Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Factors affecting the departure train station choice of cyclists in the western region of the Netherlands

Manoj Ashvin Jayaraj[a], Jullian van Kampen[b], Thomas Koch[b], Dr. Eric Pauwels[b], Prof. Dr. Rob van der Mei[b], Dr. Elenna R. Dugundji[a,*]

[a]*Vrije Universiteit, Department of Mathematics, De Boelelaan 1018, Amsterdam 1018 HV, Netherlands*
[b]*Centrum Wiskunde en Informatica, Science Park 123, Amsterdam 1098 XG, Netherlands*

Abstract

With 35,000 km of bicycle pathways, cycling is common among persons of all ages less than 65 years in the Netherlands. Bicycle is often seen as a standalone travel mode but when integrated as part of a multimodal trip with train can be an important solution for long distance journeys, often offering more flexibility and faster access time than other travel modes to a railway station. With 47% of people not preferring to depart from the nearest train station, we investigate in this paper which factors influence departure station choice on combined bicycle-train trips in the western region of the Netherlands, exploring effects of individual socio-economic characteristics of residents, their multi-modal trip attributes, as well as neighbourhood characteristics and station attributes. Observations from the Dutch National Travel Survey over years 2015-2017 where a train ride precedes a bicycle ride, or train ride is followed by a bicycle ride are considered for analysis. The choice set consists of four alternatives: the first alternative corresponds to the first closest train station from each departure postcode, and the second, third and fourth alternative accordingly correspond to the second, third and the fourth closest stations. Estimated results show that the distance to all four stations is negatively significant for departure station choice, whereas in-train travel time, expendable household income and household composition in different categories were not even significant at 10% level. Working hours per week, residential municipality outside a city region and sprinter station type are found to be positively significant.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: Departure train station choice, socio-economic characteristics, trip attributes, station attributes, neighbourhood characteristics, multinomial logit model

E-mail address of authors:ashvinmanoj@gmail.com, Jullian.van.Kampen@cwi.nl, Koch@cwi.nl, Eric.Pauwels@cwi.nl, R.D.van.der.Mei@cwi.nl, E.R.Dugundji@vu.nl.

1. Introduction

Bicycle is considered as an important mode of transport in the Netherlands: per the 2015 factsheet *(1)* of the Central Bureau of Statistics in the Netherlands (CBS), with 35,000 kilometers there are almost enough cycle paths to go around the world. The average Dutchman cycles on average 1000 kilometers a year and Dutch teens cycle even twice that adults. Cycling is not only a sport but also a part of life, used to commute to work (41% of trips made) and school (15%), to visit friends (41%) and go shopping (17%) *(1)*.

With cycling being an important part of daily life for Dutch people, it can bring along many challenges like insufficient parking spots for bicycle, increasing the complexity of exiting traffic system and many more, especially in areas near train or metro stations where cyclists can be found more in numbers. It will thus be interesting to analyze cycling behavior as part of multimodal trip. In this research, we analyze the relation between different socio-economic characteristics and the departure train station choice made by people traveling by bicycle as part of their train journey in four provinces of the Netherlands namely North Holland, South Holland, Flevoland and Utrecht.

Our paper is structured as follows. First we describe background on cycling in the Netherlands and briefly review literature relevant to our topic. Then we elaborate on the data used, the data selection process and patterns we observed in the data and how we imputed unknown values. Next we outline the choice set generation approach that we used and explain the model calibration procedures. Finally we describe our results and conclude with a discussion of those results and a recommendation for future research.

2. Background

The western region of the Netherlands host the four biggest cities Amsterdam, Rotterdam, The Hague and Utrecht and will be the focus area in our analysis. Over the past half century there has been a significant rise in bicycle use among the Dutch *(2)*. This growth is attributed to what may be called a bicyclist paradise: flat landscapes, safe separate bicycle infrastructure and an environment increasingly hostile to cars due to high gas prices, lack of parking and traffic jams. As per a transport and mobility report *(3)* prepared for the Ministry of Infrastructure and Environment in the Netherlands, there are more bicycles (22 million) than there are people (17 million); the Dutch own more bicycles per household than the number of members in the household. Pucher and Buehler *(4)* explain the reasons for Netherlands being unique when it comes to cycling which is because of the bicycle friendly policies implemented from the late 20th century. With such a high number of bicycles, a challenge is imposed when it comes to bicycle parking, especially at multimodal junctions like train stations and metro stations. In this paper we analyze the departure train station choice made by train passengers who bicycled to the train station. Since about half of the Dutch people do not simply choose the station nearest to their residence, we are interested to determine what factors strongly influence the departure train station choice in the western region of the Netherlands.

3. Literature Review

With our aim to understand the factors affecting the choice of departure train station among cyclist, we review below some key literature on factors affecting bicycle mode choice and railway station choice modeling. We will draw on these contributions when estimating our model.

