European Transport \ Trasporti Europei n. 37 (2007): 78-92

Choice of departure station by railway users

G. Debrezion [1*], E. Pels [1], P. Rietveld [1]

[1] *VU University Amstedam, Department of Spatial Economics*
De Boelelaan 1105, 1081 HV Amsterdam
The Netherlands

Abstract

This paper applies a multinomial logit model to the choice of a departure railway station by Dutch railway passengers. This is a relevant theme since about 50% of Dutch railway passengers do not travel via the nearest railway station. The passengers' choices for departure stations are aggregated at the four digit postal code area level. We applied three functional forms for the underlying systematic utility of a station, namely a linear effect of attributes, cross effect of distance and frequency of service, and a translog formulation on distance and frequency of train services. With 3,498 post code areas and 360 railway stations our analysis found consistent effect sizes for distance, frequency of service, intercity status of the station and the presence of park-and-ride facility on the choice of departure station. The effect of distance on the choice of a departure station declines smoothly. The effect of frequency of service is relatively small compared to the effect of distance. A frequency of service increase by a hundred trains per day is equivalent to being 600 m closer to the station. The Intercity status of the station plays the biggest role in the choice of departure station. It has an equivalent effect of a change in 2 km distance or about a frequency of service of 300 trains per day. In addition, the presence of park-and-ride facility in the station poses a sizable effect in the departure station choice. In most cases its effect reaches about 35% of the intercity status effect.

Keywords: Railway station choice; Logit model.

1. Introduction

Railway transport constitutes a sizable share of the daily travel made by the Dutch travellers. The figures from the central bureau of statistics (CBS) in 2002 reveal that railway transportation accounts for about 8% of the over all passenger kilometres. This figure is among the highest shares of railway transport in Europe and the world. In the US the overall public transit share (which includes railway and bus services) is about 2% (bureau of transportation statistics 2005). The modal split of passenger kilometres shares for the fifteen members of the European Union are given also in Table 1. Following Austria and France, the Netherlands has the third highest market share for

[*] Corresponding author: Ghebreegziabiher Debrezion (gdebrezion @feweb.vu.nl)

rail transport. Railway transport is, therefore, expected to be an important travel alternative for the Dutch households.

Table 1: Modal split by country for passenger transport: EU-15 (5 modes) year 2002.

	Passenger kilometer (percentage)				
	CAR	BUS	RAILWAY	TRAM & METRO	AIR
BELGIUM	79.8	9.9	6.0	0.7	3.6
DENMARK	74.3	11.1	6.8	0.0	7.8
GERMANY	78.8	8.6	7.8	0.9	3.9
GREECE	65.9	17	1.4	1.0	14.6
SPAIN	71.2	10.6	4.5	1.2	12.5
FRANCE	83.1	4.5	8.2	1.2	3.0
IRLAND	72.8	12.4	3.2	0.0	11.5
ITALY	80.2	11	5.3	0.6	3.0
LUXEMBOURG	74.7	12.8	5.1	0.0	7.4
NETHERLANDS	81.5	4.1	8.1	0.8	5.5
AUSTRIA	70.7	13.6	8.4	2.8	4.5
PORTUGAL	79.7	8.3	3.1	0.5	8.3
FINLAND	77.7	aw10.3	4.4	0.7	7.0
SWEDEN	74	8.0	7.2	1.8	9.0
UNITED KINGDOM	80.9	5.9	5.1	1.1	7.1

Adapted from EU energy and transport in figures: statistical pocket book 2004.

Once one has decided to travel by train, the logical next question that follows is which station to use as a departure point. The decision on this issue is expected to be based on the assessment of relevant access and station features. Easily accessible railway stations are expected to be chosen more often as departure stations than similar stations that are less accessible. Moreover, the choice of a departure station also depends on the quality of the station itself. The quality of railway station is generally explained by the quality of rail and supplementary services provided by or at the station. The frequency of train services, network connectivity and coverage are some examples of the rail service. The presence of other supplementary facilities such as availability of parking spaces, the park-and-ride possibility, the availability of bike stand and safes also boosts the attractiveness of a station as a departure station.

The revealed choice data for departure stations shows that in the Netherlands about 47% of the cases passengers choose a departure station other than the nearest station to their place of residence. This reveals that there is a need to understand the decision process for a departure station. Understanding the valuation and decision mechanism leading to the choice of a particular railway station as an access (departure) station has several practical implications. In the first place, it enables us to define the catchment area (or market area) for stations. This enhances the predictions of travel demand at the station level. Based on the sensitivity of travellers towards access and station features, it gives a basis to the station operator to increase travellers' turnover. In addition, it can be used as a basis for site selection for new line development or extension planning for existing lines, as well as parking, park and ride facility and feeder public transport operation planning.

In this paper we analyze the choice for departure railway station made by Dutch railway travellers. This will in turn be used in the calculation for a general railway accessibility index for zones where people live. In most real estate price studies railway station accessibility is just given by the distance to the nearest of the railway station from the property in question. However, railway station accessibility encompasses all aspects that are involved in the choice process for a departure station. The general accessibility index for areas based on the choice analysis is expected to perform better than the simple distance proxy for the accessibility to the railway station.

