delta_{ai}$ $\mu_m = 1 - \frac{\sum_{j\epsilon C_n} a_{mi}}{\sum_{j\epsilon C_n} \delta_{mi}}$	
C-Logit with Commonality Factor (CF)	$U_{in} = V_{in} + \beta_{CF} CF_{in} + \varepsilon_{in}$ $CF_{\boldsymbol{in}} = -\beta_0 ln \sum_{l\epsilon C_n} \left(\frac{L_{ij}}{\sqrt{L_i L_j}}\right)^y$ $CF_{\boldsymbol{in}} = -\beta_0 ln \sum_{a\in\Gamma_i} \left(\frac{l_a}{L_i} \sum_{l\epsilon C_n} \delta_{aj}\right)$ $CF_{\boldsymbol{in}} = -\beta_0 \sum_{a\in\Gamma_i} \left(\frac{l_a}{L_i} ln \sum_{l\epsilon C_n} \delta_{aj}\right)$ $CF_{\boldsymbol{in}} = ln\left[1 + \sum_{\substack{j\in C_n \\ j\neq i}} \left(\frac{L_{ij}}{\sqrt{L_i L_j}}\right)\left(\frac{L_i - L_{ij}}{L_j - L_{ij}}\right)\right]$ $P(i\|C_n) = \frac{e^{V_{in}+CF_{in}}}{\sum_{j\epsilon C_n} e^{V_{jn}+CF_{jn}}}$	(Cascetta et al., 2002)
Path Size Logit (PSL)	$U_{in} = V_{in} + \beta_{PS} lnPS_{in} + \varepsilon_{in}$ $P(i\|C_n) = \frac{e^{V_{in}+lnPS_{in}}}{\sum_{j\epsilon C_n} e^{V_{jn}+lnPS_{jn}}}$ $PS_{in} = \sum_{a\in\Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \delta_{aj}}$	(Moshe Ben-Akiva & Ramming, 1998)
Path Size Logit second formulation (PSL2)	$U_{in} = V_{in} + \beta_{PS} lnPS_{in} + \varepsilon_{in}$ $P(i\|C_n) = \frac{e^{V_{in}+lnPS_{in}}}{\sum_{j\epsilon C_n} e^{V_{jn}+lnPS_{jn}}}$ $PS_{in} = \sum_{a\in\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \frac{L^*_{C_n}}{L_j} \delta_{aj}}$	(Ben-Akiva & Bierlaire, 1999)
Generalized Path Size	$U_{in} = V_{in} + \beta_{PS} lnPS_{in} + \varepsilon_{in}$ $P(i\|C_n) = \frac{e^{V_{in}+lnPS_{in}}}{\sum_{j\epsilon C_n} e^{V_{jn}+lnPS_{jn}}}$ $PS_{in} = \sum_{a\in\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \left(\frac{L_i}{L_j}\right)^{\varphi} \delta_{aj}}$	(Ramming, 2002)
Path Size Correction	$U_{in} = V_{in} + \beta_{PSC} PSC_{in} + \varepsilon_{in}$ $PSC_{in} = \sum_{a\in\Gamma_i} \frac{l_a}{L_i} ln\left(\frac{1}{\sum_{j\epsilon C_n} \delta_{aj}}\right)$	(Bovy, Bekhor, & Prato, 2008)
Mixed Logit (ML)	$U_n = X\beta + \varepsilon = FT\zeta + v$	(Moshe Ben-Akiva & Bolduc, 1996)

	$$Var(\varepsilon) = FTT^{T}F^{T} + \left(\frac{g}{\mu^2}\right)I_{J_n}$$ $$\sum n = Var(FT\zeta) = \sigma^2 \begin{bmatrix} L_1 & L_{12} & \cdots & L_{1J_n} \\ L_{12} & L_2 & \cdots & L_{2J_n} \\ \vdots & \vdots & \ddots & \vdots \\ L_{1J_n} & L_{2J_n} & \cdots & L_{J_n} \end{bmatrix}$$ $$P_n(i) = \int \Lambda\,(i\vert\zeta) \prod_{m=1}^{M_n} \phi(\zeta_m)\, d\zeta_m$$	
Random Coefficients model (RC)	$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni}$$ $$L_{ni}(\beta) = \frac{e^{\beta_n' x_{ni}}}{\sum_j e^{\beta_n' x_{nj}}}$$ $$P_{ni} = \int L_{ni}\,(\beta_{ni}) f(\beta_{ni}\vert\theta) d\beta_{ni}$$	(Moshe Ben-Akiva & Bolduc, 1996)

3.0 Methodology

This chapter describes the application of the method based on the literature study in the previous chapter. First the method of data collection is discussed. Next, the MNL, the PSL and the RC model that are applied, are described. Third is the data analysis, both descriptive and statistical. For the analysis Econometric Software's Nlogit version 5 software was used.

3.1 Data collection

The dataset used to estimate the route choice models consists of GPS traces collected through a smartphone app that is part of the B-riders project. This is the bicycle stimulation program of the Province of Noord Brabant, the Netherlands. The purpose of the program is not only stimulating bicycle usage, by offering rewards to cyclists, but also collecting data for detailed analysis to provide policy insights. The scientific purpose of the data collection has the advantage that the dataset contains the sociodemographic variables 'age' and 'gender', which have been found in previous studies to have an influence. While the B-riders project started in 2013 and is still running, this study only uses the data collected for one month, March 2014. Respondents had to register on the B-riders website and download the smartphone app in order to participate in the reward program of B-riders, they also had to be 18 years or older, have a minimum commute of 4 km to a destination in the Noord Brabant region and make their commute at least 50% of the time by car during the last 3 months. The self-selection of the sample creates a bias with respondents who intend to make more bike and/or e-bike trips and know their way around a computer and a smartphone. This must be taken into account when interpreting the results of the model estimation.

Between 1 March 2004 and 31 March 2004, 901 respondents making a bike and/or e-bike trip with an origin and/or destination in the Noord Brabant region were recorded by the app. The GPS traces, obtained from the app, were transformed into activity-travel diaries by the Trace Annotator software developed by the Urban Planning group of Eindhoven University of Technology. The Trace Annotator is used to recognise the transportation mode and segment the GPS traces into trips. An advantage of the Trace Annotator is the included imputation model, which is trained in the recognition of the transport mode by combining accelerometer data with GPS data, and predicts the bicycle mode with an accuracy 97% and the e-bike mode with 99%. The GPS points were connected sequentially and matched to transportation network data acquired from OpenStreetMap (openstreetmap.org). For this an algorithm was used which first looks at the possible road segments around a GPS point and then identifies the most probable one. After the transportation mode was determined and the chosen route was matched to the transportation network, alternatives were generated and added to the dataset.

Both the map matching of the GPS coordinates and the generation of alternatives was done prior to this study. Detailed information on the process of mode identification and map matching can be found in Feng & Timmermans (2013a) and Feng & Timmermans (2013b), respectively. For the alternative generation the K-shortest path method was used based on Dijkstra's algorithm, up to 5 alternatives were generated for each trip and added, along with the actual route to the dataset. This method resulted in a dataset in which not all trips have the same number of alternatives, in some cases the generation algorithm didn't provide more than 1, 2, 3 or 4 alternatives. The algorithm produced routes that were also actual choices, adding the alternatives along with the actual choices to the dataset resulted in some having a route as both an actual choice and as an alternative. In such cases the generated alternative

was removed from the choice set. As mentioned in the previous chapter, large GPS datasets can contain random variation and noise. Due to the dataset being very large, inspection of this manually was not considered feasible, however that these errors were present in this dataset became apparent when doing the first descriptive analysis. An example of this is a travel time of 1082 minutes or a distance of 111 km, both seem infeasible. Checking the cross tabulations and testing for outliers proofed to be an effective way for detecting and excluding such errors without having to check each entry manually.
The list of variables that are collected for the estimation of the route choice models is based on the literature study. As a start, the sociodemographic variables 'age' and 'gender' are included, these are extracted directly from the B-riders dataset. Previous studies found these attributes to be of influence however they were often excluded from previous studies using GPS data, because they were simply not available. It is also not clear whether these attributes influence route choice decision by travellers using an e-bike and therefore should be included. 'Weekday-weekend', 'peak-off peak' and 'daylight' are attributes selected to be included as characteristics of the trip, these were calculated based on the start and end time of each trip found in the B-riders dataset, compared to the peak times defined in the Publieksrapportage Rijkswegennet (Rijkswaterstaat, 2017) and the times of sunrise and sunset obtained from the database of the Koninklijk Nederlands Meteorologisch Instituut (KNMI, 2014). There are also two attributes that somewhat represent trip purpose. 'PC work' and 'PC home' indicate whether the endpoint of the trip corresponds with the given work or home location in the B-riders dataset. Further indication of trip purpose is not possible due to the method of data collection and privacy reasons. Characteristics of the route are represented by 'travel time' and 'distance'. The B-riders dataset contained the information on these variables for the chosen routes, the travel time and distance for the alternatives were included in the network data extracted from OpenStreetMap. Unfortunately, it was not possible to extract more attributes describing the route characteristics from the data set. Table 3 shows the collected attributes and their number of levels.

Table 3 - Collected attributes

Characteristics of the	*Mode of transport*	*E-bike*	*Bike*
Traveller	**Age**	4 levels	4 levels
	Gender	2 levels	2 levels
Trip	**Weekday-weekend**	2 levels	2 levels
	Peak-off peak	2 levels	2 levels
	Daylight	2 levels	2 levels
	PC work	2 levels	2 levels
	PC home	2 levels	2 levels
Route	**Travel Time**	Continuous	Continuous
	Distance	Continuous	Continuous

Prior to the analysis the available dataset had to be coded. Effect coding was used for this. Table 4 shows a short example of what the dataset looks like. The description of the variables and the data coding can be found in appendix A.

Table 4 - Data example

Respondent	*Trip*	*Alternative*	*Travel time*	*Daylight*	*Peak*
A	1	1	2	-1	1
A	1	2	4	-1	1
A	2	1	3	1	1
A	2	2	2	1	1
A	2	3	4	1	1
B	3	1	4	1	-1
B	3	2	5	1	-1
B	4	1	2	-1	1
B	4	2	3	-1	1
B	5	1	3	1	-1
B	5	2	4	1	-1

3.2 Method

In order to determine what is of influence on route choice decisions during e-bike trips and make a comparison to trips with a traditional bicycle, a route choice model is applied. Since the alternative generation method is based on a k-shortest path algorithm the alternatives in the choice set are unlabelled. This means that they are described by the values of their attributes. When a respondent makes a choice from the alternatives only travel time and distance vary over the alternatives and the rest of the attributes are fixed. The choice set per trip is called 'variable' since the number of alternatives differs per trip. Each respondent has the option to make as many trips as they want, this may cause overestimation of the respondent's characteristics of a person that makes more trips compared to those who only make a few trips. Treating the data as panel data prevents this problem. In this case the panels are called unbalanced due to different numbers of entries per respondent. In short, the dataset consists of unlabelled variable choice set with unbalanced panels.