Lee and Ko *(5)* explain the relationships between neighboring environment and residents bicycle mode choice with Seoul as their geographical scope for analysis. Their analysis shows that bicycle lane density and preference to travel shorter distance are some of the influencing factors that affect the bicycle mode choice in denser cities like Seoul. Although it is hard to compare a city like Seoul with a country like the Netherlands, the study by Pucher and Buehler *(4)* show that the extensive planning in the past half century to building a good infrastructure stimulated cycling in the Netherlands. Since Netherlands have an excellent facility for cycling, it will be interesting to see if the distance to the train station is an influencing factor in the train station choice among the Dutch cyclist and whether cyclist are willing to travel to farther stations.

Debrezion et al. (6) applied a multinomial logit model on the choice of departure railway station made by Dutch railway passengers, looking at which variables impact that choice: the distance to a station, the availability of park & ride (car parking), bicycle stands/safes/rental, taxi, car rental, parking, international service, the frequency of service at a station and the availability of an intercity (express train) service to each of the Dutch provinces with the cities Leeuwarden and Groningen separately. Per 4 digit postal code a choice set of three alternatives station was determined. Using data from the main rail operator in the Netherlands and statistics aggregated on four digit postal codes they determined that in 47% cases the passenger does not pick the nearest station, so distance is an important factor in the probability of a station being chosen. Moreover, the higher the frequency of service at a station the higher the probability of being picked, but this probability decreases when the distance increases. The intercity status of a station is the biggest role in explaining the choice of a departure station: The author's state that the intercity status of a station has on average an equivalent effect of a decrease of 2 km in distance or an increase in frequency of 300 trains per day. Additionally the presence of a park-and-ride facility poses sizable effect with about 35% of the intercity status effect.

Young and Blainey *(7)* cover a broad range of past research dating back to 1970s. They compare several researchers works based on their statistical approach used for modeling and also discussed on the drawbacks of different modeling technique. Their research provides an overview for the possible factors that can have an influence in station choice among rail passengers. It will be interesting to analyze whether these factors explained by Young and Blainey are influencing the station choice among cyclist in the Netherlands.

Kager et al *(8)* demonstrate the need to analyze the synergy between bicycle and public transport by considering Netherlands as a case study. They proposed a research agenda to analyze the synergy between bicycle and train in a single trip which can generate an integrated transport system that is both fast (because of train) and flexible (because of bicycle) for both short and long distance travels. The author's reports that cycling to a train station is faster than other feeder transit access to a train station through simulated study. However they expect this synergy to be highly sensitive to shorter cycling distance and less sensitive to longer train distances when compared to car-based mobility practice.

Seeing the study by Debrezion et al (6) and Kager et al (8) we see an interesting research gap, in the effect of distances on cyclists, especially since the group of train passengers that arrive by train is steadily growing.

4. Data

Yearly mobility survey from the Netherlands Central Bureau of Statistics (CBS) called Onderzoek Verplaatsingen in Nederland (OViN) was used for our research. OViN data includes both domestic and foreign trips among the Dutch people based on sampling to ensure the survey details records trips throughout the year in every province. The OViN data for 2017 contains 166 variables (attributes) which can be separated into three layers:

- *Layer1*: Information on the person surveyed, including their household composition, household income class, age group of household members, type of vehicles owned, main transport mode, residential information across different boundaries namely province, COROP, municipality, with the postcode only representative on the first 4 digit.

- *Layer2*: Details on each journey taken by the surveyed person, where the journey from home to work is consider as one trip and from work to home is considered as a different trip, with a unique id for each trip. Trips are also categorized under different motives, the duration and distance travelled for each trip is also recorded along with the origin and destination geographical boundaries (4 digit postcodes) and the time of travel. Since each trip can have multiple rides,

- *Layer3*: Providing more information on individual legs of that form each trip. It includes information such as length of the leg, duration, distance, mode of transport for each leg and their order of travel, but their geographical information is not available. In case of a train ride it includes a unique station code for both origin and destination train station is available.

Every ride has a primary key, which is an extension of the trip id. Combining the three we can imagine a data structured from right to left, where layer-3 having individual ride information, layer-2 contains the accumulated information for all rides under each trip and layer-1 contains repeated observations of socio-economic characteristics corresponding to each person. Every individual has been distinguished by means of a unique id.

Initially, the data contained missing entries on some variables or the respondents who filled in the survey did not proved enough information about their travel. Information like departure and arrival municipality code, their province code, corop region number and travel distance was imputed based on other travel information by CBS. Illogical answers like journey by bicycle with average speed of more than 30kms hour, journey by walk for more than 40 km are also removed by CBS during pre-processing. Since our research was focused only on the departure station choice where train ride is preceded or followed by a bicycle ride, we only retained the rides with train as the mode of transport which are followed or preceded by a bicycle ride. But the bicycle travel time to the train station and bicycle travel distance to the train station are retained as part of the train journey entry, because the distance travelled by bicycle can possibly have an influence in determining the station choice (5). The OViN data was available from 2011 till 2017 for our research but the train station code variable was only available from 2015, so the research was performed using three years data (2015-2017).