This paper is organized as follows. Section 2 briefly reviews the literature in the area. Section 3 gives the specification of the multinomial logit model that we apply in our choice analysis. This is followed by the description of the data used in our analysis. Section 5 presents the methodology for our analysis. Various specifications for the utility function are considered. Section 6 gives the estimation results followed, by the discussion of these results. Section 7 concludes the analysis.

2. Literature review

The literature in this area is generally scarce. One of the early rail transit station choice models was developed by Kastrenakes (Kastrenakes 1988) in an effort to prepare a basis for forecasting rail ridership in New Jersey area. With origin-destination pair data, Kastrenakes analyzed the choice for a departure station based on access time required to reach the station, frequency of service at the boarding station, whether the boarding station is located in the locality of the residence of the passenger, and generalized cost of the train trip between the departure station and destination station (Kastrenakes 1988). The study found, as expected, positive effects for frequency of service and location of the station in the locality of the passenger's residential area on the choice of departure station choice. Similarly, negative effects were found for access time and the generalized cost of the rail trip as expected. In another study, Wardman and Whelan (1999) studied railway station choice for the inter-urban trips to London. This study was done in relation to parking attractiveness for station choice. It is indicated that availability of parking area in a station and other station facilities are important features for station choice (Wardman and Whelan 1999).

Some studies on this theme have also incorporated access mode choice in a nested structure (Fan et al. 1993; Wardman and Whelan 1999; Davidson and Yang 1999). Generally speaking, the preferred model structure has the access mode choice in the upper level. Fan el al. (1993) included several variables for the lower level transit station choice part. Travel time including access and in vehicle time, fare, peak hour frequency of trains, and the number of parking places were among the included variables. Expected positive signed coefficients for frequency of service and parking and negative signed coefficients for travel time and fare were found. Wardman and Whelan (1999) on the other hand compared the access mode-station choice for business and leisure travels. They found the value of time is highest for business trips and lower for leisure trips. Other variables included were journey time, journey headway, facilities at station and parking availability. They all show expected signs and significant effects on the choice of the departure station.

Choice analysis of this form has been popular in the literature on the choice for a departure airport (Ashford and Bencheman 1987; Hess and Polak 2004; Pels et al. 2001;

Pels et al. 2003; Basar and Bhat 2004). Fares (airport tax), access time, frequency of service, and other facilities are important features used in airport choice. Some studies also include time series historic data in the choice feature of those commuters who tend to use an airport that they have previously used. The analyses on departure airport choice have some relevance to the railway station choice. Most of the time we do not see fare difference between railway stations, so the fare does not play a relevant role in the choice among stations. However, access features like access time and access cost obviously are relevant for the railway station analysis. The frequency of service, as indicated by the number of trains leaving the station per given time and/or the number of destinations served directly from the station, plays an important role in the station choice analysis. The same holds for the nature of the station and facilities at the station. Obviously, international and intercity stations are expected to enjoy higher choice probabilities compared to express or stop train stations[1]. Stations with better public and passenger related facilities are also expected to be attractive compared to stations with lesser or no facilities. As the access time increases, the attractiveness of the station for departure declines.

3. Multinomial Logit model for a station choice

The choice made on which departure station to use forms an important choice decision for the household. In the short run the nature of a station is not expected to vary. Thus, generally speaking, the households' choice remains unchanged. However, in the long run a railway station can undergo major changes that can cause the travellers to look for different departure stations. At the same time households can have multiple destinations and multiple trip purposes. Thus, it is natural to observe different railway stations to be chosen as departure stations for different trip purposes and destination combinations. However, our data does not include information on trip purpose and destination. The analysis in this paper, therefore, will be limited to the aggregate choice of departure stations made by households in a post code area without looking at the purpose of the trip and destination.

We assume that any household's choice for a departure station is based on the assessment of an underlying utility function. That means that depending on the nature of the origin of the trip, each railway station provides certain utility level. These utilities may differ not only on the relative location from the origin, but also because stations differ from each other in their nature. As rational choice makers, travellers select a departure station among a set of feasible alternatives so that the utility is maximized. Different features can enter the utility function. The access distance, i. e. from the origin to the railway station, is an important feature in the departure station choice model. In addition, the service levels provided at the station determined by the frequency of trains, the number of destinations that can be reached from the station and other station facilities like parking and bicycle stands are potentially relevant station characteristics

[1] In the Netherlands there are four types of railway services namely: the all station rail services called stop train; semi fast also called express rail services which call at main and medium cities, inter-city service rail services that only call at main cities, and international trains that only stop at a very limited number of stations.

in the station choice utility function. The utility function can assume several functional forms. Later in our model specification we will give three functional forms for the utility, namely a simple linear additive, linear additive utility with cross distance frequency of service product and transcendental logarithmic function. For the purpose of outlining the multinomial logit model, in this section, we specify the two general components of the total utility function as follows:

$$U_j = V_j + \varepsilon_j \qquad (1)$$

Where U_j is the total utility level of alternative j; V_j represents the systematic component of the utility for alternative j and ε_j is the stochastic component of the utility for alternative j.

