



# Understanding the relation between travel duration and station choice behavior of cyclists in the metropolitan region of Amsterdam

Jullian van Kampen<sup>1,2</sup> · Eric Pauwels<sup>1</sup> · Rob van der Mei<sup>1,2</sup> · Elenna R. Dugundji<sup>1,2</sup>

Received: 15 July 2020 / Accepted: 15 November 2020  
© The Author(s) 2021

## 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, it can be an important solution for long distance journeys, offering increased flexibility and faster access time compared to other travel modes. In this paper we investigate which factors influence departure station choice on combined bicycle–train and bicycle–metro trips in the metropolitan region of Amsterdam. Data from a mobile app was used to track an individual's travel behavior over the years 2018 and 2019. A discrete choice model was estimated to see whether people prefer to park their bicycle at the station with the shortest travel duration or one of the stations with a longer travel duration. The final results show that level of education and age negatively influence the choice for cycling to the second closest station. Furthermore, the results show that people with an origin inside Amsterdam prefer to travel to a train station regardless of their destination.

**Keywords** Departure station choice · Socio-economic characteristics · Trip attributes · Station attributes · Panel logit model · Multinomial logit model

## 1 Introduction

The use of motorized vehicles has increased at a rapid rate over the years, resulting in increased traffic casualties and traffic problems in urban environments. Therefore, policies were implemented to discourage car use and encourage other modes of transport like bicycle and public transit. These policies made bicycle one of the most important modes of transport in the Netherlands. Over the past half century there has been a significant rise in bicycle use among the Dutch according to CBS (2006) (Central Bureau of Statistics). Additionally, CBS (2015) reports that with a length of 35,000 km, there are almost enough cycle paths to go around the world. Furthermore, they report that the average Dutchman cycles on average 1000 km a year while Dutch teenagers cycle even twice the amount of adults. Accordingly, cycling has become not only a sport but also a modality used

to commute to work (41% of trips made) and school (15%), to visit friends (41%) and go shopping (17%).

The Netherlands may be called a bicyclist paradise due to its flat landscapes, safe separate bicycle infrastructure, hostile car environment, lack of parking and traffic jams. According to a transport and mobility report from the CBS (2016), 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 there are number of members in the household. Unsurprisingly, bicycles have become the most important travel mode in urban environments illustrating that bicycle has become an important access mode in multimodal transportation networks over the years.

As cycling becomes an important part of daily life for Dutch people it brings along many challenges, like insufficient parking spots for bicycle, increasing complexity of existing traffic systems and many more. Especially in urbanized areas near train or metro stations, providing sufficient parking facilities can be challenging. In this research, we analyze the relation between different socio-economic characteristics and the departure station choice made by people using bicycle as an access mode in the Netherlands. Since about half of the Dutch people don't choose the station

✉ Jullian van Kampen  
Jullian.van.Kampen@cwi.nl

<sup>1</sup> Centre for Mathematics and Computer Science, Science Park 123, Amsterdam 1098 XG, The Netherlands

<sup>2</sup> Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam 1018 HV, The Netherlands

closest to their residence (Debrezion et al. 2007), we are interested to determine which factors strongly influence the departure train station choice.

Our paper is structured as follows. First we briefly review literature relevant to our topic. Then we elaborate on the data used, data selection process and how we imputed unknown values. Next we outline how the choice set is generated for each individual, and explain the model calibration procedures. Finally we describe our results and conclude with a discussion of those results including recommendations for future research.

## 2 Literature review

Our aim of this study is to understand which factors affects the choice of departure station among cyclists. Therefore, we review below some key literature on factors affecting bicycle mode choice and station choice modeling. An overview of this literature, their modelling methodologies and influential factors can be found in Table 1.

Young and Blainey (2018b) cover a broad range of past research about station choice modelling and influencing factors dating back to the 1970s. They compare several research works based on their statistical approach used for modeling and also discussed the drawbacks of different modeling technique. They found that the most common used statistical models are closed-form, where multinomial logit models are used to model station choice and nested logit models are commonly used to model the combined choice of access mode and station choice. However, these modelling approaches have a weakness, namely the inability to account for spatial correlation and patterns of substitution which represent competition between stations. Additionally, they also provide an overview of the possible factors that can have an influence in station choice among rail passengers. They found that access and rail service factors have been consistently reported in previous research works. These works conclude that station utility decreases as the journey contains more transfers, has higher distance, has increased rail leg journey time, a higher fare and lower service frequency.