Initially, 1689 observations are filtered from 2015-2017 OViN data. This data consist of train rides inside the four chosen provinces as well as train rides going out of the four chosen provinces in The Netherlands and vice versa. Each train ride chosen has a bike ride before or after or at both ends of ride. The travel patterns identified from the first selection as follows:

- Train rides originating from the other provinces into any one of the four provinces and vice versa contributes to 16.9% (8.8% and 8.1% respectively), with only 6.2% of those train rides precede or follow a bike ride inside the four provinces. In the case of train ride originating from outside the four provinces, 54.7% of the observations does not have bike ride after their train journey and these observations were discarded from the first selection.

- People travelling within the four provinces contribute to remaining 83.1% of the first selection.

- Train rides with bicycle activity recorded at both ends of the journey accounts for 14.5% in the filtered observations of which 13.2% of the rides were inside the four provinces and 1.3% contributes to rides in and out of the four provinces.

- 14.8% of the train rides from the filtered observations were identified as return trip. Return trips within a survey response are identified if and only if the response consisted of more than one trip, with the train as the main mode of transport in each trip. However there are overlaps between these four patterns.

Bicycle journey reported at both ends of the train ride can be classified into three types, where 70% of the respondents with bike-train-bike journey in a trip are travelling with work motive.

- Respondents parked a bicycle at the origin as well as the destination train stations, possibly due to a daily commute to the same destination.

- Respondent travelled in train with a folding bike which is allowed in the trains in The Netherlands at no extra cost.
- Respondents hired a bike at a rent shop (possibly OV-fiets *(9)*) at anyone end of the journey.

Since there is no detailed information in the OViN data about the usage of portable bikes we assumed that those respondents with bike-train-bike recorded in a trip owned bicycles at their departure train station as well as at their destination train station. In case of OV-fiets, it comes with a cost of €3.84 per 24 hours and no monthly subscription available on OV-fiets at least until the beginning of 2018. So this option looks expensive even for a person working five days a week. Thus, we ignored the third possibility and assumed that the OViN respondents who reported a bike-train-bike trip had bicycles at both ends of their journey. This assumption can be justified with the number of bicycles exceeding the population in the Netherlands by 6 million.

4.1. Imputation of destination train station code and data filtering

The origin and destination station code in the OViN data is a key variable for our station choice analysis, but 3.2% and 62.3% of the origin and destination station codes were reported as unknown in the first selection. Since OViN data has detailed information on every ride, the destination train station code can be imputed to some extent in case of entries with more than one trip with train as the main mode of transport. This is possible only when the sequence number of the train ride in the first trip matches to the sequence number of train ride in the second trip in reverse order. However, there were only ten instances with more than two trips being reported; these instances were manually checked for possible imputation after which the percentage of unknown destination station code remains at 40%.

Since we are only focused on the departure station choice we also removed the return trips from the first selection, which are equivalent of having a duplicate entry and can affect the station choice probabilities. In case of train rides originating from outside the four provinces but not followed by a bike ride at their destination station inside the four provinces are also removed from the first selection, but if the train rides follow a bike ride then the destination station is considered as the departure station and their corresponding return trips are discarded, this results in 1361 observations for our analysis and will be referred to as second selection which consist of 1361 observations.

4.2. Respondent Characteristics

The second selection comprises of 51.7% females and 48.3% males. Distribution of number of bicycle in a household shows: 10% accounts for single bicycle; 18.2% for two bicycles; 17.9% for three bicycles; 16.5% for four bicycles; and rest for more than four bicycles. In case of Household composition, singles person share is 22%; couples share 23.5%; couples with children and other members share 44%; single parent with children shares 5.5%; and other compositions contributes to 5%. Working hours per week distribution consist of 57.2% working more than 29 hours a week; 12-30 hours work per week shares 13%; working less than 12 hours per week accounts to 5.2%; no paid work shares 22.1%; not asked or children less than 15 years corresponds to 2% and remaining as unknown. As for spendable income of the Household, 51% of the observations spend 50,000 Euros or more. Trip motives distribution from the first selection consist of 57.5% for work and work-related trips, 19% for education, 9.7% for social/recreational purposes, 7% towards visit/stay and remaining contribute to motives like services/personal care, shopping, touring and other motives. Thus working and education class contributes to three quarters of second selection.

4.3. Station Characteristic

NS being the backbone of the Dutch railway system operates two types of trains, the sprinter and the intercity where 78% of the respondents prefer stations with intercity stops and 17.5% prefer stations with sprinter stops and the remaining 5.5% contributes to stations not operated by NS rails. Sprinter train stops at all station between its origin and destination but not the intercity.