The probability that alternative j is chosen is given by the probability that the utility corresponding to alternative j exceeds the utility levels of other alternatives. If we assume that ε_j is independently and identically distributed and follows the Gumbel extreme value distribution (see Ben-Akiva and Lerman 1985 for a detailed description of Gumbel extreme value distribution) the probability of choosing station j is:

$$P(j) = P(U_j > U_k) \text{ For all } k \neq j \qquad (2)$$

McFadden (1973) has shown that if the distribution of the stochastic component of the utility (ε) follows an extreme value distribution function, the choice situation results in multinomial logit model. The multinomial logit model is a family of discrete choice model, which allows a choice situation with multiple alternatives. The probability of choosing alternative j is thus given by the multinomial logit model given by (3):

$$P(j) = \frac{\exp(V_j)}{\sum_{k \in J} \exp(V_k)} \qquad (3)$$

An important property of the multinomial logit model is the independence from irrelevant alternatives (IIA). That means that the ratio of the station choice probabilities of two stations is not affected by the systematic utilities of any other alternatives. This property holds true for the departure station choice situation analyses in this paper.

4. Data

The data used in our analysis were acquired from the Dutch national Railway Company (NS). We employed two data sets: post code area related data and railway station related data. The choice outcome for the departure station choice is aggregated at the post code area level. The Netherlands is composed of 4,004 post code areas. However, because of the data incompleteness our final analysis is based on 3,498 post-code areas. This accounts for 87% of the country. For each of the post code areas a set of three most frequently chosen stations is identified. The overall number of stations covered in the analysis is 367. Due to the aggregation process, the choices for the departure station are explained by the share of each of the three stations receives at the post code area.

GIS information on the location of the centroid for post code areas and the railway stations was used to calculate the distance between the centroid of the post-code centre and the railway stations in the choice set. The distance assumes a Euclidean measure. The data set combines the shares of the choice each of the three stations receive in the postcode area, and the railway station features including the access distance.

Descriptive of Station and Accessibility Characteristics

Railway station accessibility is generally explained by two factors namely, the ease of reaching the stations and the service levels provided at the stations. The ease of reaching the stations is linked to the distance between the departure point (the centroid of the post code area in this case) and the railway station. The level of services provided at the stations is related to the frequency of trains leaving the station per time period and network connectivity as determined by the number of destinations that can be reached without a transfer. In addition, it includes facilities that supplement railway transport. Table 2 below, gives the descriptive statistics of the railway station characteristics and the accessibility indicators for the post code areas. For the purpose of showing the variation, in Table 2 below, we only give the distance to the first most frequently chosen station from the post code area. In addition, Table 2 gives the station features. Included are the indicators of railway service, type of station, and facilities in the station.

Table 2: descriptive statistics for the railway station characteristics (2001/2002).

	Number of stations/ post code areas	*Min*	*Max*	*Mean*	*Std. Deviation*
Rail service					
Frequency of trains per day		18	788	113	103
Destinations reached without a transfer		1	114	16	14
Station type					
Inter-city stations	64			0.18	
Station Facilities (dummies)					
Train taxi	109			0.30	
Bicycle stand	96			0.27	
Bicycle safe	264			0.74	
Bicycle rent	114			0.31	
Park-and-ride	49			0.14	
Parking	326			0.91	
Taxi	163			0.45	
Car rent	1			0.00	
Luggage deposit	64			0.18	
International connection	22			0.06	
Accessibility from Post code areas					
Distance to the most frequently chosen station (m)	3498	95	34250	6756	6129

5. Methodology

Our analysis identifies several factors which have an impact on the choice for a departure station. One of these factors is the distance from the postcode area to the railway station. We apply categories of distance classes in our analysis. To see the smoothness of the effect we use a 500 meters range categories, except in the two inner circle categories of the station, which are 250 meters each. Thus we have 21 categories of distances up to 10,000 meters, where the category above 10,000 meters is taken as reference group. Each of these categories is represented by dummy variable. We identify a number of station facilities that have utility bearing nature in the general sense. However, these facilities can be departure station related or destination station related in nature from the passenger's point of view. For instance a car rental facility is a typical destination station feature. A parking lot on the other hand is a departure station feature. As our main focus here is the choice of a departure railway station, we stick to the departure station related features of the station.