However, they also conclude that limited attention has been given to land-use factors in station choice modelling and state that these factors may influence predictive performance significantly.

Young and Blainey (2018a) describes the development of a station choice model which can be used to generate probabilistic station catchment area's, which could be integrated in trip end or trip flow models. They used a revealed preference survey from the Welsh and Scottish governments for model calibration and a trip planner for mode-specific station access variables and train leg measures. They used multinomial and mixed logit models to calibrate their models on each of the datasets. Their alternative choice set is created by considering the ten closest stations by road distance from the observed choice. Finally, they assess model transferrability by analyzing the parameter estimates along with the confidence intervals, and using the model parameters from each dataset and predict choice in the other dataset. As a result, they show that it is possible to calibrate station choice models applicable in both trip-end and flow rail demand models.

Debrezion et al. (2007) applied a multinomial logit model on the choice of departure railway station made by Dutch railway passengers, looking at which variables impact that choice including the distance to a station, the availability of park and 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. A choice set of three alternatives station was determined for each 4 digit postal code. Using data from the main rail operator in the Netherlands and statistics aggregated on four digit postal codes they determined that in 47% of the cases the passengers do not select the nearest station, making distance an interesting factor in the probability of a station being chosen. Additionally, as the frequency of service at a station increases, the probability of choosing that station also increases, but as distance increases the probability of choosing a station decreases. The intercity status of a station was the most important factor in explaining the choice of a departure station. The author's state that the intercity

**Table 1** Factors affecting station choice

Author	Methodology	Influential factors
Adnan et al. (2019)	Hybrid choice model	Station characteristics, trip characteristics, socio-economic characteristics
Chakour and Eluru (2014)	Multinomial logit	Rail service, station characteristics, socio-economic characteristics
Debrezion et al. (2007)	Multinomial logit	Rail service, station characteristics, trip characteristics
Givoni and Rietveld (2014)	Nested logit	Rail service, station characteristics, trip characteristics
Lee and Ko (2014)	Random intercept logit	Trip characteristics, socio-economic characteristics, land-use
van Kampen et al. (2020)	Multinomial/panel logit	Station characteristics, trip characteristics, socio-economics characteristics
Young and Blainey (2018a)	Multinomial/mixed logit	Rail service, station characteristics, trip characteristics

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.

Givoni and Rietveld (2014) studied the amount of railway station required in urbanized environments and whether reducing the amount of stations in an urbanized environment has an impact. In their research Amsterdam is brought up as a case study. They use a nested logit approach to estimate the access mode and departure station of an individual. The variables used in the model include rail journey time, quality of the station perceived from the access mode, travel time, and access mode distance. Results of the nested logit model were used to estimate the effects of closing a station based on welfare gains and losses using a logsum approach. They conclude that no justification could be found in reducing the number of stations, but point out that their analysis shows that adding additional stations, or rail services might have a positive welfare effect.

Chakour and Eluru (2014) made an attempt to create a framework to better understand access mode and train station choice of train commuter behavior. By using a latent segmentation approach they jointly model access mode and station choice decisions in which the order of choices is irrelevant. Apart from rail service and a few station characteristics they also include socio-economic characteristics as explanatory variables affecting station and mode choice. Their results show that as distance from the station increases people are more likely to select access mode first, while presence of parking and train frequency increases the likelihood for choosing a station. Additionally, they found that young people, females, car owners and individuals leaving before 7:30 a.m. are more likely to drive to the train station.

Lee and Ko (2014) study the relationships between neighboring environment and residents bicycle mode choice with Seoul as their geographical scope for analysis. They used neighbourhood environment, and socio-demographic factors as explanatory variables in a random intercept logit model. Their analysis shows that bicycle lane density affects the bicycle mode choice in denser cities like Seoul, implying that the accessibility of bicycle lanes is an important factor for planners in order to encourage bicycle use. Additionally, socio-economic characteristics like gender, income, occupation, vehicle ownership, shorter travel distances and housing type all showed statistical significant correlation with bicycle use. Moreover, the study showed that neighbourhoods with high levels of mixed land-use result in more bicycle travel. On the other hand, residential density did not show any statistically significant correlation.