4.4. Travel Attributes

The minimum and maximum distance traveled (in kilometers) by bicycle to the train station are 0.10 and 13.50 with median value at 2.5 kilometers and their distribution consist of 96% travelling less than 6.1 kilometers. Although OViN data is preprocessed by CBS, we looked for possible outliers especially in case of continuous variable and in the case of distance variable a maximum of 13.5 km which was reported by a respondent as a motive to work might looks like an potential outlier, but as per the transport and mobility report of CBS-2016 *(3)* a maximum distance of 7.5 km is realistic among Dutch people and with the introduction of new routes with less obstacles specially dedicated for bicycles to connect two places, a distance up to 15km are achievable and as a result we do not look into those distances as outliers in our analysis. Also, 59.5% of the respondents have an in-train travel time less than or equal to 30 minutes; 30.5% travel between 31-60 minutes; remaining 10% travel more than an hour in train.

5. Station Choice Modeling

The mathematical framework for estimating the departure train station choice for commuters travelling to train stations by bicycle which is discussed in this paper is based on the theory of utility maximization as discussed in Debrezion et al *(6)*. Readers are advised to refer to Ben-Akiva and Lerman *(10)* for mathematical details. The utility function corresponding to an alternative j in the choice set C_n for an individual n is divide into two components: $U_{jn} = V_{jn} + \varepsilon_{jn}$, where U_{jn} is the total utility; V_j represents the systematic component of the utility (consist of a constant term and observed heterogeneity) and ε_j is the random part of the utility (also referred to as error term, which accounts for the unobserved heterogeneity). McFadden *(11)* shows that if ε_j follows an extreme value distribution function, the choice situation results in multinomial logit model (MNL). In this research, using the OViN data, we will estimate the coefficients of different attributes corresponding to every alternative by means of MNL model.

5.1. Choice Set Generation:

From the available 1361 observations, 183 unique departure station codes were identified where 25% of the departure train stations have only one respondent traveling by bicycle to those stations and 7% with more than 30 people choosing the same train station. Considering the 7% alone will leave us with a disaggregate choice set of 13 alternatives, but some of the categorical attributes might not have enough observations which can result in an unreliable estimate due to smaller sample size. Instead of having a station specific alternative with smaller sample sizes we define a disaggregate choice set consisting of four generic alternatives where the alternative one to four corresponds to the first, second, third and the forth closest station in every departure postcode reported in the OViN data and the stations are chosen to be present within a 10 km radius from the postcode centroid. Since the OViN data has information about departure postcode (4-digit) on trip level and departure train station code, we combined the share of the choices made by the respondents to each of the first four closest stations in a postcode, resulting in four alternatives namely first closest station contributing 58.2%, second closest station contributing to 18.5%, third closest station adding 5.4% and the fourth closest station share is 3% of the second selection, as part of the choice set and availability for an alternative is based on the position number of the chosen departure train station in their departure postcode. Out of 1361 observations 15% of the data does not fit into any alternatives, reason being that the station reported by the user was not among the first four closest stations from the departure postcode centroid or inconsistency in the data reported by the respondent, resulting in 1158 observations valid for model estimation.

5.2. Model Calibration Procedure

For estimating the coefficients associated to each attribute of an alternative we used Biogeme (12) software packages which uses the maximum likelihood estimation technique to estimate the coefficients. Using the choice set defined in the previous section, an iterative approach was followed where the first model only had the alternative specific constant in the utility function and observed heterogeneity has been added along with the constant term in an iterative process to obtain the final model. In every iteration the following measure are taken care to reduce the margin of uncertainties:

- High standard error: regroup the segmentation of categorical variable for possible reduction in standard error or exclude that variable with small sample size from the corresponding alternative.

- Correlations between coefficients (not considering the alternate specific constant) that are above 0.69 are removed by assigning a unique coefficient.

The utility function represented in equation 2 is not always linear as it resembles. Though several non-linear transformations exist, we applied one form of non-linear transformation namely piecewise linear approximation, on continuous variable and every new iteration was based on the findings from the previous or past iterations.

5.3. Multinomial Logit Model

Table-1 lists the factors that can possibly affect the bicycle parking choice at the departure train station. For each model the following specifications are listed in the table 7.1, coefficients corresponding to each explanatory variable and their t-test values within parenthesis, rho squared, adjusted rho squared, final log likelihood and the number of parameters estimated under each model. Column-3 corresponds to a base model with only the constant term. Column 4-13 corresponds to models with observed heterogeneity and alternate specific constant and the effect of adding observed heterogeneity can be seen in the value of the final log likelihood, which increases as the number of estimated parameters increases compared to the base model leading to a model with reasonable estimates.