The first model applied is the MNL model, this provides some insight into which variables may be of influence on both travel time and distance. Travel time and distance are expected to be highly correlated, checking this shows a very high correlation coefficient of 0.95. Therefore, these attributes are separated into two different models. Since the MNL model cannot account for the overlap between alternatives, the MNL is expanded to the PSL model, proposed by Ben-Akiva & Bierlaire (1999), by adding a Path Size (PS) attribute to utility function.

$$U_{in} = V_{in} + \beta_{PS} lnPS_{in} + \varepsilon_{in}$$

And

$$P(i|C_n) = \frac{e^{V_{in}+\beta_{PS} lnPS_{in}}}{\sum_{j\in C_n} e^{V_{jn}+\beta_{PS} lnPS_{jn}}}$$

With

$$PS_{in} = \sum_{a\in\Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j\epsilon C_n} \delta_{aj}}$$

Where Γ_i is the set of links in path i; L_a and L_i are the length of link a and path i. δ_{aj} is the link-path incidence variable that is 1 if link a is on path j and 0 otherwise. A limitation of the MNL and the PSL model in this case is that it cannot handle the panel setup of the data and the taste heterogeneity amongst travellers, this means that each entry is considered as an individual respondent which increases the risk of overestimation. To overcome these problems, a ML model is applied. The size and setup of the dataset in combination with the research question makes the Random Coefficients logit (RC) model suitable for this study. This model allows a panel data setup and makes the examination of which of the explanatory variables cause the preference heterogeneity around the mean of the travel time and distance attribute possible. The utility U of alternative i of respondent n is in the RC model is described by Train (2003) by:

$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni}.$$

Where x_{ni} are the observed variables and β_n represents the taste of respondent n and is given by a vector of coefficients of these variables. ε_{ni} is a random term independent and identically distributed across individuals, choice situations and alternatives. If β_n was observed, the probability of choosing alternative i for respondent n would be:

$$L_{ni}(\beta) = \frac{e^{\beta_n x_{ni}}}{\sum_j e^{\beta_n x_{nj}}}$$

Since the repeated choice in the data has to be taken into account, the conditional probability is given by:

$$L_{ni}(\beta) = \prod_{t=1}^{T} \frac{e^{\beta_n x_{ni}}}{\sum_j e^{\beta_n x_{nj}}}$$

Where T is the number of choices made. However, β_n is unobserved and varies over the population with density $f(\beta_{ni}|\theta)$, where θ are the parameters of the distributions. The distribution can be normal, lognormal, uniform or triangular. The probability of choosing alternative i for respondent n for all possible β_n is given by:

$$P_{ni} = \int L_{ni}\,(\beta_{ni}) f(\beta_{ni}|\theta) d\beta_{ni},$$

With the log-likelihood function:

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} d_{ni}\, lnP_{ni}.$$

Where d_{ni}=1 if n chose i and 0 otherwise. Since the integral does not have a closed form the probability cannot be calculated directly. Simulation is used to approximate the probabilities. During simulation a value for β_{ni} is drawn from $f(\beta_{ni}|\theta)$ and probability is calculated. This step is repeated, and the results are averaged. By repeating this process many times, the variance decreases, and the outcome becomes more accurate. The number of draws that is sufficient to come to a stable result depends on the complexity of the model in terms of number of random parameters, correlation of attributes and on the type draws. The Halton sequence draws outperform the random sequence draws, in terms of minimum number of draws necessary to acquire a stable result and computation time (Bhat, 2001). Since the number of draws not only effects the results but also the time necessary for the estimation, the model is first estimated with 50 Halton draws, once the model provides output the final ML models are estimated with 1000 Halton draws. This number will provide an accurate result (Bhat, 2001; K. Train, 1999). The Nlogit software, used to estimate the model, automatically handles the unbalanced panels caused by respondents not all having the same amount of entries, trips made, in the dataset. This also prevents the other fixed variables being overestimated in the model. In order to assess the performance of the models a comparison between the log-likelihood function of the estimated model and the log-likelihood of the base model, which is a constant only model, is made. This is done by calculating the McFadden pseudo-rho squared given described by:

$$R^2 = \frac{LL_{base\ model} - LL_{estimated\ model}}{LL_{base\ model}}.$$

The greater the pseudo-R^2, which ranges from 0 to 1, the better the model describes the data. A pseudo-R^2 above 0.3 is considered a decent model fit (Hensher, Rose, & Greene, 2005).

4.0 Data analysis and results

The original dataset was very large and included 104,741 entries. After cleaning up during which entries where any of the variables had missing values and outliers were excluded, the final data set contained 80700 entries (17626 trips made by 742 respondents). First a descriptive analysis is conducted, second a statistical analysis using the methods mentioned in chapter 3. Next, the results of different models are compared and discussed.

4.1 Descriptive analysis

The personal characteristics 'age' and 'gender' were included in the dataset. In order to determine whether the sample population is representative, its distribution is compared to the distribution of the overall population of the Netherlands. Table 5 shows the distribution of gender. The comparison of percentages of the sample to that of the overall population in the Netherlands shows that the males are slightly under represented and females overrepresented.

Table 5 – Gender sample-population

Gender	*Sample*		*Overall Population* (CBS, 2017)	
Male	354	48%	8334385	50%
Female	388	52%	8494904	51%
Total	742	100%	16829289	100%

Table 6 shows the age distribution of the sample and the overall Dutch population. A rule of the B-riders project was that participants had to be over the age of 18 and focused on commuters, this causes the age category of <35 years to be very small compared to the overall population. The oldest respondent in the sample was 66 years old this means that age category > 55 years in the sample does not represent the overall population containing people over 66. The rest of the age distribution of the sample is not representative of the overall Dutch population either.

Table 6 - Age sample-population

Age	*Sample*		*Overall Population* (CBS, 2017)	
< 35 years	33	4%	6952004	41%
35-44 years	130	18%	2244003	13%
45-54 years	346	47%	2538745	15%
> 55 years	233	31%	5094537	30%
Total	742	100%	16829289	100%

The number of entries varies per respondent therefore the remaining independent variables are described in combination with the number of trips.

4.1.1 Descriptive analysis of characteristics of the traveller

Since the main topic of this study is the route choice decisions of e-bike users and the comparison of it with decision of traditional bicycle users, it is necessary to make sure sufficient data is available. This is done by combining the frequency data of the transport mode with the other independent variables. Combining the transport mode with the gender of respondents, see Table 7, shows that sum of bikes and e-bikes exceeds the total number of respondents shown in Table 5. This is caused by respondents owning both a bike and an e-bike. For this reason, the independent variables are described in combination with the number of trips made with the mode of transportation.

Table 7 - Mode of transportation per gender

Gender	E-bike	Bike	Total
Male	346	254	600
Female	386	298	684
Total	732	522	1285

Table 8 shows number of trip made per mode of transportation per age category and per gender of a respondent.

Table 8 – Trip frequencies per traveller characteristic and mode

Mode	Age	Male		Female		Total	
Bike	<35 years	23	2%	56	4%	79	3%
	35-44 years	175	16%	338	22%	513	19%
	45-54 years	431	39%	710	46%	1141	43%
	>55years	485	44%	431	28%	916	35%
	Total	1114	100%	1535	100%	2649	100%
E-bike	<35 years	135	2%	352	4%	487	3%
	35-44 years	1198	17%	1215	15%	2413	16%
	45-54 years	3045	43%	4178	53%	7223	48%
	>55years	2760	39%	2094	27%	4854	32%
	Total	7138	100%	7839	100%	14977	100%
Total	<35 years	158	2%	408	4%	566	3%
	35-44 years	1373	17%	1553	17%	2926	17%
	45-54 years	3476	42%	4888	52%	8364	47%
	>55years	3245	39%	2525	27%	5770	33%
	Total	8252	100%	9374	100%	17626	100%

4.1.2 Descriptive analysis of characteristics of the trip

Table 9, Table 10 and Table 11 show the number of trips per mode of transportation in combination with 'weekday', 'peak' and 'daylight' respectively. The number of trips per mode of transportation that end on the given work location and on the given home location are shown in Table 12 and Table 13.

Table 9 - Trip frequencies per trip characteristic and mode - Weekday

Weekday	Bike		E-bike		Total	
Weekend	721	27%	1306	9%	2027	12 %
Weekday	1928	73%	13671	91%	15599	88 %
Total	2649	100% (15%)	14977	100% (85%)	17626	100% (100%)

Table 10 - Trip frequencies per trip characteristic and mode - Peak

Peak	Bike		E-bike		Total	
Off peak	1808	68%	6412	43%	8220	47%
Peak	841	32%	8565	57%	9406	53%
Total	2649	100% (15%)	14977	100% (85%)	17626	100% (100%)

Table 11 - Trip frequencies per trip characteristic and mode - Daylight

Daylight	*Bike*		*E-bike*		*Total*	
No daylight	416	16%	1766	12%	2182	12%
Daylight	2233	84%	13211	88%	15444	88%
Total	2649	100% (15%)	14977	100% (85%)	17626	100% (100%)

Table 12 - Trip frequencies per trip characteristic and mode - PC work

PC work	*Bike*		*E-bike*		*Total*	
Work	312	12%	4248	28%	4560	26%
Other	2337	88%	10729	72%	13066	74%
Total	2651	100% (15%)	14981	100% (85%)	17632	100% (100%)

Table 13 - Trip frequencies per trip characteristic and mode - PC home

PC home	*Bike*		*E-bike*		*Total*	
Home	1327	50%	6456	43%	7783	44%
Other	1322	45%	8521	57%	9843	56%
Total	2651	100% (15%)	14981	100% (85%)	17632	100% (100%)

4.1.3 Descriptive analysis of characteristics of the route

The variables travel time and distance are continuous, to be able to compare the two models created by separating these variables, these variables are natural log-transformed entering the model. Table 14 shows the descriptive statistics of the travel time, distance and their transformations per transport mode.

Table 14 - Trip frequencies per route characteristic and mode

	Minimum	*Maximum*	*Mean*	*Std. deviation*
Bike				
Travel time (min)	1.00	109.00	22.61	21.28
Distance (km)	0.08	29.05	5.46	5.04
ln (travel time)	0.00	4.69	2.57	1.16
ln (distance)	-2.50	3.37	1.19	1.10
Path Size	0.28	1.00	0.84	0.14
ln (Path Size)	-1.29	0.00	-0.19	0.19
E-bike				
Travel time (min)	1.00	110.00	36.62	23.18
Distance (km)	0.12	31.34	9.04	5.51
ln (travel time)	0.00	4.70	3.31	0.89
ln (distance)	-2.15	3.44	1.93	0.84
Path Size	0.22	1.00	0.82	0.16
ln (Path Size)	-1.50	0.00	-0.23	0.22

Table 14 also describes the Path Size variable, created by using the equation given in section 3.1. Since a Path Size of 1 means a unique route and of 0 complete overlap, a mean of 0.84 (0.82) and standard deviation of 0.14 (0.15) indicates that the alternative generation algorithm performed rather well in creating different alternatives with only some overlap.