Departure related features of a station that can be used in our analysis include the frequency of train, the number of direct destinations, the availability of a bicycle stand, the presence of a park-and-ride facility, the availability of parking and the type of railway station, and inter city status. Frequency of service and number of direct destinations are expected to be highly correlated; we, therefore, only include frequency in our analysis. Bicycle stands are not expected to affect the choice of departure station, since in general passengers will find a place for their bicycle anyhow. In addition, we include the presence of park-and-ride facility, and the intercity status of the station. All the features are expected to influence the choice of a departure station positively. However the effect of intercity status of a station can differ regionally. Thus, we use it as specific to the provinces of the country. The Netherlands is made up of 12 provinces. The province of Utrecht, North Holland and South Holland constitute the most urbanized area of the country. This region is mostly called as the Randstad (rim city), since it constitutes an extended chain of cities. The effect of the inter city status of a station on the choice of a departure station is expected to be less in these regions as compared to the effect of inter city status of a station for the choice of a departure station in the more peripheral provinces.

Specification of the systematic utility function

For comparison purposes we give three functional specifications for the systematic utility of a station: the linear additive, linear additive utility with cross distance frequency of service product and transcendental logarithmic formulation. The first is aimed at capturing the separate effects of distance and other station features. The second is aimed at determining the effect of train service frequency on the choice of a departure station at different distance segments. The transcendental logarithmic station utility function is aimed at determining the general smoothed utility function for the station choice. They are given respectively in equations 4, 5, and 6.

$$V_j = \sum_{c=1}^{21} \beta_c * Dcateg_{jc} + B_f * freq_j + B_{P\_R} * P\,\&\,R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \quad (4)$$

$$V_j = \sum_{c=1}^{21} \beta_{cfreq} * Dcateg_{jc} \times freq_j + B_{P\_R} * P\,\&\,R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \qquad (5)$$

$$\begin{aligned} V_j = {} & \beta_{dist} \ln(dist_j) + \beta_{distsq} (\ln dist_j)^2 + \beta_{freq} \ln(freqT) + \beta_{freqsq} (\ln freqT_j)^2 \\ & + \beta_{distfreq} (\ln dist \times \ln freq_j) + \beta_{P\&R} P\,\&\,R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \end{aligned} \qquad (6)$$

Where $Dcateg_{jc}=1$ if station j is located in the distance category c from the centre of the post code area and zero otherwise; dist is the distance between the centroid of the post code area and the railway station in continuous measure; $freq_j$ is the frequency of trains at station j; $P\,\&\,R_j$ is an indicator for the presence of park-and-ride facility at station j; IC_j is an indicator for an intercity status of station j; $Prov_{jp}$ is a dummy variable for the province in which station j is located. It takes the value 1 if station j is located in province p, and 0 otherwise.

6. Estimation results

1. Linear additive utility function: Piecewise distance measure

Here we formulate the utility function for the railway stations in a linear additive way. Distance is measured in a piecewise fashion. This gives a detailed effect of distance compared to other continuous treatment of distance, because the area over which averaging is made is quite limited. In addition we include the frequency of trains at the stations, the availability of park-and-ride and the IC status of the station into the utility function. The systematic utility specification is given by equation 4 above. The estimation output for the multinomial logit model for the station choice is given below in Table 3. All coefficients are significant, with a dominant role for the distance effect. The effect of distance as expected has a positive sign. The values of the coefficients for the distance categories are relative to the zero reference value for areas beyond 10 km.

The value of the pseudo R-square (given as $1-\log L_{\text{estimated model}} / \log L_{\text{base model}}$), used as a measure of goodness of fit, shows the model has a good prediction power. According to the empirical relationship drawn (Domencich and McFadden 1975) against the R-square of the linear models, the R-square of the above logit model is equivalent to an R-square close to 0.8, in the linear regression models.

From the value of the coefficients we observe a smooth decline in the effect size with distance. That means that the closer the postcode area is to the station, the higher the probability of choosing that station as a departure station. The value found for the frequency of service effect is relatively small compared to the effect of distance. A frequency of service increase by a hundred trains per day is equivalent to being about 600 meters closer to the station.

Table 3: Multinomial logit model estimation for station choice.

Variable	*Coefficient*	*t- value*	*p-value*
DIST250	7.379	7.360	0.000
DIST250_500	6.852	16.190	0.000
DIST500_1000	6.328	27.154	0.000
DIST1000_1500	5.734	27.195	0.000
DIST1500_2000	5.147	26.058	0.000
DIST2000_2500	4.797	25.488	0.000
DIST2500_3000	4.150	22.916	0.000
DIST3000_3500	3.723	21.676	0.000
DIST3500_4000	3.411	19.758	0.000
DIST4000_4500	2.940	17.630	0.000
DIST4500_5000	2.765	16.837	0.000
DIST5000_5500	2.471	15.274	0.000
DIST5500_6000	2.282	13.505	0.000
DIST6000_6500	1.953	11.354	0.000
DIST6500_7000	1.902	11.776	0.000
DIST7000_7500	1.748	9.901	0.000
DIST7500_8000	1.626	9.273	0.000
DIST8000_8500	1.487	8.635	0.000
DIST8500_9000	1.143	6.347	0.000
DIST9000_9500	1.237	6.682	0.000
DIST9500_10000	0.955	4.873	0.000
Frequency	0.004	12.633	0.000
Park & Ride	0.419	6.226	0.000
IC-Groningen	1.947	7.875	0.000
IC-Friesland	1.785	8.318	0.000
IC-Drenthe	1.510	3.841	0.000
IC-Overijssel	1.101	5.038	0.000
IC-Gelderland	1.052	5.751	0.000
IC-Utrecht	-0.272	-0.970	0.332
IC-North Holland	0.724	3.427	0.001
IC-Zuid-Holland	0.534	3.041	0.002
IC-Zeeland	3.235	5.436	0.000
IC-Noord-Brabant	0.971	6.340	0.000
IC-Limburg	1.769	9.173	0.000
IC-Flevoland	-3.054	-3.446	0.001
Number of observations 3396			
Log likelihood function -2312.613			
R-sqrd 0.38014 RsqAdj 0.37693			