5.3.1. Models with travel attributes

Model-2 represents a constant term and the bicycle travel distance to the train station from the departure postcode centroid. The distance variable gives a significant increase in the rho squared and log likelihood value. However, the negative sign on the t-test indicates that the distance has a negative effect in departure station choice. Also, compared to the closest station people preferring the farthest station gives more importance to the distance traveled by bicycle to the train station. The fact that the distance is negatively significant even for the farthest station can be surprising but this shows the willingness of the people to travel far in bicycle which is possible because of the infrastructure available for bicycle commuters with 35000km of dedicated bicycle path in the country plus The Netherlands is almost flat in its landscape making it comfortable to bicycle a longer distance. In our model estimation we used a unique coefficient for the distance variable of the first two closest station because of high correlation between the individual coefficients.

A log likelihood ration test on model-2 with unique coefficients satisfies the null hypothesis that the model holds true in case of unique coefficients for station-1 and station-2. Log likelihood ration test statistics: $-2[L(\beta_R) - L(\beta_U)]$where $\beta_R$denotes the estimated coefficients of the restricted model- the model that is true under the null hypothesis- and β_U denotes the coefficient estimates of the unrestricted model. This statistic is x^2distributed with ($K_U - K_R$ degrees of freedom, where K_U and K_R are the numbers of estimated coefficients in the unrestricted and restricted models, respectively. The x^2test statistic with one degree of freedom for model-2 equal to 0.014, so we cannot the reject the null hypothesis even at a 0.5 level of significance $x^2_{1,0.5}$=0.45

Adding the in-train travel time variable along with the bicycle travel distance does not produce a significant difference to the rho squared and log likelihood value and also the t-test of the in-train travel time variable is not significant at 10% level. However, this variable makes the alternate specific constant insignificant for station-1 and station-2 thereby nullifying the effect of the constant term for each alternative; we will get back to this while discussing the final model results.

Table 1 : Iterations of multinomial logit model

	Name	Estimated coefficients (t-test statistics)											
		Model-1	***Model-2***	***Model-3***	***Model-4***	***Model-5***	***Model-6***	***Model-7***	***Model-8***	***Model-9***	***Model-10***	***Model-11***	***Model-12***
	Constant for closest_Station_1	2.89 (18.00)	1.78 (3.65)	1.48 (1.58)*	2.14 (3.90)	1.63 (3.29)	1.56 (3.11)	1.78 (3.64)	1.71 (3.50)	1.57 (2.60)	1.51 (2.49)	1.53 (2.52)	1.33 (1.31)*
Alternate Specific Constant	Constant for closest_Station_2	1.77 (10.48)	1.79 (3.68)	1.49 (1.59)*	2.15 (3.93)	1.66 (3.37)	1.57 (3.15)	1.79 (3.69)	1.86 (3.80)	1.61 (2.66)	1.69 (2.78)	1.72 (2.82)	1.52 (1.50)*
	Constant for closest_Station_3	0.549 (2.81)	0.89 (1.56)*	0.587 (1.60)*	1.26 (2.03)	0.768 (1.33)*	0.668 (1.15)*	0.590 (1.00)*	0.9 (1.58)*	0.73 (1.09)*	0.753 (1.13)*	0.474 (0.70)*	0.276 (0.26)*
	coefficients of Distance traveled to closest_station_1 and closest_station_2 (distance travelled to station by bicycle in km)		-0.843 (-11.97)	-0.845 (-11.97)	-0.845 (-11.98)	-0.842 (-11.94)	-0.84 (-11.94)	-0.853 (-12.02)	-0.821 (-11.58)	-0.845 (-11.93)	-0.824 (-11.55)	-0.834 (-11.61)	-0.835 (11.62)
	coefficients of Distance traveled to closest_station_3 (distance travelled to station by bicycle in km)		-0.707 (-7.82)	-0.709 (-7.82)	-0.709 (-7.84)	-0.708 (-7.82)	-0.704 (-7.80)	-0.747 (-8.12)	-0.678 (-7.49)	-0.711 (-7.83)	-0.682 (-7.50)	-0.723 (-7.82)	-0.724 (-7.82)
	coefficients of Distance traveled to closest_station_4 (distance travelled to station by bicycle in km)		-0.524 (-5.58)	-0.525 (-5.58)	-0.525 (-5.39)	-0.503 (-5.28)	-0.513 (-5.43)	-0.522 (-5.62)	-0.496 (-5.26)	-0.486 (-5.08)	-0.457 (-4.73)	-0.463 (4.77)	-0.461 (-4.74)
Travel characteristics													
	Train travel time (<20 minutes)			0.0184 (0.37)*									0.0110 (0.21)*
	Train travel time (21-40 minutes)			-0.00668 (-0.26)*									0.00119 (0.04)*
	Train travel time (> 40 minutes)			0.00563 (0.48)*									0.000977 (0.09)*
	Single person household (Common coefficient for stations_1, station_2 and station_3)				-0.744 (-1.93)					-0.436 (-1.01)*	-0.436 (-1.01)*	-0.443 (-1.02)*	-0.443 (-1.00)*
	Household with only couples (Common coefficient for stations_1, station_2 and station_3)				-0.212 (-0.51)*					0.104 (0.24)*	0.104 (0.26)*	0.112 (0.26)*	0.108 (0.24)*
	Household with Single parent, children and others (available only for station_1 and station_2)				0.0489 (0.14)*					0.117 (0.32)*	0.117 (0.35)*	0.0.0573 (0.16)*	0.0558 (0.15)*
Socio-economic characteristics	No Working hours (Common coefficient for 3 stations)					0.794 (1.72)				0.869 (1.84)	0.845 (1.78)	0.855 (1.80)	0.847 (1.78)
	Working hours less than 12 hours per week (available only for closest_station_1)					0.548 (1.45)*				0.531 (1.60)*	1.14 (1.67)	0.597 (1.77)	1.14 (1.53)*
	Working hours between 12-30 hours per week (Common coefficient for 3 stations)					1.07 (1.65)				1.12 (1.51)*	0.563 (1.53)*	1.14 (1.53)*	0.596 (0177)
	Spendable income of €50,000 or more (Common coefficient for 3 stations)						0.606 (1.80)			0.501 (1.32)*	0.502 (1.32)*	0.477 (1.25)*	0.467 (1.22)*
Neighbourhood and station characteristics	Residential municipality not part of a city(coeffecient estimated for station_3 alone)							0.763 (2.91)				0.764 (2.90)	0.764 (2.90)
	Train station type as sprinter (available only for station_1)								1.04 (4.93)		1.04 (4.93)	1.04 (4.92)	0.764 (4.92)
	Rho squared	0.349	0.422	0.422	0.423	0.425	0.423	0.425	0.431	0.426	0.435	0.438	0.438
Model parameters	**Adjusted Rho squared**	0.347	0.418	0.416	0.418	0.419	0.419	0.420	0.427	0.418	0.426	0.429	0.427
	Final Log Likelihood	-1013.54	-899.471	-899.281	-897.551	-895.476	-897.793	-895.013	-885.605	-892.679	-878.846	-874.427	-874.371
	No of parameters estimated	3	6	9	9	9	7	7	7	13	14	15	18