Although it is hard to compare a city like Seoul with a country like the Netherlands, the study by Pucher and Buehler (2008) show that the extensive planning in the past half

century to building a good infrastructure stimulated cycling in the Netherlands. In their study they explain how bicycling is promoted in countries like the Netherlands, Denmark and Germany. They conclude that the key for success in these countries is a mixture of mutually reinforcing policies encouraging cycling. The most important approach is a combination of providing cycling facilities along busy roads and measures involving traffic calming in residential neighbourhoods. Additionally, traffic education, integration of public transport, bike parking and promotional events create a wide public support in these countries for cycling. Finally they note that car use in these countries is considerably more expensive than it is in other countries like the USA, due to taxes and restrictions on car parking, ownership and use.

Kager et al. (2016) demonstrates the need to analyze the synergy between bicycle and public transport by considering Netherlands as a case study. Their study explores the distinct characteristics of the bicycle-train combination and how these modalities can complement each other. They found that these two modalities have a strong synergy when considered as a single trip chain due to the high speed of the train, the high accessibility of the bicycle and the flexibility in combining both modes. Finally, They propose 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. 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.

Adnan et al. (2019) focus their study on public bicycle sharing systems in Belgium where bike sharing was implemented to reduce last-mile travel. In their research, a web-based survey questionnaire was used to gather data on peoples' socio-economic characteristics, travel habits, attitudes towards friendliness-to-cycling and responses to stated preference scenario's. They estimated a hybrid choice model to determine whether bicycle sharing was used as an access or egress mode to the rail station. As a result, they found out that bike rental costs, bad weather conditions, trip distances smaller than 1 kilometer and car availability from colleagues/parents/friends have a negative influence on the choice for a bike share. On the other hand, bike parking availability at destination and low bus service frequencies increase the chances of selecting a bike share.

This research is an extension to our previous work in van Kampen et al. (2020) where we researched the behavior of cyclists travelling to the train station in the western region of the Netherlands. The data that was used is from a national travel survey which was aggregated over the years 2015–2017. A multinomial logit model was used to estimate where an individual has the choice to travel to the closest station or a station further away. The choice set was

determined by looking at a 10 km radius around the place where the person lived. The results showed that people are willing to travel as far as the fourth closest station, prefer to travel to the closest station if that station is skipped by the intercity train, and that municipalities not part of a city are prepared to travel to the 3th closest station.

In this research we attempt to explain how travel duration impacts station choice. Additionally, although Young and Blainey (2018b) mentions the use of socio-economic characteristics, the literature on railway service and access factors is way more extensive compared to the literature on socio-economic characteristic. Therefore, in this paper we attempt to further investigate the effects of socio-economic characteristics in station choice.

### 3 Methodology

#### 3.1 Data description

Previous research has so far relied on survey data as a means to understand station choice. However, due to advancements in location tracking devices, GPS tracking data is a potentially better candidate in estimating station choice model. Boukhechba et al. (2018) shows that with the help of GPS tracking data and recent developments in data mining technologies it is possible to predict the next location of an individual even when irregular patterns are present in the data. They developed an algorithm capable of predicting a users' next location by learning from his habits and account for changes in their routine. The algorithm analyzes an individual his points of interests and stores these as a sequence. They test their approach using synthetic data and GPS trajectories. They found that their algorithm performs better compared to similar algorithms in terms of mobile resource usage and the ability to support users' habit changes.

For our research we will be using data from a mobile app which tracks an individual's travel behavior using GPS tracking. From these GPS records mapmatching and modality deduction has been applied to infer individual trip details. As a result, the data we work with is processed data containing information on an individual's trips on a specific date. From these trips we know the origin, destination, travel distance, duration, and modality of a trip. Additionally, the pre-processed data also records whether a trip was performed multimodal or unimodal.

#### 3.2 Data selection

Our research focuses on people travelling to from or within Amsterdam and was collected in three separate waves. The first wave started in 2018 from the 20th of June until 18th of August. The second wave also started in 2018 from the

16th of September until 28th of October. The last and third wave started in 2019 from the 31th of May until the 15th July 2019. For our research we focus on individuals that performed a train or metro trip with bicycle or electric bicycle as access mode. After removing records that did not involve a train or metro trip following a bicycle trip an initial dataset was created where we have 155 unique users that performed 600 unique bicycle trips. This data was further enriched with socio-economic characteristics of people using the app. These characteristics include information about an individual's income, age education and gender in the form of a categorical variable. However, the final dataset contained less trips which will be explained in the Sect. 3.4 on choice set generation.

#### 3.3 Cross variable: Station type x trip O-D type

For our modelling we also created dummy variables indicating origin or destination of a trip inside or outside of Amsterdam, and whether the parking location involved a metro station or a train station. From these dummy variables several cross-variables were made. Table 2 shows a count of these cross-variables where the rows show the chosen station and the columns show an origin-destination pair where "in" indicates in Amsterdam and "out" indicates outside of Amsterdam.

