HIGH-SPEED RAIL'S IMPACT ON THE LOCATION OF OFFICE EMPLOYMENT WITHIN THE DUTCH RANDSTAD AREA

45th Congress of the European Regional Science Association

Amsterdam, August 23-27, 2005

Jasper Willigers
Utrecht University, Faculty of Geosciences
P.O. Box 80115, 3508 TC Utrecht, The Netherlands, j.willigers@geo.uu.nl

Han Floor
Utrecht University, Faculty of Geosciences
P.O. Box 80115, 3508 TC Utrecht, The Netherlands, h.floor@geo.uu.nl

Bert van Wee
Delft University of Technology, Faculty of Technology, Policy and Management
P.O. Box 5015, 2600 GA Delft, The Netherlands, g.p.vanwee@tbm.tudelft.nl

Abstract

With the upcoming implementation of high-speed railway infrastructure in the Netherlands, interest has arisen in the spatial-economic effects this might have. Experiences with high-speed rail outside the Netherlands have shown that effects at a local or regional level can be important, due to relocation of employment within regions and cities. This paper focuses on this issue by presenting the results of discrete choice models for office location choice. Both stated and revealed choice data are used. Two approaches to combine these data sets have been evaluated; an approach using full-information maximum likelihood estimation is found to give the best results. The discrete location choice models give information on to what extent the introduction of high-speed rail in the Netherlands can change the attractiveness of individual cities within the Randstad area on the one hand and of places within these cities on the other hand. Distributive effects of offices within urban regions are expected to be larger than relocations between urban regions. Attention is also given to the specification of accessibility indicators. Furthermore, the paper focuses on a segmentation of employment that reflects this paper's purpose of studying the influence of (high-speed) rail on location choices. Accessibility by high-speed rail in particular seem important to a distinct group of office employment that regularly makes international business trips.

Keywords: high-speed rail, office location choice, accessibility, nested logit

This paper is based on research funded by Connekt-NWO as part of the stimulation program "Gebruik en waardering van vervoersnetwerken" (Use and valuation of transport networks).

1 Introduction

With the upcoming implementation of high-speed railway infrastructure in the Netherlands, interest has arisen in the spatial-economic effects this might have. Many of the reasons for governments to build high-speed railway infrastructures have an interregional or international scope: to decrease travel times on long-distance relations, to improve interregional accessibility of remote regions, or to stimulate European integration (see also Vickerman, 1997). In contrast to this scope of high-speed railway projects, experiences with high-speed rail outside the Netherlands have shown that effects at a local or intraregional level can be important, due to relocation of employment within regions and cities. For many cases urban and intraregional effects are even more evident than the regional development effect.

However, even when substantial intraregional effects can be expected for high-speed railway implementation, compared to interregional effects little attention is given to modelling the possible intraregional effects of high-speed rail. This paper focuses on this topic by studying the possible impact of the high-speed railway line Amsterdam – Brussels – Paris on the location choices of offices in the Dutch Randstad region. A novice element thereby is the use of a combined model of revealed choice and stated choice data on location choices to take account of high-speed railway connectivity. Additionally, attention is given to how the effect of high-speed rail differs depending on spatial scale, by using a nested logit structure to distinguish the choice of an urban region from the choice of a location within an urban region.

The remaining of this paper is organized as follows. Section 2 gives an overview of how accessibility influences location choices and what effects have been observed at cities with existing high-speed railway connections. Subsequently, section 3 describes the data and methodology used in this paper. Two methods for merging stated choice and revealed choice data are considered. In section 4 the results of the logit model estimations are presented. Section 5 builds upon these results by discussing the possible effects of high-speed railway implementation in the Netherlands. Finally, in section 6 some conclusions are drawn.

2 Impact of high-speed railway accessibility on the location of offices

The current section focuses on existing theories and modelling concepts of how accessibility influences location choices. Thereafter some observed developments related to high-speed railway accessibility are described. In this paper the analysis is restricted to the location of offices, as a special type of economic activity. Compared to other types of economic activities, office employment is in general less restricted in their location choice, because it does not depend on the location of natural resources, on environmental regulations or on the possibility to move goods, for example via inland waterways. As business travel is a major activity for many offices, the importance of accessibility for personal travel can be expected to be high. Furthermore, office employment is relatively mobile since it often does not have high sunk costs for its location. These factors and the importance of railway accessibility for offices are illustrated by the concentration of offices at (especially the larger) railway stations in the Netherlands. The following subsection describes the role of accessibility for the location choices of offices. Subsequently some examples are presented of observed effects at the implementation of existing high-speed railway lines.

2.1 Office location choices and accessibility indicators

The location of offices is influenced by many characteristics of both the offices themselves and the locations where they could potentially be located. In this paper a discrete choice approach based on random utility maximization is adopted to take account of this variety of factors. From this perspective decision makers of offices choose a location for their office on the basis of the attributes of the locations that are available (see McFadden, 1978; Train, 2002). Since it is not feasible to attain all relevant location attributes and characteristics of the offices only the probability that a location will be chosen can be estimated. Because of the large uncertainties, interpretation of results can only take place on an aggregated level.

An assumption that is made in several location choice studies (see e.g. McFadden, 1978) is that locations are more similar to other locations within the same city or region than to locations in another city/region. Therefore the probability of a location being chosen depends stronger negatively on the attractiveness of locations within its own region. As a result, for example, the main city in a peripheral region may attract more economic activities than a city of similar size in a more urbanized region (assuming the latter region has more competing locations), even when the second city has comparable or even better characteristics e.g. in terms of accessibility. Accordingly, if the attractiveness of a single location changes then this affects the probability of locations in the same region more than locations in other regions. In the context of high-speed rail this is relevant because it explains why high-speed rail can have a larger effect on the urban scale than on an interregional level.