5.3.2. Models with Socio-economic Characteristics:

Model-4 which adds the household composition variable to model-2 does not have a significant difference in the rho squared and log likelihood values. Household composition being a categorical variable is divided into the following categories, category-1 represents single person household, category-2 comprises only of couples, category-3 consist of couples with children and couples with children and other members in the household and category-4 consist of household members who are single parent, single parent with children and single parent with children and other members. The estimated coefficients are compared to household composition in category-3

Results show that people belonging to a single person household has as significant effect in choosing all the three stations in a negative way, whereas the households with only couples does not have a significant effect in determining the station choice across all three stations i.e. their coefficients are equal to zero at 10% significance level. In case of household with only single parent, single parent and children and single parent, children and others the estimated coefficients are not significant at the first two closest stations when compared to category-3. Because of very few observations under category-4 we did not estimate the corresponding coefficient for station-3.

Model-4 with household composition variable does not affect the significance level of the bicycle travel distance or its coefficient's and the reason for considering unique coefficients for the household composition categories is due to high correlation between the coefficients across alternatives. The x^2test statistics with 5 degrees of freedom for model-4 is equal to 4.92, so we cannot the reject the null hypothesis at 0.95 level of significance$x^2_{5,0.95}$=11.07. Thus, the reduced model-4 is valid with unique coefficients.

Model-5 is another extension of model-2 by adding the number of hours of working in a week. Working hours per week has been divided into six categories in the OViN data but we has grouped them four categories with category-1 corresponds to people with no paid work, children's less than 15 years of age who were not asked to answer this question and respondents who did not shared the information about working hours, category-2 contains responded with less than 12 hours of work per week, respondents working for 12-30 hours per week belong to category-3 and respondents working more than 30 hours per week belongs to the last category. The estimated coefficient for category-1 is positively significant but the coefficient for category-3 is somewhat close to significance at 10% level across all three stations. Category-2 which is only applicable for station-1 is just significant at 10% level, the significance are compared to respondents under category-4. Also, respondents in category-2 who work for less than 12 hours per week shows less importance to the first closest station than respondents under category-1. Again high correlation between coefficients across alternatives is the reason for considering a unique coefficient. Also, x^2test statistics with 4 degrees of freedom for model-5 is equal to 4.064, so we cannot the reject the null hypothesis at 0.95 level of significance $x^2_{4,0.95}$= 9.48.