4.2 Statistical analysis

As described earlier, the variable 'travel time' and 'distance' have such a high correlation that they are separated into two models. As a starting point an MNL model is applied for the two variables. First the MNL model is estimated with only the main effects of 'travel time' and 'distance'. Next, since all the other independent variables are fixed over the alternatives in the choice set, they are entered in the model as interaction variables, interacted with 'travel time' or 'distance'. The third model applied is the PSL model including the Path Size variable. Followed by the ML model, which allows the panel data setup. Since the pseudo rho-squared of the MNL models improved considerably by adding the Path Size variable, the next model applied is the ML in combination with the Path Size variable. The results from the Nlogit software for all the models can be found in Appendix B and Appendix C. Table 15 shows an overview of the pseudo rho-squared values calculated using the formula given in paragraph 3.1. The pseudo rho squared values greater than 0.3 are considered 'good' (Hensher et al., 2005). The pseudo rho-squared values for the bike models cannot be compared directly to the pseudo rho-squared values of the e-bike models. This is because the samples used to estimate the models and calculate the pseudo rho-squared values are different. The PSL model performs better than de MNL model for all combinations bike, e-bike, travel time and distance. This shows that the Path Size attribute is not only necessary to account for the overlap between alternatives but also improves the model fit. The ML models, which are estimated considering the repeated choice in the data, have a higher pseudo rho-squared value than the MNL models but lower than the PSL models. The ML model in combination with the PSL model is also estimated using the repeated choice data and performs best for both bike, e-bike, travel time and distance. The models including distance better describe the data than the models including travel time.

Table 15 -Overview R^2 all models

	Bike				*E-bike*			
	MNL	PSL	ML	ML + PSL	MNL	PSL	ML	ML + PSL
Distance								
pseudo rho-squared	0.11	0.31	0.12	0.34	0.03	0.42	0.10	0.48
Travel time								
pseudo rho-squared	0.02	0.20	0.04	0.25	0.03	0.40	0.12	0.46

Table 16 and
Table 17 show the results of the MNL, PSL, ML and the ML + PSL model with interaction effects on distance for respectively bikes and e-bikes. In the next part these results are discussed. The results for the models including travel time can be found in Appendix C.

Table 16 - Results distance models for bike

Attribute	*Bike*			
Main effects	MNL	PSL	ML	ML + PSL
ln(distance)	3.655***	4.082***	3.934***	4.786***
Interaction effects				
Gender	-0.026	-0.098	-0.013	-0.113
Age <35	2.439**	2.547**	2.037*	2.246
Age 35-44	-1.029**	-0.782*	-1.384**	-1.550*
Age 45-55	-0.378	-0.256	-0.709	-0.826
Weekday	0.387**	0.454***	0.260	0.221
Peak	-0.090	-0.214	-0.029	-0.103
Daylight	0.641***	0.398*	0.764***	0.938***
PC work	1.226***	0.855***	0.655*	0.420
PC home	0.184	0.275*	0.086	-0.041
ln(Path Size)		8.102***		10.20***
Std. dev. of random ln(dist) coeff.			2.190***	4.366***
log-likelihood model	-3009	-2339	-2981	-2241
log-likelihood base	-3394	-3394	-3394	-3394
pseudo rho-squared	0.113	0.311	0,122	0.340
Note: ***, **, * ➔ significance at 1%, 5%, 10% level				

Table 17 - Results distance models for e-bike

Attribute	*E-bike*			
Main effects	MNL	PSL	ML	ML + PSL
ln(distance)	3.299***	4.253***	5.009***	4.889***
Interaction effects				
Gender	-0.452***	-0.349***	-0.125	-0.024
Age <35	-0.795*	-0.282	-2.098	-1.605
Age 35-44	-1.161***	-1.311***	-1.964*	-1.731
Age 45-55	-1.297***	-1.701***	-1.700**	-1.931**
Weekday	-0.272**	-0.674***	-0.911***	-0.831***
Peak	-0.135*	-0.379***	-0.301**	-0.484***
Daylight	-0.017	-0.455***	-0.067	-0.171
PC work	0.136	0.035	0.628***	0.608***
PC home	-0.074	0.062	0.230*	0.194
ln(Path Size)		13.08***		15.50***
Std. dev. of random ln(dist) coeff.			8.130***	8.780***
log-likelihood model	-21831	-13059	-20184	-11600
log-likelihood base	-22389	-22389	-22389	-22389
pseudo rho-squared	0.025	0.417	0.098	0.482
Note: ***, **, * ➔ significance at 1%, 5%, 10% level				

For all models ‘Distance’ is the only variable in the model which has a main effect since all independent variables are fixed over the alternatives in the choice set. Its effect is significant in all models however its sign is counterintuitive. The positive sign of the estimated coefficient means that if the distance of a route increases so does the probability the route being chosen, suggesting that people prefer longer routes. This is not in line with findings in existing literature, this contradiction will be discussed in section 4.3.
The Path Size variable estimate is positive and significant at a 1%-level when included in the model, which is as expected. The sign needs to be positive in order to correct the utility for the overlap amongst alternatives and the fact that the estimates have a significant effect proves that there is indeed correlation between the alternatives which have overlap.

Besides distance and Path Size only the age category ’35-44 years’ and the ‘daylight’ estimate are significant in the ML + PSL model for bikes. The age category estimate has a negative sign and is significant at only 10%-level. The ‘daylight’ estimate is highly significant, at 1%-level, and has a positive sign. This means that the likelihood of a longer route being chosen at night is smaller than during the day. Moreover, respondents between 35 and 44 years old prefer a longer distance to a lesser extent than respondents over 55 years old, which is the base category.
In the e-bike ML + PSL model other variables, compared to the bike model, besides distance and Path Size are significant. The estimates for age category ’45-55 years’, ‘weekday’ and ‘peak’ have a negative sign and the estimate for ‘PCwork’ has a positive sign. All are significant at 1%-level, only age category ’45-55 years’ is significant at 5%-level. The negative estimate of ‘peak’ indicates that when respondents are making a trip using an e-bike during peak hours they prefer a longer distance to a lesser extent than when making an e-bike trip outside peak hours. The negative estimate for ‘weekday’ indicates that during weekdays respondents prefer a longer distance to a lesser extent than during the weekend when using an e-bike. The negative estimate for the age category ’45-55 years’ indicates that respondents in this age category prefer a longer distance to a lesser extent during e-bike trips than respondent over 55 years old, which is the base category. The positive sign for the ‘PCwork’ estimate suggests that if the end location of the trip is the work location, when using an e-bike, respondents have a higher preference for longer distance routes than when the end location is not the work location.
All variables, besides distance and Path Size, are interaction variables, interacted with distance. Because the results vary for the different transport modes, the results for the MNL, PSL, ML and ML+PSL model are addressed separately per transport mode, starting with bike and then e-bike.

4.2.1 Results bike models

‘Gender’ has in none of the bike models a significant effect on the preferred distance. This is in line with results from other studies who found no significant effect of gender on preferred distance (Aultman-Hall, 1996; Hood et al., 2011; Sener et al., 2009; Shafizadeh & Niemeier, 1997) but contradicts significant negative effect found by Heinen et al., (2013).
The age category ‘<35 years’ has a positive sign for all bike models but is only significant at 5%- level in the PSL and in the MNL model. In the ML model this effect is significant only at the 10%-level and in the ML + PSL model the estimate is no longer significant. The positive sign for this estimate indicates that people younger than 35 years old have a stronger preference for a route with a longer distance than people in the base category of 55 years and older. The

sign for the age category ’35-44 years’ estimate is in all models negative. This estimate is significant at 5%-level in the MNL and ML + PSL model and significant at 10%-level in the other two models. The negative sign of this estimate indicates that respondents in the between 35 and 44 years old prefer a longer distance to a lesser extent than respondents over 55 years old. The age category ’45-55 years’ has no significant effect on the preferred distance for the bike models. In other studies age is often not included in the statistical analysis but only used to describe the sample. When age is included, the effects on trip length, distance or travel time, is found to be either not significant (Aultman-Hall, 1996; Hood et al., 2011) or positively related to travel time, older respondents have a lower sensitivity to travel time (Shafizadeh & Niemeier, 1997; Stinson & Bhat, 2003).

The variable ‘weekday’ has a significant positive effect on preferred distance in the MNL and PSL model for bike but loses significance in both ML models. This suggests that the panel data setup of the ML model prevents the overestimation of this effect, caused by some respondents having many and other just a few trips reported. During the literature review no other studies including the effect of the day of the week on distance for route choice were found.

The estimates for ‘peak’ are in all models negative, as expected, but not significant. In the bike models the estimates for ‘daylight’ are all positive and significant, in the PSL model only at 10%-level. This positive sign means respondents have a higher preference for a longer distance during the day than during the night. A sensitivity to daylight and darkness for cyclists is also found by Gatersleben & Appleton (2007) and Heinen et al., (2013).

If the end location of the trip corresponds to the work location it only influences the preferred distance in the MNL and PSL model at 1%-level. In the ML model this estimate is only significant at 10%-level and in the ML + PSL model it is no longer significant. The positive sign of this estimate means that people who are cycling to work have a higher preference for a longer distance route than people who cycle to another location, this is not as expected because commuters have a more constrained time schedule (Ton, Cats, et al., 2017).

The effect of the end location being the home location on the sensitivity to distance is only significant in the PSL model and at a 10%-level. The sign is positive which suggests that respondents that are travelling home have a higher preference for a longer distance than respondents cycling to a different location.

In both the ML models the spread of the distance variable is significant, this indicates that taste variation exists in the sample that cannot be captured by the mean of distance.

4.2.2 Results e-bike models

In the e-bike MNL and PSL model ‘gender’ has a significant negative effect, meaning that women prefer a longer distance to a lesser extent than men when riding an e-bike. Women’s increased sensitivity to distance is also found by (Heinen et al., 2013; Stinson & Bhat, 2004) but as mentioned earlier other studies found no significant effect. The ML and the ML + PSL model the estimates are no longer significant, this suggest that the effect of gender is over estimated due to the multiple response data being treated as individual data in the MNL and PSL models.

In the e-bike models the sign of age category ‘<35 years’ are all negative, however only the MNL model is significant and at the 10%-level. The age category ’35-44 years’ has a negative estimate that is significant for the MNL and PSL model, significant in the ML model at a 10%-level. The negative sign of the estimate indicates that respondents in the group between 35 and 44 years prefer a longer distance to a lesser extent than respondents over 55 years old.

The age category ’45-55 years’ has a similar result, however this negative estimate is also significant at 5%-level in the ML + PSL model. The results for the different age categories show that a higher age, 55 years and older, means a higher preference for longer distances. This result was also found by Shafizadeh & Niemeier, (1997) and Stinson & Bhat, (2003).

The estimate for ‘weekday’ in the e-bike models is significant in all models and has a negative sign, which indicates that during weekdays respondents using an e-bike during a trip prefer a longer distance to a lesser extent than during the weekend.

Travelling during peak hours has a significant negative effect on the utility, this indicates that during peak hours respondents prefer a longer distance to a lesser extent, this is in line with findings of other studies. During peak hours a lot of commuters travel to work, they have a more constrained time schedule and don’t like to detour a lot (Broach et al., 2012; Li et al., 2017; Ton, Cats, et al., 2017)

‘Daylight’ has a negative estimate which suggests that when respondents use an e-bike they have a lower preference for a longer distance during the night. However, this effect is only significant in the PSL model.

Similar to the bike models, the estimate for ‘PC work’ is positive in the e-bike models, but the effect is highly significant in both the ML and the ML + PSL model. This indicates that respondents traveling to work prefer longer distance routes to a higher extent than when travelling to another location.

The estimate for ‘PC home’ is only significant at a 10%-level in the ML model. The positive sign indicates that when the end location of the trip is the home location, respondents have a higher preference for a longer distance compared to trips that have a different end location.