#### 3.4 Choice set

In order to estimate a model we need to generate a choice set for the trips that could have been chosen by the individual. Therefore, a choice set was generated using an open source library which was developed by Conway et al. (2017). Given a certain time window several alternative travel options were generated for a trip based on the origin, destination and available travel options. In this research the generated alternatives are metro and train stations that could have been chosen. We limit our choice set to 4 stations to choose from based on travel duration. The algorithm was able to generate additional alternatives for

**Table 2** Chosen station versus origin-destination

Station	out-in	in-out	in-in	out-out
Train 1	67	70	17	38
Train 2	38	50	9	23
Train 3	12	9	5	1
Train 4	0	32	8	3
Metro 1	3	23	54	0
Metro 2	1	2	14	0
Metro 3	1	2	9	0
Metro 4	1	1	12	0

most trips but for 95 trips no metro or train trips could be generated. Therefore, these trips were removed from the dataset resulting in a final dataset of 505 total trips with 141 unique users. Out of these 505 total trips there were 184 cases where 3 alternatives could be generated and there were 290 cases where only 2 alternatives could be generated.

### 3.5 Model framework

The mathematical framework for estimating the departure station choice for commuters using bicycle as access mode is based on the theory of utility maximization as discussed in Debrezion et al. (2007). Readers are advised to refer to Ben-Akiva and Lerman (1985) for mathematical details.

$$U_{in} = V_{in} + \epsilon_{in} \tag{1}$$

The utility function shown in Eq. 1 corresponding to an alternative  $i$  in the choice set  $C_n$  for an individual  $n$  is divided into two components where  $U_{in}$  is the total utility;  $V_j$  represents the systematic component of the utility (consist of a constant term and observed heterogeneity) and  $\epsilon_j$  is the random part of the utility (also referred to as error term, which accounts for the unobserved heterogeneity). McFadden (1973) shows that if  $\epsilon_j$  follows an extreme value distribution function, the choice situation results in a multinomial logit model (MNL). The probability that an individual chooses an alternative is given by Eq. 2.

$$P(i|C_n) = \frac{\exp V_{in}}{\sum_j \exp V_{jn}} \tag{2}$$

An important limitation of the MNL model is that it does not allow correlation between alternatives or correlation between individuals. In our data it is very likely that individuals perform the same trip every day as part of their daily schedule. Therefore, in this research we will estimate a panel logit model in order to allow correlation between these individual trips. This correlation is captured in an additional normal error term which allows us to account for repeated choices by the same individual. As a result the utility function changes as shown in Eq. 3 where  $t$  represents choice at time  $t$ . In this equation our  $\sigma_{in}$  is the same for each individual while the other error term is still varying over trips. Readers who are interested in the full mathematical details are referred to Train (2009).

$$U_{int} = V_{int} + \epsilon_{int} + \sigma_{in} \tag{3}$$

Consider a sequence of alternatives, one for each time period,  $i = i_1, \dots, i_T$ . Conditional on  $\beta$  the probability that an individual makes this sequence of choices is a product of multiple logit formulations as shown in Eq. 4.

$$L_{in} = \prod_{t=1}^T \frac{\exp V_{int}(\beta)}{\sum_j \exp V_{jnt}(\beta)} \tag{4}$$

Since we are dealing with normal error terms, the unconditional probability is an integral over all possible values of  $\beta$  as shown in Eq. 5. Simulation is required to generate the normal error terms to solve the integral in Eq. 5.

$$P_{in} = \int L_{in}(\beta)f(\beta)d\beta \tag{5}$$

We will start estimating the coefficients of different attributes corresponding to every alternative by means of an MNL model. Since the panel logit model does not have a unique optimal solution, we will use the coefficients of the MNL model as a starting point for simulated estimation of our panel logit model.

For estimating the coefficients associated to each attribute of an alternative we used the software package biogeme, developed by Bierlaire (2003), which uses the maximum likelihood estimation technique to estimate the coefficients. For the MNL model the default setting were used to estimate the model. For the Panel logit model Halton (1960) draws are used to generate random error terms, since we expect this will reduce simulation time as shown in Train (1999). The optimization algorithm that was used is CFSQP, which was developed by Lawrence et al. (1994).