Model-6 includes spendable household income as an observed heterogeneity in the utility function. OViN data has seven categories of spendable income starting from up to €10,000 till €50,000 or more with unknown income category also a part of it and because of relative low number of records in unknown income category we added them to the high income category which forms the major share of spendable income. Experiments with different grouping of income categories reported no significance among the spendable income category, however when comparing the spendable income of €50,000 or more with other category we observe that the respondents with spendable household income of €50,000 or more shows positive significance in making a station choice across all alternatives. As in the previous models, correlation is the reason for considering a unique coefficient for expendable household income across all alternatives. The x^2test statistics with 2 degrees of freedom for model-6 is equal to 0.44, so we cannot the reject the null hypothesis even at 0.5 level of significance $x^2_{2,0.5}$=1.38.

After adding different socio-economic characteristics to the model-2 we saw no significant change in the estimated coefficients and the t-test statistics of the bicycle travel distance, also the correlation between the travel characteristics and socio-economic characteristics variable are very low indicating that they are independent of each other in the model estimation.

5.3.3. Model with neighborhood and station characteristics

Model-7 and model-8 includes the city region and station type as the observed heterogeneity in addition to the bicycle travel distance variable in model-2. 57.9% of the respondents in the second selection does not belongs to a resident municipality which is part of a city. Considering the neighborhood characteristics across all the alternatives resulted in high correlation among the estimated coefficients with coefficient on the third alternative being positively significant in case of resident not part of city region, however a model with unique coefficients (to remove the correlation) for all three alternatives does not produce any significant result, which can be due to the percentage of observations choosing station-1 and station-2 that are higher than station-3, so as part of model-7 we only estimated the coefficients for station-3 alone and the estimated coefficient for residents not being part of a city is positively significant in choosing the third closest station compared to other stations. With two types of train stations in The Netherlands namely intercity and sprinter, estimating the coefficients on station type indicate that people prefer sprinter station over intercity in case of the first closest station compared to all other stations as observed in model-8.

5.3.4. Road to final model

Model-9 combines the distance traveled by bicycle to the train station and the socio-economic characteristics together, the estimate coefficients on the bicycle travel distance still remains significance at 10% level as in the model-2. However, the single person household which was significant in model-4 became insignificant when other socio-economic characteristics are added. Also, in case of working hours per week, significance of respondents under category-1 increases slightly but in case of category-3 the significance drops below the 10% level which was just significant at 10% level in model-5 and the significance of category-2 moved slightly towards the 10% level but remains insignificant. The spendable household income variable also becomes insignificant at 10% level.

Model-10 which is an extension of model-9 by adding the station type variable does not shows much significant changes except for the respondents who work less than 12 hours per week whose estimated coefficient is now significant at 10% level. Model-11 includes residential municipality location to model-10 and this new variable does not affect the significance of other coefficients but the rho squared and the log likelihood values has improved in model-11 compared to other models. It is also important to note that the constant term is still significant in model-11 on the first two alternatives even at 5% significant level. From the previous models we observed that in model-3 the constant term completely loses its significance on all alternatives with the addition of in-train travel time, so in model-12 we add the in-train travel time along with other variables of model-11 and the result is not surprising that the in-train travel time nullifies the significance of the constant term even though the in-train travel time is not significant at 10% level. Thus model-12 was the final model for our analysis.

6. Summary

This paper focuses on the different socio-economic characteristics of people, train station type, neighborhood and travel characteristics that can possibly have an influence in choosing a departure train station among those people travelling by a bicycle to the train station, in the four provinces of The Netherlands and OViN data from 2015 till 2017 has been used in this research. A disaggregate choice set consisting of four alternatives based on the four closest station near the postcode centroid, within a radius of 10km was considered for this research. A Multinomial logit model estimated the coefficients of the observed heterogeneity that are available for each alternative through multiple iterations and the results are reported in table 1.

However, the final MNL model does not includes the observed heterogeneity on all alternatives, where station type was only estimated for the first closest station and category-3 of household composition was estimated for the first two stations only. The reason for estimating the coefficients for specific alternatives alone is primarily due to the small sample size on the two farther stations under those categories. Also high correlation was observed between the estimated coefficients, with the help of log likelihood ratio test statistics we verified that the restricted model with unique coefficients still follow a Chi-squared distribution, thereby eliminating high correlations between the estimated parameters. Moreover, the motive of respondents travelling to work and education attributes to 75% of the data used for modeling and so the results discussed in the last section can be biased towards the choices made by the respondents travelling to work and education purposes.

7. Discussion

The results analyzed in this paper did not surprise our expectation, however for readers outside the Netherlands these results might be surprising, especially the effect of distance variable on the departure station choice and in this section we try to relate our results with the literatures reviewed in the earlier sections so readers outside the Netherlands can understand the difference.