For the e-bike ML models the spread of the distance variable is significant. Again, this indicates that taste variation exists in the sample that cannot be captured by the mean of distance.

4.3 Discussion

Comparing the models for the different transport modes shows that the significant positive estimate for distance is similar for both modes of transportation. This indicates that the probability of a route being chosen increases when distance increases, this contradicts the base assumption of route choice modelling, and that people prefer the shortest route. A less negative estimate for the distance coefficient might be explained by a respondents imperfect knowledge of the network, they might not know a shorter path exists (Prashker & Bekhor, 2004; C. G. Prato, 2009), but this cannot explain a positive sign for the estimate. The positive estimate for distance may be caused by the alternative generation algorithm in combination with the few route specific variables included. The generation algorithm used is a k-shortest path algorithm which generated alternatives that are clustered around the shortest path, with little variety on other road attributes. The high Path Size variable coefficient, which corrects the alternatives for overlap, indicates that there are unobserved variables that have a negative influence in the utility function (Broach, Gliebe, & Dill, 2011). An example of this is that many of the shortest paths generated might all pass through a high density urban area and share a busy road segment with high motorized traffic volumes and no separate bicycle lane, while the chosen longer routes detours around the high density urban area, through an more rural landscape, and have little motorized traffic or a separate bicycle path, these variables are not included in the model while other studies found these attributes to have a significant effect. The effect of such unobserved variables is captured by the attributes that are included in the model, hence the positive estimate for ‘distance’ and the large Path Size estimate. The sample bias created by the data collection approach is also a reason for the higher preference for longer distance routes of participants. Part of the B-riders project is that participants are rewarded for every kilometre they travel by bike or e-bike, travelling longer distances will result in a bigger reward. The estimate for the standard deviation of distance is positive and significant for both bike and e-bike, meaning that for both bike trips and e-bike trips taste variation for distance exists in the sample. The fact that in the ML models the spread of the random variable distance is significant and the increase in model fit shows that the MNL and the PSL model do not account for taste variation and correlation between repeated choices. This complies with findings by Han et al.(2001), who apply an ML to RP data in order to accommodate for drivers' taste variations.

The age category '35-44 years' has a negative effect on the utility for both transportation modes but loses some significance when repeated choice is taken into account for e-bikes. For the age category '44-55 years' the estimate is not significant for bikes but does have a significant an effect for e-bikes. Respondents between 44 and 55 years old with an e-bike have a lower preference for a larger distance than people over 55 years old, while this difference in age category has no influence on the sensitivity to distance for trips made by bike.

Both 'peak' and 'weekday' have a significant negative effect on the utility of a route for trips made with an e-bike but not for trips made with a bike. This indicates that when respondents make a trip by e-bike they prefer a longer distance to a lesser extent during peak hours and weekdays while for respondents making a trip by bike 'peak' and 'weekday' don't have this effect. During weekday and peak hours traffic is characterized by commuters who are more time-constrained (Ton, Cats, et al., 2017). The higher average speed, one of the main reason for using an e-bike (Weinert, Ma, & Cherry, 2007), suggest that this group of travellers is more time conscience. Also there is a higher crash risk for e-bikes during peak hours (Hu, Lv, Zhu, & Fang, 2014). This difference in safety and crash risk with bikes explains why during peak hours respondents making e-bike trips are less willing to travel longer distances.

The variable 'daylight' has a significant positive effect in the bike model but not in the e-bike model. This suggests that people who travel by bike are indeed have a different preference for distance when it comes to whether it is day or night, but people who travel by e-bike don't have this difference in preference. The change in preference for different daylight conditions for bike trips is explained by the perceived unsafe conditions found by Stinson & Bhat (2004) and Gatersleben & Appleton (2007). This unsafety during the night is not experienced by e-bike trips. The perceived safety is affected by the actual safety. The actual safety, number of injuries, for e-bike trips is not influenced by daylight (Hu et al., 2014). This explains the lack of preference differences caused by daylight changes during e-bike trips.

The 'PCwork' estimate has a positive significant effect on the preferred distance for e-bike trips this contradicts other studies found that decreasing distance was most important for trip to work (Allemann & Raubal, 2015; Broach et al., 2012; Dill & Gliebe, 2008). The positive estimate is indeed counterintuitive when assuming that the majority of travellers during peak hours are commuters travelling to work. The positive estimate for 'PCwork' would contradict the estimate found for 'peak', a correlation test shows that a relation between these two variables is not that clear. The positive estimate might be explained by trip-chaining, people drop their kids off at school and travel further to work, but majority of people travelling to work do this directly (Department for Transport, 2014). Why the estimate is only significant for e-bike trips and not for bike trips is likely because respondents making a trip by e-bike in general travel, and prefer, longer distances than respondents making a trip by bike. The significance of the 'PCwork' estimate indicates that when respondents are using an e-bike travelling to work other route characteristics play a more important role in their route choice decisions. Which causes them to detour even further compared to when they are not travelling to work.

5.0 Conclusions

5.1 General conclusion

The main goal of this research was to analyse route choice behaviour of e-bike users. First a literature study was conducted to determine the factors that play a role in route choice decisions by bike users and could influence route choice decisions by e-bike users. An overview of these variables can be found in Table 1. The literature review also contains an outset of several approaches to data collection and model estimation for route choice decisions, these are described in part 2.2 and 2.3. Extra attention was given to the GPS data collection and which applications it has for research into route choice decisions. Finally, 4 route choice models, MNL, PSL, ML and MNL + PSL, were estimated for both bike and e-bike trips and the results were compared. For the estimation of the models a GPS dataset acquired from the 'B-riders project', combined with network data acquired from OpenStreetMap was used. Extra variables were added with the use of secondary data on sunset-sunrise times and peak hours acquired from KNMI and Rijkswaterstaat, respectively.

This study distinguishes itself from other cyclists' route choice studies by the size of the final dataset and that there are several socio-demographic variables included that often lack in other cyclists' route choice studies. Of the socio-demographic included in this study 'gender' has no significant effect but 'age' does have a significant effect on utility of a route. This indicates the importance including socio-demographic variables in travel behaviour studies. Similar to other studies, this study applied the Multinomial logit and the Path Size logit model. Added to this are a Mixed Logit and a Mixed-Path Size logit combination model, these models take into account the correlation caused by repeated choice that is ignored in other studies and handles the overlap problem between alternatives. It also allowed to examine whether there was any taste variation within the sample. This proved to be the case for both bike and e-bike trips.

A comparison between the two transportation modes, described in full in section 4.3, indicates that the distance is an important determinant for both, but also indicates that several different variables, such as 'weekday' and 'peak hours', do have an effect on route choice when respondent make e-bike trips but don't affect the route choice decisions when travelling by bike. The other way around, 'daylight' does have an effect on the route choice decision for bike trips but not for e-bike trips. This direct comparison between e-bike and bike trips route choice decisions in combination with a large GPS dataset has rarely been made. The lack of research into route choice decisions by e-bike users and the small selective group of people who used them in the past caused them to be merged with bike users. The increasing share of e-bikes and the increasingly younger age group of e-bike users means that, even if in several occasions the same person can be both e-bike and bike user varying over different trips, merging e-bike with bike users should be supported by research. The results of this study show that there are indeed differences between these two transport modes and that policymakers should consider these differences. Furthermore, the positive estimate for 'distance' that seems counterintuitive can be explained by incentives given in the project from which the data is collected. This illustrates the value of encouragement through rewards used in the B-riders project.

Detailed design recommendations cannot be given due to the lack of route specific variables in this study. The literature study proved a lot more variables to be potential influencers on route choice, unfortunately only distance and travel time were present in the dataset. The effect of these unobserved variables likely explains the counterintuitive results, such as a preference for longer distance, in this study. Based on this study the travel behaviour of bike users and e-bike users should not be considered as the same. This means that the 'Ontwerpwijzer fietsverkeer' (Design manual for bicycle traffic) should be updated and expanded to take the increasing number of e-bikes into account if further increasing the bike/e-bike mode share is desired. This study also indicates the importance of further research in the travel behaviour of e-bike users.

5.2 Recommendation

This study has proven that the use of a large GPS dataset for the estimation of a route choice model is possible. The map matching was done prior to this study, at a time when the code for extracting route characteristics from OpenStreetMap was not available. This caused the lack of route specific variables in this study, which in turn led to counter intuitive results. In future it might become possible to extract the route characteristics. Repeating this study including the route characteristics could clarify if there is indeed a difference between bike and e-bike users as suggested by the results of this study. Such a study could also reveal different needs for infrastructure between these types of cyclists and provide more detailed design recommendations to infrastructure policymakers.

The k-shortest path algorithm that was used to generate the alternatives and create the choice sets, influenced the estimation results of the model. Future researchers need to understand that the selection of a generation algorithm will affect the choice sets of respondents and impact the results and model performance. A choice set should contain all reasonable alternatives but not more, adding unreasonable alternatives to a choice set will also affect the model fit, therefore it is necessary to spend time finding an optimal combination of generation algorithms and their calibration. Testing different alternative generation methods or combining them could lead to more realistic choice sets.

The dataset of this study only contained one month of data, March 2014, the total B-riders dataset contains several years of data. Using data of several years could reveal seasonal differences as well as the effects of changes in the infrastructure. The results of this study did show that the repeated response data cannot be ignored in such cases.

References

Allemann, D., & Raubal, M. (2015). Usage Differences Between Bikes and E-Bikes. In F. Bacao, M. Y. Santos, & M. Painho (Eds.), *AGILE 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities* (Vol. 217, pp. 201–217). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16787-9_12

Aultman-Hall, L. (1996, August 16). *Commuter Bicycle Route Choice: Analysis of Major Determinants and Safety Implications*. Doctoral dissertation.

Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2003). 80 weeks of GPS-traces: approaches to enriching the trip information. *Arbeitsbericht Verkehrs- Und Raumplanung*, (178).

Azevedo, J., Santos Costa, M. E. O., Silvestre Madeira, J. J. E. R., Vieira Martins, E. Q., A. Azevedo, J., Emília O. Santos Costa, M., … Queirós Vieira Martins de, E. (1993). An algorithm for the ranking of shortest paths. *European Journal of Operational Research*, *69*(1), 97–106.

Batty, M. (2012). Smart Cities, Big Data. *Environment and Planning B: Planning and Design*, *39*(2), 191–193. https://doi.org/10.1068/b3902ed

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), 235–247. https://doi.org/10.1007/s10479-006-0009-8

Ben-Akiva, & Bierlaire, M. (1999). Discrete Choice Methods and Their Applications to Short-Term Travel Decisions. *Handbook of Transportation Science*, (1985), 7–37. https://doi.org/10.1007/0-306-48058-1_2

Ben-Akiva, M., Bergman, M. . J., Daly, A. J., & Ramaswamy, R. (1984). Modelling inter urban route choice behaviour. In *9th International Symposium on Transportation and Traffic Theory* (pp. 299–330). Delft, Netherlands.