### 3.6 Modelling procedure

Using the choice set defined in Sect. 3.4, the modelling procedure follows an iterative approach where the first model only has the alternative specific constant in the utility function. In subsequent iterations, observed heterogeneity was added using an alternative specific coefficient for each alternative. If a coefficient showed either a high standard error or a high correlation, the following measures were taken to reduce the margin of uncertainty:

- 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 for these two coefficients.

As a result of these measures, most coefficients related to socio-economics for the third and fourth alternative were removed since these alternatives did not provide enough samples for the model to estimate. For individuals younger than 39 a high correlation was found and therefore a unique



coefficient was defined for alternatives 2 to 4. Finally, it is important to mention that the socio-economic characteristics do not vary across alternatives since these are individual specific. As a consequence, in order to avoid linear dependency, we omit the coefficient for the first third and fourth alternative to create a robust baseline. As a result, the systematic utility function  $V_{int}$  of the model is shown in Eq. 6

$$\begin{aligned}
 V_1 &= ASC_1 + duration_1 + train\ in\ out \\
 &\quad + train\ within + transfer \\
 V_2 &= ASC_2 + duration_2 + age_2 + bachelor\ or\ higher \\
 &\quad + low\ income + male + \\
 &\quad train\ in\ out + train\ within + transfer \\
 V_3 &= ASC_3 + duration_3 + age_3 + train\ in\ out \\
 &\quad + train\ within + transfer \\
 V_4 &= ASC_4 + duration_4 + age_4 + train\ in\ out \\
 &\quad + train\ within + transfer
 \end{aligned} \tag{6}$$

## 4 Model Results

Tables 3 and 4 show the results of the final multinomial model and panel model. Our alternatives are always ordered based on travel duration where station 1 is the station with the shortest travel duration and station 4 is the station with the longest travel duration. We will start by discussing the results and interpretations of the MNL model and will compare this to the results of the panel logit model. The final MNL model has an adjusted rho-square of 0.450. Since the travel duration is used as an explanatory variable, but was also used to define the ordering of the alternatives, we need to interpret the travel duration together with the alternative specific constants. More specifically, the travel duration

variables are interpreted as a correction to the alternative specific constants in the model.

### 4.1 Results trip characteristics

The cross-variable for parking at a train station originating from Amsterdam and going outside of Amsterdam show significant results at a 99% level. The positive sign in the parameter shows that it is likely that people living in Amsterdam with a destination outside of Amsterdam cycle to the train station. A possible explanation could be that these people use the train to travel outside of Amsterdam, and prefer to cycle directly to the train station instead of parking at a closer metro station. When we consider the other cross-variable we see significance at a 99% confidence level for people travelling within Amsterdam indicating that even for trips within Amsterdam the train station is likely to be chosen as a parking location. However, the impact of this dummy variable is lower compared to the other cross-variable. In the literature review of Young and Blainey (2018b) it was found that the amount of transfers required for an individual is an important indicator for station choice parking. Our results show that the amount of transfers is not significant for any of the alternatives at a 90% confidence level. Even after taking measures to reduce the margin of uncertainty, as described in Sect. 3.6, our model was not able to find significant results at a 90% confidence level. A possible explanation could be that the influence of our transfers variable may be different for metro parking and train parking. If this difference is substantial it might explain the high standard error in our model and therefore also explain the lower t-test.

### 4.2 Results socio-economics

The coefficients for the socio-economic characteristics are not as impactful as the duration or cross-variable but they

**Table 3** Model results

Name	Constants		Duration		age<39		education		income <2500	
	Value	t-test	Value	t-test	Value	t-test	Value	t-test	Value	t-test
Station 2 constant	-0.686	-6.55***	-4.06	-10.76***	-4.00	-10.11***	-3.88	-9.67***	-4.12	-9.68***
Station 3 constant	-1.20	-6.68***	-4.52	-7.88***	-4.45	-7.73***	-4.46	-7.73***	-4.47	-7.76***
Station 4 constant	-0.21	-1.27	-6.01	-7.47***	-6.15	-7.45***	-6.19	-7.48***	-6.22	-7.49***
Station 1 duration			-4.44	-5.69***	-5.20	-6.26***	-5.17	-6.20***	-5.37	-6.32***
Station 2 duration			2.19	4.35***	2.21	4.19***	2.26	4.23***	2.34	4.28***
Station 3 duration			2.09	3.06***	2.06	3.02***	2.09	3.04***	2.01	2.95***
Station 4 duration			5.17	5.19***	5.46	5.32***	5.5	5.35***	5.51	5.34***
Age<39					-1.01	-4.21***	-0.952	-3.90***	-1.07	-4.28***
Station 2 education							-0.363	-1.41*	-0.529	-1.95*
Station 2 income									0.591	2.10**