Marcus & Simon *(7)* highlighted some of the important station choice modeling dating back to 1974. They have grouped the factors that influences the utility function of a station choice by relating to the station accessibility attribute, railway service attributes and decision maker characteristics which can be related to the travel characteristics, neighborhood characteristics and station characteristics and the socio-economic characteristics variable used in our research, thereby including some of the key attributes for departure station choice among bicycle commuters. Debrezion et al *(6)*, found based on an aggregated choice set that in the Netherlands about 47% of the cases passenger does not choose the nearest station as their departure station, which can be seen from the estimated coefficients of the bicycle travel distance to the train station in table-1, which is highly negatively significant and highlighting the fact that Dutch people indeed travel far in bicycle to board a train. The fact that three out of the four provinces considered for this research constitute the most urbanized area of the country, the results of the station characteristics comparing the intercity and sprinter stations type for the first closest station holds true as expected by Debrezion et al *(6)*. However, the income which was significant in determining the bicycle mode choice *(5)* was not significant in determining the departure station among bicycle commuters in our research and the destination neighborhood and destination station characteristics was not included in our final model. Land use mix which is an important factor for bicycle model choice *(5)* as well as station choice *(7)* is not present as part of our utility function.

8. Conclusion

Summarizing the effect of observed heterogeneity in the final model, we observed that the distance to the station by bicycle from the postcode centroid has a negative influence on all four train station expressing the willingness of the Dutch train passengers to travel to a station farther away to board a train, whereas the household composition and the spendable income of a household does not have any significant effect in the departure station choice. Comparing the respondents who work more than 30 hours per week with different categories of respondents such as working hours less than 30 hours per week, the people working 12-30 hours per week and people who do not work including children under 15 years, are just significant across the first alternatives where as people working less than 12 hours per week are slightly below the significance level. In terms of station type and neighborhood characteristics: people prefer to bicycle to a train station with only Sprinter service, if it is the first closest train station; residents living in a municipality that is not part of the city are significantly interested in choosing the third closest station than other stations when compared to the residents who are part of a municipality region inside the city. Finally, the in-train travel time which as a variable of utility function and available for all alternatives is far from begin significant at 10% level but it reduces the significance of the alternate specific constant in the utility function, making the final model independent of the alternate specific constant variable.

9. Recommendations

Although this paper focuses only on the factors influencing the departure train stations choice by bicycle commuters, this research can be extended to understand the choice behavior at any multimodal junctions and also forecast the demand for a new bicycle parking facility based on the estimates analyzed at existing bicycle parking facilities of similar characteristics. Importantly there should be more focus given to start collecting information on the bicycle travel and bicycle parking behavior. Understanding the choice of bicycle parking will also help in making better decision on policies concerning bicycle travel and parking, there by promoting the usage of bicycle.

Acknowledgement

The authors confirm contribution to the paper as follows: study conception and design: MJ, EP, RM, ED; data collection: MJ, JK, TK, ED; analysis and interpretation of results: MJ, JK, ED; draft manuscript preparation: MJ, TK, ED. All authors reviewed the results and approved the final version of the manuscript.

References

1. Statistics Netherlands (CBS). *Factsheet the Netherlands: Cycling country*, 2015. Available at: https://www.cbs.nl/NR/rdonlyres/9F9F3F71-9324-46D3-AD7E-076C59F8392D/0/2015factsheetnederlandfietsland_ENG.pdf. Accessed Jul. 1, 2018.
2. Dutch Bicycling Council. *Fietsberaad publication 7. Continuous and integral: The cycling policies of Groningen and other European cycling cities*, 2016. Available at: http://www.fietsberaad.nl. Accessed Jul. 1, 2018.
3. Statistics Netherlands (CBS). *Transport and mobility report*, 2016. Available at: https://www.cbs.nl/-/media/_pdf/2016/38/2016-transport-and-mobility.pdf. Accessed Jul. 1, 2018.
4. Pucher, J., and R. Buehler. Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 2008. 28:4, 495-528.
5. Lee, K. H., and E. J. Ko (2014) Relationships between neighborhood environments and residents' bicycle mode choice: a case study of Seoul. *International Journal of Urban Sciences*, 2014. 18:3, 383-395.
6. Debrezion, G., Pels, E., and P. Rietveld. Choice of departure station by railway users. *European Transport report*, 2007. 37: 78–92.
7. Young, M., and S. Blainey. Railway station choice modelling: a review of methods and evidence, *Transport Reviews*, 2018. 38:2, 232-251.
8. Kager, R., L. Bertolini, and M. Te Brömmelstroet, Characterization of and reflections on the synergy of bicycles and public transport. *Transportation Research Part A: Policy and Practice*, 2016. 85: 208-219.
9. OV-fiets. *A rental bicycle from the dutch railway service NS*. Available at: https://www.ns.nl/en/door-to-door/ov-fiets/. Accessed Jul. 1, 2018.
10. Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis: Theory an Application to Travel Demand*. MIT Press Cambridge, MA, 1985.
11. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (P. Zarembka, ed.), Academic Press, New York, 1973.
12. Bierlaire, M. BIOGEME: A free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland, 2003.