Ben-Akiva, M., & Bolduc, D. (1996). *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*. D'epartement d''economique, Universit'e laval with Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Ben-Akiva, M., & Ramming, S. (1998). Lecture notes: Discrete choice models of traveler behavior in networks. *Prepared for Advanced Methods for Planning and Management of Transportation Networks. Capri, Italy*, *25*.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, *35*(7), 677–693. https://doi.org/10.1016/S0191-2615(00)00014-X

Bierlaire, M., Chen, J., & Newman, J. (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, *26*, 78–98. https://doi.org/10.1016/j.trc.2012.08.001

Bovy, P. H. L., Bekhor, S., & Prato, C. (2008). The factor of revisited path size: Alternative derivation. *Transportation Research Record: Journal of the Transportation Research Board*, (2076), 132–140.

Bovy, P. H. L., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, *3*(3), 173–189. https://doi.org/10.1080/18128600708685672

Bovy, P. H. L., & Stern, E. (1990). *Route Choice: Wayfinding in Transport Networks* (9th ed.). Springer Netherlands.

Bradley, M. A., & Bovy, P. H. L. (1984). A stated preference analysis of bicyclist route choice.

In *PTRC Annual meeting* (pp. 39–53). London.
Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005
Broach, J., Gliebe, J., & Dill, J. (2011). Bicycle route choice model developed using revealed preference GPS data. *TRB 2011 Annual Meeting*, *5464*. Retrieved from ftp://ftp.hsrc.unc.edu/pub/TRB2011/data/papers/11-3901.pdf
Buehler, R., & Pucher, J. (2012). Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. *Transportation*, *39*(2), 409–432. https://doi.org/10.1007/s11116-011-9355-8
Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B: Methodological*, *36*(7), 577–592. https://doi.org/10.1016/S0191-2615(00)00041-2
Casello, J. M., & Usyukov, V. (2014). Modeling cyclists' route choice based on GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, (2430), pp 155–161. https://doi.org/10.3141/2430-16
CBS. (2017). Bevolking; Kerncijfers. Retrieved January 31, 2018, from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37296ned/table?ts=1517394809306
Cherry, C., & Cervero, R. (2007). Use characteristics and mode choice behavior of electric bike users in China. *Transport Policy*, *14*(3), 247–257. https://doi.org/10.1016/j.tranpol.2007.02.005
Department for Transport. (2014). National travel survey factsheet: trip chaining. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/509447/nts-trip-chaining.pdf
Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 269–271.
Dill, J., & Gliebe, J. (2008). Understanding and measuring bicycling behavior: A focus on travel time and route choice. *Bicycling*, (December). https://doi.org/OTREC-RR-08-03
Dill, J., & Rose, G. (2012). Electric bikes and transportation policy. *Transportation Research Record: Journal of the Transportation Research Board*, *2314*(15 mi), 1–6. https://doi.org/10.3141/2314-01
Feng, T., & Timmermans, H. J. P. (2013a). Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C: Emerging Technologies*, *37*, 118–130. https://doi.org/10.1016/j.trc.2013.09.014
Feng, T., & Timmermans, H. J. P. P. (2013b). Map matching of GPS data with Bayesian belief networks. *Journal of the Eastern Asia Society for Transportation Studies*, *10*, 100–112. Retrieved from http://easts.info/on-line/proceedings/vol9/PDF/P21.pdf
Fietsberaad. (2013). *Fietsberaadpublicatie 24*, versie 1, 1–85. Retrieved from http://www.fietsberaad.nl/library/repository/bestanden/Fietsberaadpublicatie_24_Feiten_over_de_elektrische_fiets_versie_1.pdf
Fiorenzo-Calatano, S. (2007). *Choice Set Generation in Multi-Modal Transportation Networks*. Delf Univesity of Technology.
Frejinger, E. (2008). Route choice analysis: data, models, algorithms and applications, *4009*. https://doi.org/10.5075/epfl-thesis-4009
Frejinger, E., & Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological*, *41*(3), 363–378.

https://doi.org/10.1016/j.trb.2006.06.003
Gatersleben, B., & Appleton, K. M. (2007). Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A: Policy and Practice*, *41*(4), 302–312. https://doi.org/10.1016/j.tra.2006.09.002
Goudappel Coffeng. (2011). *Workshop fietssnelwegen 1 maart : Wat levert het op ?* (pp. 1–37). Retrieved from http://www.fietssnelwegen.nl/Links/WorkshopFietssnelwegen.pdf
Han, B., Algers, S., & Engelson, L. (2001). Accommodating drivers ' taste variation and repeated choice correlation in route choice modeling by using the mixed logit. *Transportation Research Board 80th Annual Meeting*, (01).
Harvey, F., & Krizek, K. (2007). *Commuter Bicyclist Behavior and Facility Disruption*. *Transportation Research Board*. Retrieved from https://trid.trb.org/view.aspx?id=811576
Heinen, E., Maat, K., & van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation*, *40*(1), 23–43. https://doi.org/10.1007/s11116-012-9399-4
Heinen, E., van Wee, B., & Maat, K. (2010). Commuting by bicycle: An overview of the literature. *Transport Reviews*, *30*(1), 59–96. https://doi.org/10.1080/01441640903187001
Hensher, D. a. (1994). Stated preference analysis of travel choice: the state of practice. *Transportation*, *21*(2), 107–133.
Hensher, D., Rose, J., & Greene, W. (2005). *Applied Choice Analysis*.
Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transportation Research*, *3*(1), 63–75. https://doi.org/10.3328/TL.2011.03.01.63-75
Hu, F., Lv, D., Zhu, J., & Fang, J. (2014). Related risk factors for injury severity of e-bike and bicycle crashes in Hefei. *Traffic Injury Prevention*, *15*(3), 319–323. https://doi.org/10.1080/15389588.2013.817669
Huisman, D. J., & Hengeveld, J. (2014). *Opmaak van een methode voor fietstellingen in de provincie Antwerpen*. https://doi.org/10.1007/s13398-014-0173-7.2
Hunt, J. D., & Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, *34*(4), 453–470. https://doi.org/10.1007/s11116-006-9109-1
KNMI. (2014). Tijden van zonopkomst en –ondergang 2014 52°00'. Retrieved from http://www.sciamachy-validation.org/klimatologie/achtergrondinformatie/zonop2014.pdf
la Barra, T., Pérez, B., & Añez, J. (1993). Multidimensional path search and assignment. In *21st PTRS Summer Meeting* (pp. 307–319).
Landis, B., Vattikuti, V., & Brannick, M. (1997). Real-time human perceptions: Toward a bicycle level of service. *Transportation Research Record: Journal of the Transportation Research Board*, *1578*(January 1997), 119–126. https://doi.org/10.3141/1578-15
Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506
Li, S., Muresan, M., & Fu, L. (2017). Cycling in Toronto: Route choice behaviour and implications to infrastructure planning. *Transportation Research*, *5998*(January 2017).
Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 134–142. https://doi.org/10.1016/j.isprsjprs.2015.11.006

MacArthur, J., Dill, J., & Person, M. (2014). Electric bikes in North America. *Transportation Research Record: Journal of the Transportation Research Board*, *2468*(2468), 123–130. https://doi.org/10.3141/2468-14

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2009). Route choice of cyclists in Zurich: GPS-based discrete choice models. *Arbeitsberichte Verkehrs- Und Raumplanung, IVT , ETH Zurich*, (544), 1–25.

Milakis, D., & Athanasopoulos, K. (2014). What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network. *Journal of Transport Geography*, *35*, 120–129. https://doi.org/10.1016/j.jtrangeo.2014.01.009

Milieudefensie. (2015). *Luchtvervuiling en de gevolgen voor onze gezondheid*. Retrieved July 25, 2017, from https://milieudefensie.nl/publicaties/factsheets/informatieblad-luchtvervuiling-en-de-gevolgen-voor-onze-gezondheid

Misra, A., & Watkins, K. (2017). Modeling Cyclists willingness to deviate from shortest path using revealed preference data (Vol. 1, pp. 1–16).

Prashker, J. N., & Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transport Reviews*, *24*(4), 437–463. https://doi.org/10.1080/0144164042000181707

Prato, C., & Bekhor, S. (2006). Applying branch-and-bound technique to route choice set generation. *Transportation Research Record Journal of the Transportation Research Board*, *1985*, 19–28.

Prato, C., & Bekhor, S. (2007). Modeling route choice behavior: How relevant is the composition of choice set? *Transportation Research Record: Journal of the Transportation Research Board*, *2003*, 64–73. https://doi.org/10.3141/2003-09

Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8

Quddus, M. A., Ochieng, W. Y., & Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, *15*(5), 312–328. https://doi.org/10.1016/j.trc.2007.05.002

Ramming, M. S. (2002). Network knowledge and route choice. *Ph.D. Thesis, Massachusetts Institute of Technology*.

Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D., & Srivastava, M. (2010). Biketastic: Sensing and mapping for better biking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, *3*, 1817–1820. https://doi.org/10.1145/1753326.1753598

Rijksoverheid. (2016). *Agenda Fiets 2017-2020*. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2017/02/16/agenda-fiets-2017-2020

Rijkswaterstaat. (2017). *Publieksrapportage Rijkswegennet*. *Publieksrapportage Rijkswegennet* (Vol. 3e periode).

Romanillos, G., Zaltz Austwick, M., Ettema, D., & De Kruijf, J. (2016a). Big data and cycling. *Transport Reviews*, *36*(1), 114–133. https://doi.org/10.1080/01441647.2015.1084067

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences using a web-based survey to examine bicycle facilities. *Transportation*, *36*(5), 511–539.

Shafizadeh, K. R., & Niemeier, D. A. (1997). Bicycle journey-to-work: Travel behavior

characteristics and spatial attributes. *Transportation Research Board*, *1578*, 84–90.
Stichting BOVAG-RAI. (2016). *Mobiliteit in cijfers: Tweewielers 2016-2016*. Stichting BOVAG-RAI Mobiliteit. Amsterdam.
Stinson, M. a, & Bhat, C. R. (2003). An analysis of commuter bicyclist route choice using a stated preference survey. *Transportation Research Record: Journal of the Transportation Research Board, 634*(512).
Stinson, M., & Bhat, C. (2004). Frequency of Bicycle Commuting: Internet-Based Survey Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *1878*(512), 122–130. https://doi.org/10.3141/1878-15
Tertoolen, G., Ruijs, K., Vree, R. de, & Stelling, C. (2016). *Change is cool. Inzichten uit fietsstimulerings projecten*.
Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How do people cycle in Amsterdam ? Estimating cyclists ’ route choice determinants using GPS data from an urban area, 1–12.
Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2017). Evaluating a data-driven approach for choice set identification, 1–17.
Train, K. (1999). Halton Sequences for Mixed Logit. *Working Papers*, 1–18.
Train, K. E. (2003). Mixed Logit. In K. E. Train (Ed.), *Discrete choice methods with simulation* (Vol. 31, pp. 138–154). Cambridge: Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511753930.007
Usyukov, V. (2013). *Development of a cyclists’ route-choice model : An Ontario case study* (Thesis).
van de Coevering, P., De Kruijf, J., & Bussche, D. (2014). Policy renewal and innovation by means of tracking technology een innovatieve schakel tussen onderzoek en fietsbeleid. In *Colloquium Vervoersplanologisch Speurwerk 20*.
van der Zijpp, N., & Fiorenzo-Catalano, S. (2005). Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B: Methodological*, *39*, 545–563.
van Overdijk, R. P. J. (2016). *The influence of comfort aspects on route- and mode-choice decisions of cyclist in the Netherlands* (Thesis).
Wachotsch, U., Kolodziej, A., Specht, B., Kohlmeyer, R., & Petrikowski, F. (2014). Electric bikes get things rolling. The environmental impact of pedelecs and their potential. Federal Environment Agency, Germany. Retrieved from http://www.umweltbundesamt.de/publikationen/e-rad-macht-mobil
Weber, T., Scaramuzza, G., & Schmitt, K. U. (2014). Evaluation of e-bike accidents in Switzerland. *Accident Analysis and Prevention*, *73*, 47–52. https://doi.org/10.1016/j.aap.2014.07.020
Weinert, J., Ma, C., & Cherry, C. (2007). The transition to electric bikes in China: History and key reasons for rapid growth. *Transportation*, *34*(3), 301–318. https://doi.org/10.1007/s11116-007-9118-8
Wen, C.-H., & Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research Part B: Methodological*, *35*(7), 627–641. https://doi.org/10.1016/S0191-2615(00)00045-X
Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and deterrents of bicycling: Comparing influences on decisions to ride. *Transportation*, *38*(1), 153–168. https://doi.org/10.1007/s11116-010-9284-y
Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data

without choice sets of paths. *Transportation Research Part C: Emerging Technologies*, *75*, 183–196. https://doi.org/10.1016/j.trc.2016.12.009