\*\*\*, \*\*, \* indicate significance at a 99%, 95%, and 90% confidence level

**Table 4** Model results cont.

Name	male		crossvariable		transfers		panel	
	Value	t-test	Value	t-test	Value	t-test	Value	t-test
Station 2 constant	- 4.23	- 9.81***	- 4.63	- 9.25***	- 4.24	- 7.08***	- 8.18	- 5.33***
Station 3 constant	- 4.48	- 7.79***	- 4.55	- 7.07***	- 4.16	- 5.71***	- 7.1	- 5.81***
Station 4 constant	- 6.27	- 7.49***	- 7.52	- 6.30***	- 7.07	- 5.64***	- 12.4	- 4.10***
Station 1 duration	- 5.38	- 6.32***	- 5.66	- 5.82***	- 5.57	- 5.67***	- 10.20	- 5.12***
Station 2 duration	2.24	4.13***	2.48	4.15***	2.51	4.17***	4.31	4.01***
Station 3 duration	2.08	3.04***	1.88	2.62***	1.81	2.46**	2.25	2.24**
Station 4 duration	5.61	5.37***	6.30	4.43***	6.18	4.28***	10.40	3.10***
Age<39	- 1.14	- 4.50***	- 1.28	- 4.50***	- 1.32	- 4.59***	- 1.20	- 2.20**
Station 2 education	- 0.458	- 1.69*	- 0.527	- 1.71*	- 0.486	- 1.57	- 0.413	- 0.61
Station 2 income	0.592	2.10**	0.749	2.39**	0.791	2.50**	0.118	0.16
Station 2 male	0.499	1.88*	0.291	0.95	0.274	0.87	0.224	0.36
train in-out			4.42	7.72***	4.26	7.45***	5.63	6.35***
train in-in			1.59	4.05***	1.52	3.87***	1.71	3.40***
Station 1 transfer					0.433	1.58	0.207	0.57
Station 2 transfer					0.173	0.93	0.348	1.15
Station 3 transfer					0.385	1.39	0.392	0.99
Station 4 transfer					0.375	1.09	0.879	1.31
SIGMA 1							1.80	4.34***
SIGMA 2							1.35	2.37**
SIGMA 4							1.92	1.52

\*\*\*, \*\*, \* indicate significance at a 99%, 95%, and 90% confidence level

still provide useful information. For many of these variables few observations were available to make any statements about travelling to station 3 or 4. Therefore, we decided to only estimate a coefficient for the second closest station, using the other alternatives as a baseline. First of all, people with a net income below 2500 a month are more likely to travel to the second station compared to the other stations at a 95% confidence level. The male and education variables were of no importance to the model. Finally, we attempted multiple specifications with the age variable but due to high correlations and low significance of other age groups, only people younger than 39 years old are added as an explanatory. The coefficient shows that people younger than 39 are less likely to travel to the second station at a 95% confidence level.

### 4.3 Panel model

Finally we will compare the results of the MNL model to the Panel model. First we observe that the rho-square improved from 0.450 to 0.491. Secondly, the socio-economic characteristics lost some of their explanatory power since their t-tests and values are closer to zero compared to the MNL model. This is a result from the fact that the model now recognizes individuals instead of only trips. Thirdly, the impacts of the coefficients for all variables other than the

socio-economics are higher compared to the MNL model. The sigma coefficients are significant at a 95% confidence level for station 1 and 2. This indicates that there is variation among the individuals that choose to park their bicycle at station 1 and 2. Additionally, the impact of this variance increases as people travel to stations with the shortest travel duration.

## 5 Discussion

If we look at the alternative specific constants of the final model we can see that the coefficients of alternative 2 are more negative compared to alternative 3. This would mean that if no explanatory variables were added to the model the probability of choosing station 3 is higher than station 2. However, the final model shows that a lot of explanatory power can be found in the station with the second shortest travel duration. A possible explanation could be that the closest station may involve an additional transfer required for the individual to reach their final destination. This might be the case when the individual travels to the closest train station but needs to get on a specific intercity train that does not stop at the closest station. Moreover, an individual might not be interested in travelling by either metro or train and would thus not consider this station even if it is the closest one.