The province specific coefficients for the intercity status of a station have generally positive effect with the exception of the province of Flevoland. A positive coefficient shows that intercity status has a positive effect on the choice of a departure railway station compared to a non intercity station. As expected the coefficients for the provinces making up the Randstad area, have lover value. For example, the coefficient for the province of Utrecht is found to be insignificant. This indicates that an intercity status of a station does not make a significant difference in the choice for a departure station in the area as compared to the non intercity stations. This finding makes much sense because this province is in the heart of the Netherlands and the most accessible one. The provinces with the highest value for the intercity status include Zeeland, Groningen and Friesland. These are the peripheral provinces of the country. Generally the intercity status of a railway station has a big effect on the choice for a departure station. On average, the effect is equivalent to the frequency effect of about 300 trains per day. Further, the availability of a park-and-ride facility has a sizable impact on the choice of a departure station. It has an equivalent effect of about 105 trains per day for the frequency of service level.

2. Linear additive utility function: piecewise distance and frequency of trains cross product

This model is a slightly adapted form of the above formulation. The formulation is aimed at assessing the effect of frequency of trains on the choice of a departure station at the different distance categories. Thus, distance categories are cross multiplied with frequency. This approach gives some flexibility in allowing the effect of frequency on the choice of a station to differentiate across distance categories. The systematic utility specification is given by equation 6 above. The estimation results for this specification are given in Table 4 below.

Table 4: Cross product of distance categories and frequency of trains.

Variable	*Coefficient*	*t-value*	*p-value*
FRQ250	0.0717	6.183	0.000
FRQ250_500	0.0549	13.171	0.000
FRQ500_1000	0.0441	17.907	0.000
FRQ1000_1500	0.0326	17.739	0.000
FRQ1500_2000	0.0249	16.651	0.000
FRQ2000_2500	0.0196	16.781	0.000
FRQ2500_3000	0.0167	16.195	0.000
FRQ3000_3500	0.0146	16.102	0.000
FRQ3500_4000	0.0103	12.564	0.000
FRQ4000_4500	0.0087	12.227	0.000
FRQ4500_5000	0.0069	9.411	0.000
FRQ5000_5500	0.0065	9.053	0.000
FRQ5500_6000	0.0053	7.599	0.000
FRQ6000_6500	0.0046	5.798	0.000
FRQ6500_7000	0.0043	6.602	0.000
FRQ7000_7500	0.0040	5.031	0.000
FRQ7500_8000	0.0036	5.691	0.000
FRQ8000_8500	0.0032	5.069	0.000
FRQ8500_9000	0.0026	3.765	0.000
FRQ9000_9500	0.0026	3.550	0.000
FRQ9500_10000	0.0016	1.825	0.068
Park & ride	0.2509	4.091	0.000
IC-Groningen	0.8691	4.304	0.000
IC-Friesland	1.0502	5.921	0.000
IC-Drenthe	1.4540	4.616	0.000
IC-Overijssel	0.3014	1.695	0.090
IC-Gelderland	0.3964	2.540	0.011
IC-Utrecht	-0.2983	-1.118	0.264
IC-North Holland	0.4322	2.138	0.033
IC-Zuid-Holland	0.4792	3.263	0.001
IC-Zeeland	2.7405	5.295	0.000
IC-Noord-Brabant	0.6605	5.071	0.000
IC-Limburg	0.7897	4.931	0.000
IC-Flevoland	-1.9936	-2.284	0.022
Number of observations	**3396**		
Log likelihood function	**-2693.256**		
R-sqrd .27812 RsqAdj .27449			

From the table we see that the frequency of trains at a station has the expected effect sign and pattern on choice of a departure station across the distance categories. The effect size at each segment is given in comparison to the base category of frequency of service at distances of more than 10 kilometres. Frequency is given in trains per day. Similarly, the regional effect of intercity status of a station on the choice of a departure station is higher on the peripheral provinces as compared to its effect on the Randstad area.

In Figure 1 below we show the smoothness of distance and frequency effects on the choice of a departure station given in tables 3 and 4. On the Y-axis we have the size of the coefficients and on the X-axis we have the distance categories.