6.0 Appendices

Appendix A: Data coding

id	Label	Value label	Value numeric
pp	respondent id number	numeric	
IdtripS	trip id number	numeric	
daylight	daylight	Night	-1
		Day	1
weekday	Weekday	Saturday or Sunday	-1
		Monday, Tuesday, Wednesday, Thursday or Friday	1
peak	During Peak hours	No	-1
		Yes	1
refRoute	route in choice set id number	numeric	
nij	Number of choices in choice set	numeric	
choice	Choice	Not chosen	-1
		chosen	1
dist	Distance in kilometres	numeric	
ttMin	Travel time in minutes	Numeric	
mode	Transporation mode	bike	-1
		e-bike	1
gender	Gender	male	-1
		female	1
A	Age	<35 years	1
		35-44 years	2
		45-54 years	3
		>55 years	4
A1	Age dummy <35 years		1
A2	Age dummy 35-44 years		1
A3	Age dummy 45-54 years		1
A4	Age dummy >55 years		-1
PCendW	End location is work location	No	-1
		Yes	1
PCendH	End location is home location	No	-1
		Yes	1
Nchoice	Number of choice sets per respondent	numeric	
Npp	Number of alternatives in all choice sets per respondent	numeric	
Nmode0	Number of choice sets per respondent for bike	numeric	
Nmode1	Number of choice sets per respondent for e-bike	numeric	
PS	Path Size	numeric	
lnDist	Natural log of distance	numeric	
lnTTmin	Natural log of travel time	numeric	
lnPS	Natural log of Path Size	numeric	

Appendix B: Nlogit results for models with distance

Appendix B.1 Bike MNL

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=DIS2
    $
Normal exit:   6 iterations. Status=0, F=    3032.367

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -3032.36702
Estimation based on N =   2649, K =   1
Inf.Cr.AIC  =   6066.7 AIC/N =    2.290
Model estimated: Jun 19, 2018, 16:02:16
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    2.95543***      .14923    19.80  .0000     2.66294   3.24791
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.2 Bike MNL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS= DIS2,GDIS2, A1DIS2, A2DIS2, A3DIS2, WDDIS2, PDIS2, DLDIS2,
PCWDIS2, PCHDIS2
    $
Normal exit:   7 iterations. Status=0, F=    3008.557

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -3008.55750
Estimation based on N =   2649, K =  10
Inf.Cr.AIC  =   6037.1 AIC/N =    2.279
Model estimated: Jun 19, 2018, 16:02:16
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    3.65460***      .42541     8.59  .0000     2.82081   4.48838
   GDIS2|    -.02554         .15241     -.17  .8669     -.32427    .27319
  A1DIS2|    2.43919**      1.04622     2.33  .0197      .38863   4.48975
  A2DIS2|   -1.02923**       .40368    -2.55  .0108    -1.82042   -.23804
  A3DIS2|    -.37768         .35045    -1.08  .2812    -1.06455    .30918
  WDDIS2|     .38682**       .16210     2.39  .0170      .06911    .70454
   PDIS2|    -.08985         .18263     -.49  .6227     -.44780    .26809
  DLDIS2|     .64112***      .21564     2.97  .0029      .21848   1.06376
 PCWDIS2|    1.22611***      .32732     3.75  .0002      .58457   1.86764
 PCHDIS2|     .18359         .15587     1.18  .2389     -.12192    .48909
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.3 Bike PSL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS= DIS2,GDIS2, A1DIS2, A2DIS2, A3DIS2, WDDIS2, PDIS2, DLDIS2, PCWDIS2,
PCHDIS2,lnPS
    $
Normal exit:   7 iterations. Status=0, F=    2339.182

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -2339.18162
Estimation based on N =   2649, K =  11
Inf.Cr.AIC  =   4700.4 AIC/N =    1.774
Model estimated: Jun 19, 2018, 16:02:17
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    4.08178***      .44371     9.20  .0000     3.21213   4.95143
   GDIS2|    -.09792         .15992     -.61  .5403     -.41136    .21552
  A1DIS2|    2.54752**      1.05530     2.41  .0158      .47918   4.61587
  A2DIS2|    -.78216*        .44169    -1.77  .0766    -1.64785    .08353
  A3DIS2|    -.25560         .35499     -.72  .4715     -.95137    .44017
  WDDIS2|     .45408***      .17098     2.66  .0079      .11895    .78920
   PDIS2|    -.21446         .18377    -1.17  .2432     -.57464    .14572
  DLDIS2|     .39782*        .23077     1.72  .0847     -.05447    .85012
 PCWDIS2|     .85547***      .32992     2.59  .0095      .20883   1.50210
 PCHDIS2|     .27530*        .16489     1.67  .0950     -.04789    .59849
    LNPS|    8.10190***      .28617    28.31  .0000     7.54101   8.66278
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.4 Bike ML

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=DIS2
    ;fcn=DIS2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE0
    $
Normal exit:   6 iterations. Status=0, F=    3032.367

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function     -3032.36702
Estimation based on N =   2649, K =   1
Inf.Cr.AIC  =   6066.7 AIC/N =    2.290
Model estimated: Jun 19, 2018, 16:02:17
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    2.95543***      .14923    19.80  .0000     2.66294   3.24791
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  22 iterations. Status=0, F=    2981.192

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                CHOICE
Log likelihood function     -2981.19248
Restricted log likelihood   -4746.37083
Chi squared [  11 d.f.]      3530.35671
Significance level               .00000
McFadden Pseudo R-squared      .3719006
Estimation based on N =   2649, K =  11
Inf.Cr.AIC  =   5984.4 AIC/N =    2.259
Model estimated: Jun 19, 2018, 16:21:23
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values -3032.3670  .0169******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     548 groups
Variable number of obs./group =NMODE0
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
    DIS2|    3.93370***      .53450     7.36  .0000     2.88610   4.98129
        |Heterogeneity in mean, Parameter:Variable
DIS2:GEN|    -.01335         .21151     -.06  .9497     -.42790    .40120
```

```
 DIS2:A1|    2.03714*       1.20299     1.69  .0904     -.32068   4.39495
 DIS2:A2|   -1.38452**       .58428    -2.37  .0178    -2.52969   -.23935
 DIS2:A3|    -.70852         .48582    -1.46  .1447    -1.66071    .24366
DIS2:DAY|     .76381***      .24617     3.10  .0019      .28132   1.24629
DIS2:WEE|     .25961         .18553     1.40  .1617     -.10403    .62324
DIS2:PEA|    -.02893         .19438     -.15  .8817     -.40991    .35204
DIS2:PCE|     .65490*        .35908     1.82  .0682     -.04889   1.35870
DIS0:PCE|     .08550         .17819      .48  .6313     -.26374    .43474
        |Distns. of RPs. Std.Devs or limits of triangular
  NsDIS2|    2.19031***      .30026     7.29  .0000     1.60181   2.77882
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.5 Bike ML + PSL

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=DIS2, lnPS
    ;fcn=DIS2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE0
    $
Normal exit:   6 iterations. Status=0, F=    2355.399

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable              Choice
Log likelihood function     -2355.39919
Estimation based on N =   2649, K =   2
Inf.Cr.AIC  =   4714.8 AIC/N =    1.780
Model estimated: Jun 19, 2018, 16:21:23
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    3.71035***      .15961    23.25  .0000     3.39753   4.02318
    LNPS|    8.06050***      .28301    28.48  .0000     7.50581   8.61519
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   26 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable              CHOICE
Log likelihood function     -2241.17237
Restricted log likelihood   -4746.37083
Chi squared [  12 d.f.]      5010.39694
Significance level               .00000
McFadden Pseudo R-squared      .5278135
Estimation based on N =   2649, K =  12
Inf.Cr.AIC  =   4506.3 AIC/N =    1.701
Model estimated: Jun 19, 2018, 16:48:52
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values -2355.3992  .0485******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     548 groups
Variable number of obs./group =NMODE0
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
    DIS2|    4.78603***      .72580     6.59  .0000     3.36348   6.20858
        |Nonrandom parameters in utility functions
```

```
    LNPS|    10.1953***      .38233    26.67  .0000      9.4460   10.9447
        |Heterogeneity in mean, Parameter:Variable
DIS2:GEN|    -.11261         .31478     -.36  .7205     -.72957    .50435
 DIS2:A1|    2.24611        1.69945     1.32  .1863    -1.08475   5.57696
 DIS2:A2|   -1.55049*        .88291    -1.76  .0791    -3.28096    .17999
 DIS2:A3|    -.82606         .72586    -1.14  .2551    -2.24872    .59660
DIS2:DAY|     .93777***      .30077     3.12  .0018      .34827   1.52727
DIS2:WEE|     .22104         .22818      .97  .3327     -.22619    .66826
DIS2:PEA|    -.10326         .22628     -.46  .6481     -.54677    .34024
DIS2:PCE|     .41968         .42986      .98  .3289     -.42283   1.26219
DIS0:PCE|    -.04076         .21963     -.19  .8528     -.47123    .38972
        |Distns. of RPs. Std.Devs or limits of triangular
  NsDIS2|    4.36578***      .42703    10.22  .0000     3.52881   5.20274
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.6 E-bike MNL

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=DIS2
    $
Normal exit:   5 iterations. Status=0, F=    21892.94

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -21892.93902
Estimation based on N =  14977, K =   1
Inf.Cr.AIC  =  43787.9 AIC/N =    2.924
Model estimated: Jun 19, 2018, 16:52:57
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    2.07657***      .07503    27.68  .0000     1.92951   2.22363
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.7 E-bike MNL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS= DIS2,GDIS2, A1DIS2, A2DIS2, A3DIS2, WDDIS2, PDIS2, DLDIS2, PCWDIS2,
PCHDIS2
    $
Normal exit:   6 iterations. Status=0, F=    21831.00