The cross-variable shows an interesting result that trips originating from Amsterdam are less likely to travel to a metro station but rather go to a train station directly. This shows that people favor the bicycle over the metro in most cases, which might be explained by the fact that people do not want to spend extra costs for the metro. Additionally, it is also interesting to see that even for trips within Amsterdam, train remains an important variable. This is probably the result of having many more bicycle-train trips compared to bicycle-metro trips. Surprisingly, the amount of transfers did not provide any insights in station choice behavior as was observed in previous studies. A possible explanation is that a transfer might be valued differently for individuals parking at a train or metro station.

The socio-economic variables shows that potentially some explanatory power can be obtained from these variables. Most of the results for our socio-economics are not significant, but our research was limited in two aspects. First of all, in our survey limited information was available about people their socio-economics since some people chose not to share some or all of their socio-economics with us. As a result, we had to segment our socio-economic variables into categories to ensure robustness of the model. Secondly, our socio-economic variables consisted mainly of categorical and not continuous values which diminishes the explanatory power and caused high correlation during the modelling phase. It would be interesting to see if the model would perform better if the age and income variables were continuous and the dataset was larger.

Adding a panel effect to the model improved the model performance since our socio-economic variables are individual related while the duration, transfers and the cross-variable are trip related. Although it would be most straightforward to apply a panel effect on the individual, it would also be interesting to apply a panel effect on one person's socio-economics characteristic. For example, it would be interesting to see if people of the same age behave similarly in their station choice. By putting a panel effect on the age variable it would be possible to test whether there is significant cohort variation among people of different ages.

Although the choice set generation as discussed in Young and Blainey (2018a) is interesting we decided to order alternatives based on travel duration for several reasons. First of all, the duration of the observed choice take into account the speed at which an individual is cycling. In a city like Amsterdam the distance between two stations might appear to be similar, but the actual duration of the trip can differ depending on the traffic involved in the trip. Secondly, by adding duration as an explanatory back in the model we can allow it to function as a correction to the alternative specific constant, since travel duration for recurring trips may vary while distances don't. Finally, in Young and Blainey (2018a) alternative choices were generated based on the ten closest

train stations by road distance. As a result, someone living in the city may have a lot of metro stations close to their origin but if their final destination involves a train leg journey they might prefer to travel all the way to the train station further away. Therefore, we decided to stick to our approach to generate alternatives

This research is an extension and an improvement to the work done in van Kampen et al. (2020). First of all, the data that was used allows us to track the travel behavior of an individual over a longer period of time compared to a specific date which is done in most surveys. Moreover, this allowed us to define a smaller study area making it possible to analyse our results on a municipal level instead of multiple provinces. Secondly, adding metro station as a choice and travel option allowed us to highlight the interaction of metro and train in the city environment. People living in Amsterdam definitely prefer to cycle to the train station, whereas the metro plays a more important role for people living outside of Amsterdam coming to the city. It would be interesting to see whether the same results could be found when bicycle is only considered as an egress mode. Furthermore, the use of Conway et al. (2017) in combination with option to travel to metro stations helped us create time-dependent, route-specific choice sets making our models more reliable. Finally, the use of tracking data allowed us to create a panel model which allows us to capture variational influences.

## 6 Conclusion

The results of the model shows that there is reason to believe that people do not always travel to the station with the shortest travel duration. Although the socio-economic characteristics were not as significant, it does illustrate the importance of these characteristics in station choice.

The results also show that people originating from Amsterdam prefer to park their bicycle at a train station and possibly do not like to travel with a metro at all. A possible explanation would be that people with destinations outside of Amsterdam use the train to leave Amsterdam mostly and therefore choose to park their bicycle directly at a train station instead of parking at a metro station. This shows that people prefer to cycle over using the metro in their mode choice behavior.

For future research it would be interesting to see to what extent the socio-economic characteristics impact the choice for a station. Although our models only found explanatory power in the two stations with the shortest travel duration, it would be interesting to see if any explanatory power could be found in stations with longer travel durations if more observations were available. Moreover, it would be interesting to see whether these results could be replicated for other



cities. Finally, the use of access factors and railway service factors is limited in this research which could potentially further improve the results that are presented here.