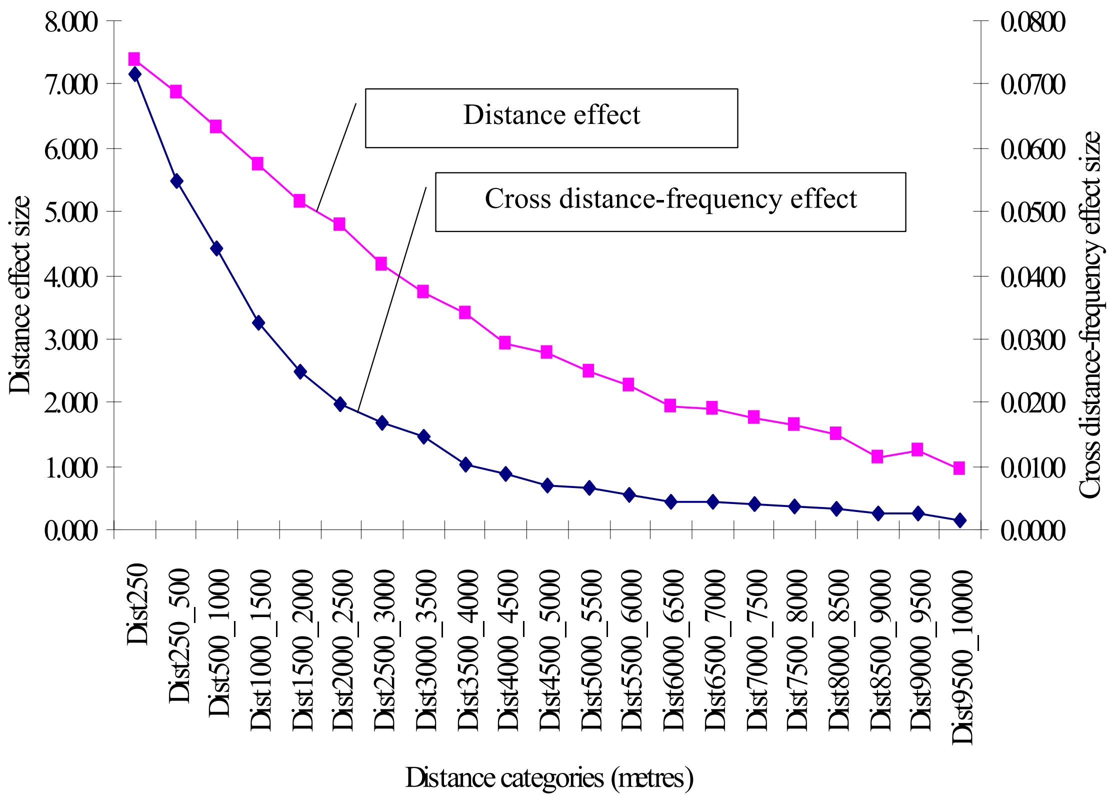

Figure 1: Distance and frequency effect on the departure station utility.

Comparing the two models we conclude that a tendency can be discerned that the effect of frequency improvements on the utility of a station is larger for residents that live nearby a station compared with living further away. However, it should be noted that the fit of the model without the cross effect is substantially higher than that of the model with the cross-effect (.29 versus .38). In order to shed more light on the issue we explore a third specification via the translog function which allows a more refined analysis of interaction effects.

3. Transcendental logarithmic formulation

In addition to the linear additive formulation of the utility function we used the transcendental logarithmic formulation. Here, distance, frequency of trains and park-and-ride variables are included. Distance is treated in the continuous form. The use of the translog function does not give a detailed treatment of the effect of distance as the stepwise treatment of distance does. The translog model is better in dealing with the effect of frequency, in particular the extent to which frequency effects are different for a traveller close to stations and a traveller further away. The systematic utility for this formulation is given by equation 7 above. The multinomial logit estimation output based on this specification is given below.

Table 5: Transcendental logarithmic utility function.

Variable	*Coefficient*	*t-value*	*p-value*
Log (distance)	1.341	2.305	0.021
Log (square distance)	-0.826	-1.224	0.221
Log (frequency)	-0.124	-1.954	0.051
Log (square frequency)	-0.192	-4.771	0.000
Log (distance) * log (frequency)	0.327	5.867	0.000
Park & ride	0.369	5.149	0.000
IC-Groningen	2.067	7.990	0.000
IC-Friesland	2.111	9.360	0.000
IC-Drenthe	1.743	4.027	0.000
IC-Overijssel	0.483	2.121	0.034
IC-Gelderland	0.584	2.968	0.003
IC-Utrecht	-0.507	-1.786	0.074
IC-North Holland	0.446	2.034	0.042
IC-Zuid-Holland	0.306	1.660	0.097
IC-Zeeland	3.260	5.305	0.000
IC-Noord-Brabant	0.611	3.814	0.000
IC-Limburg	1.362	6.627	0.000
IC-Flevoland	-2.623	-2.951	0.003
Number of observations	**3396**		
Log likelihood function	**-2195.868**		
R-sqrd .41144 RsqAdj	**.40987**		

Mapping the above output enables us to see the utility level for the different levels of frequency at different distances from the station. The graphical illustration is given in Figure 2 below. This confirms the relevance of the cross-product specification used above. Note also that the fit of model 3 is clearly better than that of models 1 and 2, even though the number of parameters is smaller. From the figure we see that the utility level for the stations smoothly declines with distance for all frequency levels. In addition, the curve corresponding to higher frequency of trains assumes a flatter pattern, indicating that the catchment area for the station expands with an increase in frequency of service at the station. The curves represent the combined utility contribution of the distance and frequency of service from the translog station utility formulation.