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -21831.00196
Estimation based on N =  14977, K =  10
Inf.Cr.AIC  =  43682.0 AIC/N =    2.917
Model estimated: Jun 19, 2018, 16:52:58
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    3.29914***      .20560    16.05  .0000     2.89618   3.70211
   GDIS2|    -.45227***      .07741    -5.84  .0000     -.60399   -.30056
  A1DIS2|    -.79530*        .40788    -1.95  .0512    -1.59473    .00412
  A2DIS2|   -1.16102***      .23980    -4.84  .0000    -1.63101   -.69103
  A3DIS2|   -1.29676***      .17575    -7.38  .0000    -1.64122   -.95229
  WDDIS2|    -.27234**       .11556    -2.36  .0184     -.49883   -.04585
   PDIS2|    -.13491*        .07909    -1.71  .0880     -.28992    .02009
  DLDIS2|    -.01729         .12597     -.14  .8909     -.26418    .22960
 PCWDIS2|     .13629         .10556     1.29  .1967     -.07061    .34319
 PCHDIS2|    -.07385         .08303     -.89  .3737     -.23658    .08888
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.8 E-bike PSL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS= DIS2,GDIS2, A1DIS2, A2DIS2, A3DIS2, WDDIS2, PDIS2, DLDIS2, PCWDIS2,
PCHDIS2, lnPS
    $
Normal exit:   7 iterations. Status=0, F=    13058.57

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -13058.57453
Estimation based on N =  14977, K =  11
Inf.Cr.AIC  =  26139.1 AIC/N =    1.745
Model estimated: Jun 19, 2018, 16:52:59
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
```

CHOICE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
DIS2	4.25252***	.21312	19.95	.0000	3.83481	4.67023
GDIS2	-.34857***	.08168	-4.27	.0000	-.50865	-.18849
A1DIS2	-.28202	.47343	-.60	.5514	-1.20991	.64588
A2DIS2	-1.31081***	.26219	-5.00	.0000	-1.82468	-.79694
A3DIS2	-1.70109***	.18674	-9.11	.0000	-2.06709	-1.33508
WDDIS2	-.67389***	.12371	-5.45	.0000	-.91635	-.43144
PDIS2	-.37935***	.08423	-4.50	.0000	-.54444	-.21426
DLDIS2	-.45502***	.12931	-3.52	.0004	-.70846	-.20157
PCWDIS2	.03547	.11128	.32	.7499	-.18264	.25357
PCHDIS2	.06224	.08812	.71	.4800	-.11047	.23496
LNPS	13.0816***	.15660	83.54	.0000	12.7747	13.3885

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Appendix B.9 E-bike ML

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=-1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=DIS2
    ;fcn=DIS2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE1
    $
Normal exit:   5 iterations. Status=0, F=    21892.94

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function    -21892.93902
Estimation based on N =  14977, K =   1
Inf.Cr.AIC  =  43787.9 AIC/N =    2.924
Model estimated: Jun 19, 2018, 16:53:00
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    2.07657***      .07503    27.68  .0000     1.92951   2.22363
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   19 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable               CHOICE
Log likelihood function    -20184.25105
Restricted log likelihood  -26835.18157
Chi squared [  11 d.f.]     13301.86104
Significance level               .00000
McFadden Pseudo R-squared      .2478437
Estimation based on N =  14977, K =  11
Inf.Cr.AIC  =  40390.5 AIC/N =    2.697
Model estimated: Jun 19, 2018, 19:10:55
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values **********  .0645******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     732 groups
Variable number of obs./group =NMODE1
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
    DIS2|    5.00920***      .66279     7.56  .0000     3.71016   6.30823
        |Heterogeneity in mean, Parameter:Variable
DIS2:GEN|    -.12515         .34756     -.36  .7188     -.80636    .55606
```

```
 DIS2:A1|   -2.09838        1.81466    -1.16  .2475    -5.65505   1.45828
 DIS2:A2|   -1.96379*       1.02573    -1.91  .0556    -3.97419    .04660
 DIS2:A3|   -1.69989**       .80225    -2.12  .0341    -3.27226   -.12751
DIS2:DAY|    -.06721         .18027     -.37  .7093     -.42053    .28612
DIS2:WEE|    -.91067***      .17824    -5.11  .0000    -1.26001   -.56132
DIS2:PEA|    -.30095**       .11977    -2.51  .0120     -.53569   -.06621
DIS2:PCE|     .62808***      .15953     3.94  .0001      .31540    .94076
DIS0:PCE|     .22984*        .12666     1.81  .0696     -.01842    .47810
        |Distns. of RPs. Std.Devs or limits of triangular
  NsDIS2|    8.13025***      .32995    24.64  .0000     7.48356   8.77694
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix B.10 E-bike ML + PSL

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=-1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=DIS2, lnPS
    ;fcn=DIS2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE1
    $
Normal exit:   7 iterations. Status=0, F=    13176.62

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function    -13176.61850
Estimation based on N =  14977, K =   2
Inf.Cr.AIC  =  26357.2 AIC/N =    1.760
Model estimated: Jun 19, 2018, 19:10:56
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
    DIS2|    2.17460***      .08153    26.67  .0000     2.01479   2.33440
    LNPS|    12.9263***      .15509    83.35  .0000     12.6224   13.2303
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   21 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable               CHOICE
Log likelihood function    -11599.62456
Restricted log likelihood  -26835.18157
Chi squared [  12 d.f.]     30471.11401
Significance level               .00000
McFadden Pseudo R-squared      .5677456
Estimation based on N =  14977, K =  12
Inf.Cr.AIC  =  23223.2 AIC/N =    1.551
Model estimated: Jun 19, 2018, 21:32:19
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values **********  .1127******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     732 groups
Variable number of obs./group =NMODE1
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
    DIS2|    4.88858***      .71404     6.85  .0000     3.48909   6.28807
        |Nonrandom parameters in utility functions
```

```
    LNPS|    15.5026***      .20002    77.51  .0000     15.1106   15.8946
        |Heterogeneity in mean, Parameter:Variable
DIS2:GEN|    -.02400         .37798     -.06  .9494     -.76483    .71683
 DIS2:A1|   -1.60537        2.00262     -.80  .4228    -5.53043   2.31968
 DIS2:A2|   -1.73061        1.10053    -1.57  .1158    -3.88762    .42640
 DIS2:A3|   -1.93130**       .87155    -2.22  .0267    -3.63950   -.22310
DIS2:DAY|    -.17139         .21012     -.82  .4147     -.58321    .24043
DIS2:WEE|    -.83160***      .18561    -4.48  .0000    -1.19539   -.46782
DIS2:PEA|    -.48419***      .13194    -3.67  .0002     -.74279   -.22558
DIS2:PCE|     .60778***      .18645     3.26  .0011      .24235    .97321
DIS0:PCE|     .19400         .14011     1.38  .1662     -.08061    .46861
        |Distns. of RPs. Std.Devs or limits of triangular
  NsDIS2|    8.78027***      .35163    24.97  .0000     8.09109   9.46944
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C: Nlogit results for models with travel time

Appendix C.1 Bike MNL

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2
    $
Normal exit:   4 iterations. Status=0, F=    3342.472

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -3342.47153
Estimation based on N =   2649, K =   1
Inf.Cr.AIC  =   6686.9 AIC/N =    2.524
Model estimated: Jun 19, 2018, 21:36:31
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|     .55576***      .05824     9.54  .0000      .44161    .66991
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.2 Bike MNL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2, GTT2, A1TT2, A2TT2, A3TT2, WDAYTT2, PTT2, DLTT2, PCWTT2, PCHTT2
    $
Normal exit:   6 iterations. Status=0, F=    3320.635

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -3320.63542
Estimation based on N =   2649, K =  10
Inf.Cr.AIC  =   6661.3 AIC/N =    2.515
Model estimated: Jun 19, 2018, 21:36:32
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|     .56220***      .18756     3.00  .0027      .19459    .92982
    GTT2|    -.10123         .06211    -1.63  .1031     -.22296    .02049
   A1TT2|    1.45189***      .51604     2.81  .0049      .44046   2.46331
   A2TT2|    -.60812***      .16576    -3.67  .0002     -.93301   -.28324
   A3TT2|    -.26684*        .14622    -1.82  .0680     -.55343    .01975
 WDAYTT2|     .01777         .06976      .25  .7989     -.11895    .15449
    PTT2|    -.10681         .06831    -1.56  .1179     -.24071    .02708
   DLTT2|     .30253***      .09353     3.23  .0012      .11920    .48585
  PCWTT2|     .05852         .11882      .49  .6223     -.17436    .29140
  PCHTT2|     .14670**       .06415     2.29  .0222      .02098    .27242
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.3 Bike PSL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2, GTT2, A1TT2, A2TT2, A3TT2, WDAYTT2, PTT2, DLTT2, PCWTT2, PCHTT2, lnPS
    $
Normal exit:   6 iterations. Status=0, F=    2700.281

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function     -2700.28081
Estimation based on N =   2649, K =  11
Inf.Cr.AIC  =   5422.6 AIC/N =    2.047
Model estimated: Jun 19, 2018, 21:36:32
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|    1.18889***      .20772     5.72  .0000      .78176   1.59601
     GTT2|   -.04528         .07076     -.64  .5222     -.18396    .09340
    A1TT2|   1.51907***      .53233     2.85  .0043      .47572   2.56242
    A2TT2|   -.92665***      .18117    -5.11  .0000    -1.28174   -.57155
    A3TT2|   -.43873***      .16322    -2.69  .0072     -.75864   -.11881
  WDAYTT2|     .00134        .07755      .02  .9862     -.15065    .15332
     PTT2|   -.06398         .07605     -.84  .4001     -.21304    .08507
    DLTT2|     .35143***     .11065     3.18  .0015      .13456    .56830
   PCWTT2|     .08715        .12911      .68  .4997     -.16590    .34020
   PCHTT2|     .17858**      .07188     2.48  .0130      .03771    .31946
     LNPS|    7.21336***     .25604    28.17  .0000     6.71153   7.71519
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.4 Bike ML

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=T2
    ;fcn=T2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE0
    $
Normal exit:   4 iterations. Status=0, F=    3342.472

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function     -3342.47153
Estimation based on N =   2649, K =   1
Inf.Cr.AIC  =   6686.9 AIC/N =    2.524
Model estimated: Jun 19, 2018, 21:36:32
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|     .55576***      .05824     9.54  .0000      .44161    .66991
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  28 iterations. Status=0, F=    3255.208

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable                CHOICE
Log likelihood function     -3255.20820
Restricted log likelihood   -4746.37083
Chi squared [  11 d.f.]      2982.32526
Significance level               .00000
McFadden Pseudo R-squared      .3141690
Estimation based on N =   2649, K =  11
Inf.Cr.AIC  =   6532.4 AIC/N =    2.466
Model estimated: Jun 19, 2018, 21:56:52
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values -3342.4715  .0261******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     548 groups
Variable number of obs./group =NMODE0
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      T2|     .27584         .32690      .84  .3988     -.36488    .91656
        |Heterogeneity in mean, Parameter:Variable
  T2:GEN|    -.00167         .14356     -.01  .9907     -.28305    .27971
```

```
   T2:A1|    2.05936***      .79048     2.61  .0092      .51005   3.60867
   T2:A2|   -1.06878***      .41032    -2.60  .0092    -1.87298   -.26457
   T2:A3|    -.70035**       .33062    -2.12  .0342    -1.34837   -.05234
  T2:DAY|     .54731***      .14297     3.83  .0001      .26709    .82752
  T2:WEE|     .03305         .10612      .31  .7555     -.17495    .24105
  T2:PEA|    -.19064*        .10458    -1.82  .0683     -.39560    .01432
  T2:PCE|    -.15269         .18422     -.83  .4072     -.51376    .20838
 T20:PCE|     .08725         .10202      .86  .3924     -.11271    .28722
        |Distns. of RPs. Std.Devs or limits of triangular
    NsT2|    2.01700***      .20995     9.61  .0000     1.60550   2.42850
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.5 Bike ML + PSL

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=T2, lnPS
    ;fcn=T2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE0
    $
Normal exit:   6 iterations. Status=0, F=    2728.235

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function     -2728.23531
Estimation based on N =   2649, K =   2
Inf.Cr.AIC  =   5460.5 AIC/N =    2.061
Model estimated: Jun 19, 2018, 21:56:52
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|    1.03793***      .06877    15.09  .0000      .90314   1.17271
    LNPS|    7.10253***      .25294    28.08  .0000     6.60679   7.59828
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Normal exit:  25 iterations. Status=0, F=    2547.670

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable               CHOICE
Log likelihood function     -2547.66988
Restricted log likelihood   -4746.37083
Chi squared [  12 d.f.]      4397.40190
Significance level               .00000
McFadden Pseudo R-squared      .4632383
Estimation based on N =   2649, K =  12
Inf.Cr.AIC  =   5119.3 AIC/N =    1.933
Model estimated: Jun 19, 2018, 22:15:27
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values -2728.2353  .0662******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     548 groups
Variable number of obs./group =NMODE0
Number of obs.=  2649, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      T2|     .96117**       .40727     2.36  .0183      .16293   1.75941
        |Nonrandom parameters in utility functions
```

```
    LNPS|    9.00199***      .32420    27.77  .0000     8.36657   9.63740
        |Heterogeneity in mean, Parameter:Variable
  T2:GEN|     .04725         .18812      .25  .8017     -.32146    .41595
   T2:A1|    2.21470**      1.00772     2.20  .0280      .23960   4.18980
   T2:A2|   -1.12002**       .54010    -2.07  .0381    -2.17859   -.06145
   T2:A3|    -.68806         .43011    -1.60  .1097    -1.53105    .15494
  T2:DAY|     .61570***      .16173     3.81  .0001      .29872    .93268
  T2:WEE|     .10456         .12178      .86  .3906     -.13413    .34325
  T2:PEA|    -.09005         .12089     -.74  .4563     -.32699    .14688
  T2:PCE|    -.10033         .21871     -.46  .6464     -.52900    .32834
 T20:PCE|     .14664         .11655     1.26  .2083     -.08180    .37509
        |Distns. of RPs. Std.Devs or limits of triangular
    NsT2|    2.90144***      .24307    11.94  .0000     2.42503   3.37784
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.6 E-bike MNL

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2
    $
Normal exit:   5 iterations. Status=0, F=    21979.46

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -21979.45942
Estimation based on N =  14977, K =   1
Inf.Cr.AIC  =  43960.9 AIC/N =    2.935
Model estimated: Jun 19, 2018, 22:15:28
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|   -1.16151***      .04452   -26.09  .0000    -1.24876  -1.07426
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.7 E-bike MNL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2, GTT2, A1TT2, A2TT2, A3TT2, WDAYTT2, PTT2, DLTT2, PCWTT2, PCHTT2
    $
Normal exit:   5 iterations. Status=0, F=    21777.52

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -21777.52059
Estimation based on N =  14977, K =  10
Inf.Cr.AIC  =  43575.0 AIC/N =    2.909
Model estimated: Jun 19, 2018, 22:15:29
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|   -1.05047***      .11182    -9.39  .0000    -1.26963   -.83130
    GTT2|    -.11562***      .04475    -2.58  .0098     -.20332   -.02793
   A1TT2|     .12557         .25170      .50  .6179     -.36776    .61889
   A2TT2|    -.30710**       .13692    -2.24  .0249     -.57547   -.03873
   A3TT2|    -.48046***      .09914    -4.85  .0000     -.67477   -.28615
 WDAYTT2|    -.43333***      .05918    -7.32  .0000     -.54932   -.31734
    PTT2|    -.47618***      .04812    -9.90  .0000     -.57049   -.38187
   DLTT2|     .10154         .06829     1.49  .1371     -.03231    .23538
  PCWTT2|    -.57181***      .06640    -8.61  .0000     -.70196   -.44166
  PCHTT2|    -.07424         .04758    -1.56  .1187     -.16750    .01901
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.8 E-bike PSL with interaction effects