**Acknowledgements** This research has been conducted in the framework of the Impact Study North/Southline, funded in part by the Municipality of Amsterdam and the regional transportation authority of Amsterdam. We would gratefully like to acknowledge the Vrije Universiteit Amsterdam (VU) legal affairs office, IT department, and university library research data management team, as well as the IT and Facilities department of CWI for support in the design and implementation of the computing infrastructure. We would like to thank CWI/VU researcher Thomas Koch for generating transit schedule-specific alternative choice sets using the Conveyal R5 algorithm with OVapi GTFS. Hereby we also thank CWI/VU intern Manoj Ashvin Jayaraj for research assistance in previous work leading up to this paper and CWI/VU guest researcher Dr. Luk Knapien for useful discussions related to the research topic and contribution to the study design and data architecture. The authors declare that they do not have any conflicts of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Annan M, Altaf S, Bellemans T, Yasar AUH, Shakshuki E (2019) Last-mile travel and bicycle sharing system in small/medium sized cities: user's preferences investigation using hybrid choice model. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-018-0849-5>
- Ben-Akiva M, Lerman S (1985) *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge
- Bierlaire M (2003) Biogeme: A free package for the estimation of discrete choice models. <http://biogeme.epfl.ch>
- Boukhechba M, Bouzouane A, Gaboury S, Gouin-Vallerand C, Giroux S, Bouchard B (2018) Prediction of next destinations from irregular patterns. *J Ambient Intell Human Comput* 9:1345–1357. <https://doi.org/10.1007/s12652-017-0519-z>
- CBS (2006) Fietsberaad publication 7. continuous and integral: the cycling policies of Groningen and other European cycling cities. <https://www.cycling-embassy.org.uk/document/continuous-and-integral-cycling-policies-groningen-and-other-european-cycling-cities-public>
- CBS (2015) Cbs factsheet 2015. <https://www.cbs.nl/en-gb/background/2015/27/factsheet-the-netherlands-cycling-country>
- CBS (2016) Transport and mobility report, 2016. <https://www.cbs.nl/en-gb/publication/2016/25/transport-and-mobility-2016>
- Chakour V, Eluru N (2014) Analyzing commuter train user behavior: a decision framework for access mode and station choice. *Transportation* <https://doi.org/10.1007/s11116-013-9509-y>
- Conway MW, Byrd A, van der Linden M (2017) Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks. *Transp Res Rec* 2653(1):45–53. <https://doi.org/10.3141/2653-06>
- Debrez G, Pels E, Rietveld P (2007) Choice of departure station by railway users. *Eur Transp* 37:78–92
- Givoni M, Rietveld P (2014) Do cities deserve more railway stations? the choice of a departure railway station in a multiple-station region. *J Transp Geogr* 36:89–97. <https://doi.org/10.1016/j.jtrangeo.2014.03.004>
- Halton JH (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer Math* 2:84–90
- Kager R, Bertolini L, Te Brömmelstroet M (2016) Characterisation of and reflections on the synergy of bicycles and public transport. *Transp Res Part A: Policy Pract* 85:208–219. <https://doi.org/10.1016/j.tra.2016.01.015>
- Lawrence CT, Zhou J, Tits A (1994) User's guide for cfsqp version 2.0: A c code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. <http://hdl.handle.net/1903/5496>
- Lee KH, Ko EJ (2014) Relationships between neighborhood environments and residents' bicycle mode choice: a case study of Seoul. *Int J Urban Sci* 18(3):383–395. <https://doi.org/10.1080/12265934.2014.957715>
- McFadden D (1973) *Conditional Logit Analysis of Qualitative Choice Behavior*. In *Frontiers in Econometrics*, Academic Press, New York
- Pucher J, Buehler R (2008) Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transp Rev* 28(4):495–528. <https://doi.org/10.1080/01441640701806612>
- Train K (1999) *Halton sequence for mixed logit*. University of California, Berkeley
- Train KE (2009) *Discrete choice methods with simulation*. Cambridge University Press, Cambridge
- van Kampen J, Jayaraj MA, Pauwels E, van der Mei R, Dugundji ER (2020) Understanding the relation between distance and train station choice behavior of cyclists in the western region of the Netherlands. *Procedia Computer Science* 170:98 – 106. <https://doi.org/10.1016/j.procs.2020.03.013>, the 11th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops
- Young M, Blainey S (2018) Development of railway station choice models to improve the representation of station catchments in rail demand models. *Transp Plan Technol* 41(1):80–103. <https://doi.org/10.1080/03081060.2018.1403745>
- Young M, Blainey S (2018b) Railway station choice modelling: a review of methods and evidence. *Transp Rev* 38(2):232–251. <https://doi.org/10.1080/01441647.2017.1326537>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.