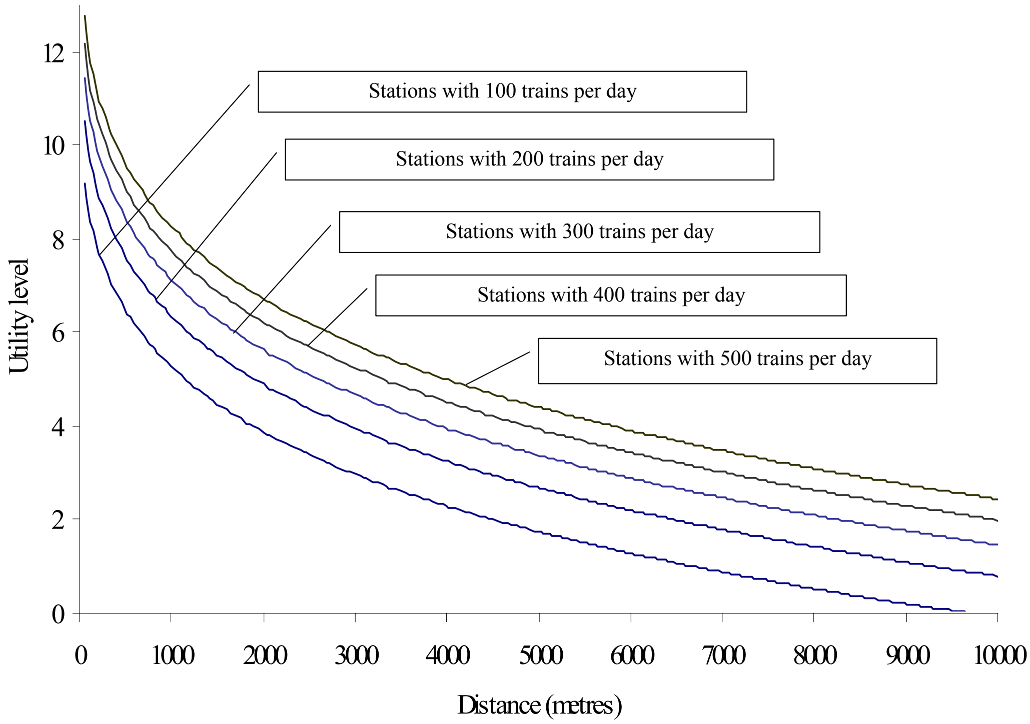

Figure 2: Railway station utility for different utility levels.

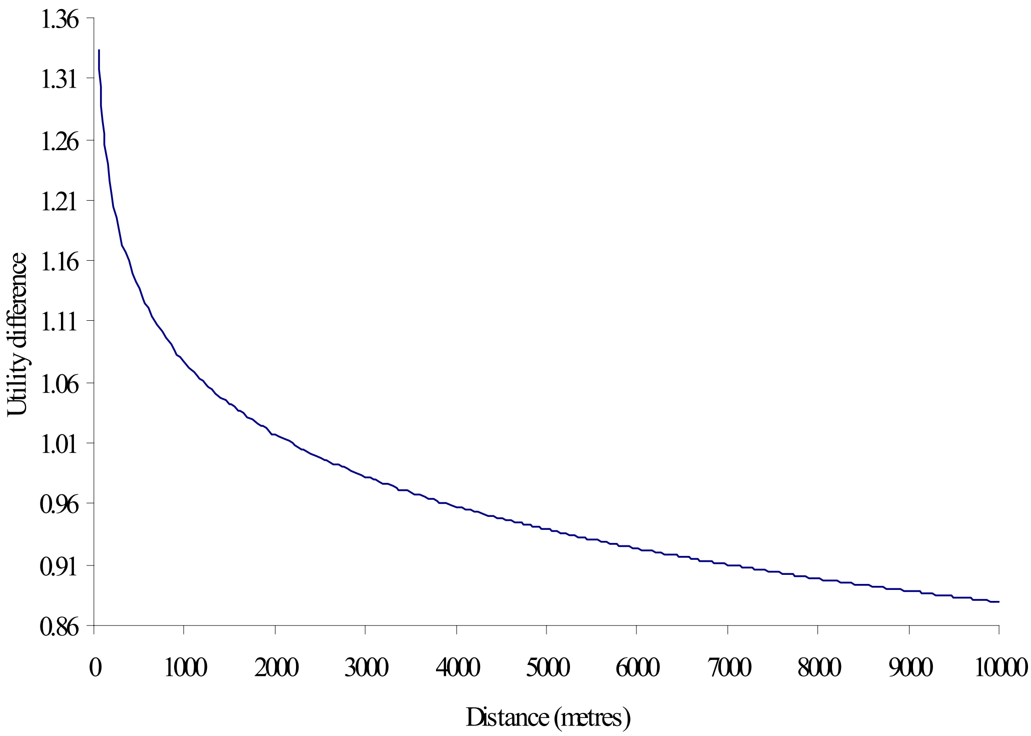

Figure 3: The effect of an increase in frequency from 100 to 200 trains per day on utility for a range of distances.