```
|-> SAMPLE
    ;All
    $
|-> REJECT
    ;MODE=-1
    $
|-> NLOGIT
    ;LHS=choice,nij,refroute
    ;RHS=T2, GTT2, A1TT2, A2TT2, A3TT2, WDAYTT2, PTT2, DLTT2, PCWTT2, PCHTT2, lnPS
    $
Normal exit:   7 iterations. Status=0, F=    13398.91

-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function    -13398.90906
Estimation based on N =  14977, K =  11
Inf.Cr.AIC  =  26819.8 AIC/N =    1.791
Model estimated: Jun 19, 2018, 22:15:31
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
```

CHOICE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
T2	-.26489**	.12422	-2.13	.0330	-.50837	-.02142
GTT2	-.05329	.04750	-1.12	.2619	-.14639	.03981
A1TT2	.08651	.28551	.30	.7619	-.47307	.64609
A2TT2	-.37066***	.14356	-2.58	.0098	-.65204	-.08929
A3TT2	-.50852***	.10647	-4.78	.0000	-.71719	-.29985
WDAYTT2	-.42449***	.06729	-6.31	.0000	-.55637	-.29261
PTT2	-.36981***	.05022	-7.36	.0000	-.46825	-.27138
DLTT2	-.10029	.07362	-1.36	.1731	-.24458	.04400
PCWTT2	-.62249***	.07077	-8.80	.0000	-.76120	-.48377
PCHTT2	-.02353	.05084	-.46	.6435	-.12318	.07612
LNPS	12.5742***	.15176	82.85	.0000	12.2768	12.8717

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Appendix C.9 E-bike ML

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=-1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=T2
    ;fcn=T2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE1
    $
Normal exit:   5 iterations. Status=0, F=    21979.46

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function    -21979.45942
Estimation based on N =  14977, K =   1
Inf.Cr.AIC  =  43960.9 AIC/N =    2.935
Model estimated: Jun 19, 2018, 22:15:32
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|   -1.16151***      .04452   -26.09  .0000    -1.24876  -1.07426
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   24 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable               CHOICE
Log likelihood function    -19682.88070
Restricted log likelihood  -26835.18157
Chi squared [  11 d.f.]     14304.60175
Significance level               .00000
McFadden Pseudo R-squared      .2665270
Estimation based on N =  14977, K =  11
Inf.Cr.AIC  =  39387.8 AIC/N =    2.630
Model estimated: Jun 20, 2018, 01:02:31
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values **********  .0932******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     732 groups
Variable number of obs./group =NMODE1
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      T2|   -2.63494***      .44001    -5.99  .0000    -3.49735  -1.77253
        |Heterogeneity in mean, Parameter:Variable
  T2:GEN|     .05975         .23929      .25  .8028     -.40925    .52876
```

```
   T2:A1|    -.49069       1.26990     -.39  .6992    -2.97965   1.99828
   T2:A2|    -.41419        .71400     -.58  .5619    -1.81360    .98522
   T2:A3|    -.59758        .54672    -1.09  .2744    -1.66913    .47397
  T2:DAY|     .01634        .09621      .17  .8652     -.17223    .20490
  T2:WEE|    -.76355***     .09803    -7.79  .0000     -.95568   -.57141
  T2:PEA|    -.38512***     .06710    -5.74  .0000     -.51663   -.25360
  T2:PCE|    -.44056***     .09131    -4.82  .0000     -.61953   -.26159
 T20:PCE|    -.11093        .07160    -1.55  .1213     -.25127    .02941
        |Distns. of RPs. Std.Devs or limits of triangular
    NsT2|    5.69894***     .22373    25.47  .0000     5.26043   6.13745
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Appendix C.10 E-bike ML + PSL

```
|-> SAMPLE
    ;All$
|-> REJECT
    ;MODE=-1
    $
|-> RPLOGIT
    ;LHS=Choice,nij,refroute
    ;RHS=T2, lnPS
    ;fcn=T2(n)
    ;rpl=gender, A1, A2, A3, Daylight, Weekday, Peak, PCendW, PCendH
    ;halton
    ;pts=1000
    ;pds=NMODE1
    $
Normal exit:   7 iterations. Status=0, F=    13554.28

-----------------------------------------------------------------------------
Start values obtained using MNL model
Dependent variable               Choice
Log likelihood function    -13554.28470
Estimation based on N =  14977, K =   2
Inf.Cr.AIC  =  27112.6 AIC/N =    1.810
Model estimated: Jun 20, 2018, 01:02:32
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
Response data are given as ind. choices
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
      T2|    -.51847***      .04527   -11.45  .0000     -.60719   -.42975
    LNPS|    12.5585***      .15127    83.02  .0000     12.2620   12.8550
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Line search at iteration   24 does not improve fn. Exiting optimization.

-----------------------------------------------------------------------------
Random Parameters Logit Model
Dependent variable               CHOICE
Log likelihood function    -12029.37223
Restricted log likelihood  -26835.18157
Chi squared [  12 d.f.]     29611.61869
Significance level               .00000
McFadden Pseudo R-squared      .5517313
Estimation based on N =  14977, K =  12
Inf.Cr.AIC  =  24082.7 AIC/N =    1.608
Model estimated: Jun 20, 2018, 03:30:30
Constants only must be computed directly
               Use NLOGIT ;...;RHS=ONE$
At start values **********  .1055******
Response data are given as ind. choices
Replications for simulated probs. =1000
Used Halton sequences in simulations.
RPL model with panel has     732 groups
Variable number of obs./group =NMODE1
Number of obs.= 14977, skipped    0 obs
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
  CHOICE|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Random parameters in utility functions
      T2|   -1.33504***      .40138    -3.33  .0009    -2.12173   -.54835
        |Nonrandom parameters in utility functions
```

```
    LNPS|    13.7578***      .18087    76.07  .0000     13.4033   14.1123
        |Heterogeneity in mean, Parameter:Variable
  T2:GEN|     .07139         .21491      .33  .7398     -.34982    .49259
   T2:A1|    -.56980        1.13734     -.50  .6164   -2.79894   1.65934
   T2:A2|    -.54960         .64093     -.86  .3912   -1.80580    .70660
   T2:A3|    -.73299         .49019    -1.50  .1348   -1.69374    .22776
  T2:DAY|     .00792         .10410      .08  .9394     -.19612    .21195
  T2:WEE|    -.69422***      .10088    -6.88  .0000     -.89194   -.49649
  T2:PEA|    -.38603***      .07156    -5.39  .0000     -.52630   -.24577
  T2:PCE|    -.51443***      .10273    -5.01  .0000     -.71579   -.31308
 T20:PCE|     .03223         .07561      .43  .6699     -.11596    .18043
        |Distns. of RPs. Std.Devs or limits of triangular
    NsT2|    4.98313***      .19898    25.04  .0000     4.59313   5.37314
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```