The lower curve corresponds to a frequency level of 100 trains per day, whereas the outer curve corresponds to a frequency of 500 trains per day. The frequency interval between the curves is fixed to 100 trains per day to facilitate comparison on the effect of an additional train. As one moves upward, the curves get closer. Thus, the graph reveals there is a diminishing effect of increasing frequency of trains on the utility level.

We can also show that the effect of frequency on utility declines with distance. Taking the difference in utility between frequency level 100 and 200 and mapping it with distance gives the curve in Figure 3 below.

7. Conclusion

This paper discusses the choice of a railway station as a departure station. Aggregated choices of households at the post code level were analyzed. For each post code area a set of three alternative stations was determined. We applied a multinomial logit model to determine the choice process. A number of access and station features were included in the station utility function. Distance between the centroid of the postcode area and the railway station was taken to include all access features. The station features considered in the utility function include frequency of service at the station, availability of park-and-ride facility and the intercity status of the railway station. Applying three model specifications for the utility function (linear additive, linear additive with the cross product of distance and frequency of service and the transcendental logarithmic), we found that all access and station features have significant expected effect sign. The probability of choosing a particular station declines with distance: a nearby station has a higher probability of being chosen. Moreover, the higher the frequency of service at a station, the higher is the probability that the station is chosen for departure. On the other hand, we found that the effect of frequency of a change in service declines with distance. The effect of frequency is higher for closer post code areas than for post code areas farther away. The intercity status of the station plays the biggest role in explaining the choice of a departure station. 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. In addition the presence of a park-and-ride facility in the station poses a sizable effect in the departure station choice. In most cases its effect reaches about 35% of the intercity status' effect.

Our model may however suffer from endogeneity problem. The choice of the station stems on the characteristics of the station and other facilities in the station. Most of the time the facilities at the station can be explained by the demand for travel the station generates, which in return corresponds with the choice variable. For example decisions for parking lots around stations are made on the basis of the number of commuters accessing the station. Endogeneity issues will be explored in a more advanced model.

These results may be further developed in various directions. One is to correct for possible endogeneity in the explanatory variables such as the park-and-ride variable. A second and more drastic development would be the use of more detailed data on access mode, which would allow for the estimation of a joint access mode and railway station choice model.

Acknowledgement:

The authors thank Mark van Hagen and his colleagues at the Dutch railway company (NS) for providing us with the necessary data and comments on an earlier draft. This paper was written in the context of three research programs: NWO-VEV, Transumo and Habiforum.

References:

Ashford, N. and Benchemam, M. (1987) "Passengers' choice of airport: an application of the multinomial logit model". *Transportation Research Record,* 1147, pp. 1-5.
Basar, G. and Bhat, C. (2004) "A Parameterized Consideration Set Model for Airport Choice". *Transportation Research Part B,* 38, pp. 889-904.
Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press Cambridge, MA.
Davidson, B. and Yang, L. (1999) "Modeling commuter rail station choice and access mode combinations". Presented at the *Transportation research Board* 1999 annual meeting Washington D.C.
Domencich, T. and McFadden, D. (1975) *Urban Travel Demand a behavioural Analysis.* Elsevier, New York and Oxford.
European Commission (2004) *EU Energy and Transport in Figure: statistical pocketbook 2004*
Fan, K., Miller, E. and Badoe, D. (1993) "Modeling Rail Access Mode and Station Choice". *Transportation Research Record,* 1413, pp. 49-59.
Hess, S. and Polak, J. W. (2005) "Mixed Logit Modelling of Airport choice in Multi-Airport Regions". Presented at the 45th annual congress of the European regional Science Association.
Kastrenakes, C. R. (1988) "Development of a Rail Station Choice Model for NJ Transit" *Transportation Research Record,* 1162, pp. 16-21.
Lythgoe, W. F., Wardman, M. and Toner, J. P. (2005) *Enhancing Rail Passenger Demand models to Examine Station Choice and Access to the Rail Network.* (draft).
McFadden, D. (1973) "Conditional Logit Analysis of Qualitative Choice Behavior". In: Zarembka, P. (ed.), *Frontiers in Econometrics*. Academic Press, New York.
Pels, E., Nijkamp, P. and Rietveld, P. (2003) "Access to and competition between airports: a case study for the San Francisco Bay area". *Transportation Research Part* A 37, pp. 71-83.
Wardman, M. and Whelan, G. A. (1999) "Using Geographical information systems to improve rail demand models". Final report to Engineering and Physical science Research council.