In general, accessibility can be seen as an important factor for the attractiveness of a location for offices. From better accessible locations a geographically larger market for selling products can be reached, as well as a larger labour market and a larger pool of potential other business contacts. If a significant gain in travel time and/or cost can be achieved then a transport improvement might be seen to integrate previously separate labour or product markets into a single functional region (Blum *et al.*, 1997). Furthermore, new transport infrastructure can decrease average travel times and/or costs to destinations within the target market.

Several types of accessibility indicators can be used to quantify the concept of accessibility. Within a land-use/transport interaction context potential accessibility indicators derived from a gravity-type spatial-interaction model (Wilson, 1971; Vickerman, 1974) are often used. These can be interpreted as the size of a market. At a local level accessibility can additionally be represented by more easily interpretable connectivity indicators (see e.g. Waddell and Ulfarsson, 2003) such as the distance or travel time to a network node (railway station, motorway ramp) or the quality of a nearby network node (level-of-service of a railway station).

Within this perspective, an important issue is how offices take account of competition. At a large spatial scale, competition can have a divertive effect when offices aim to find a location with an optimal proportion of the market size to the number of competitors. Accessibility indicators exist that take account of competition effects in this way, such as the balancing factors of a doubly-constrained spatial-interaction model (Geurs and van Wee, 2004). At a lower spatial scale however, competition may result in clustering of economic activities. Offices focussing on the same geographical market may end up in choosing locations at a central place (or better: at a good accessible place such as near a transportation node) close to each other if following a game theory approach, as in the tradition of Hotelling (1929). Furthermore, agglomeration economies (see e.g. Krugman, 1991) can be a reason for offices to form clusters within a city or region. Competition effects can have a great impact on the eventual impact of high-speed railway implementation. However, the outcome of competition in interaction with other accessibility aspects is not easily comprehensible.

For high-speed rail in particular also more subjective factors play a role. The location decisions of firms are often based on perceptions, which are not necessarily corresponding to reality (Pellenbarg, 1985; Rietveld, 1994). The availability of high-speed train services can raise the image or status of a location, which is an extra factor attracting activities (see e.g. van den Berg and Pol, 1998). Furthermore, in location choices access to transport modes can be treated differently from the travel behaviour of employees and visitors. For example, the opportunity for visitors to use the train to come to an office might make train connectivity more important to this office than would be expected solely on the basis of actual use of the train. These effects are an extra motive to include connectivity indicators in location choice models. However, when interpreting the result of location choice models it is difficult (and often impossible) to make a distinction between the accessibility aspect, perception aspect and image aspect of connectivity.

2.2 Evidence for high-speed rail

A question to remain is to what extent the above-mentioned concepts can be observed in practice. In the last few decades several descriptive and statistical studies have been carried out on the local and regional effects of high-speed rail, which provide information on these topics. On an urban scale, entrepreneurial surveys can shed light on the motives of location decisions and the role of high-speed rail. This type of research has been carried out mainly in France, such as studies reported by Bonnafous (1987), Sands (1993) and Mannone (1997). In general, high-speed rail accessibility is just one of a series of factors that influence location decisions. In a sample of entrepreneurs located near the Lyon Part-Dieu high-speed railway station Mannone (1997) found only about one-third of the respondents indicating that the high-speed train services had been a predominant factor in their location choice.

These cases also show that the impacts of high-speed rail are to a large extent intraregional distributive effects. For example in Grenoble – where accessibility impacts of the TGV where much smaller than in Lyon and therefore less important in location choices – the revitalized station area did attract several firms and institutions from other places within the city but not from outside the region (Mannone, 1997). For the case of Grenoble Mannone (1997) suggests image effects to be relevant as well, as is also mentioned by Sands (1993) for the city of Nantes. However, the importance of image effects in location choices is difficult to assess from these studies. Sands (1993) also describes the relocation of companies from small towns to cities with a high-speed train station, which is another indication that location choices can be seen as having a spatially hierarchical structure.

On a regional scale, several studies have found positive statistical relationships between high-speed railway connectivity and regional development. These are mostly studies on the Japanese Shinkansen, for example Brotchie (1991) describes studies by Hirota (1984) and Nakamura and Ueda (1989). An overview of these studies is given by Haynes (1997). Some comments should be made about the interpretation of these results. Firstly, other factors are found to yield higher correlations than high-speed rail (such as expressway connectivity, Nakamura and Ueda, 1989; Brotchie, 1991). Secondly, the statistical relationship does not fully reveal the causal relationship between high-speed rail and regional development: the Shinkansen might as well have been connected to cities in anticipation to an expected growth of the city. Although these figures are commonly used in the discussion on high-speed rail's spatial-economic impact, the real impact of high-speed rail on regional-economic development is still difficult to assess.

3 Methodology

The remaining of this paper describes the results of a discrete choice model for office locations using both stated and revealed choice data. The purpose of this model is to study what factors are important the location choices for different types of offices, thereby taking into account the future high-speed railway accessibility. The revealed choice data is used to attain a model of the current condition of office location choices. The stated choice data enriches these data by reducing correlations between attributes concerning station level-of-service and introducing high-speed rail as a new level-of-service component. Two different methods to combine the stated and revealed choice data are used: a full-information maximum likelihood method using a nested logit structure and a two-step method using consecutive models for the stated and revealed choice data. Analyses are carried out with commercially available software (Nlogit version 3.0, Econometric Software, 2003).

The next subsection gives a description of the data set used and the study area. Subsequently, separate models of revealed and stated choice data are described. These are the building blocks of the combined models of stated and revealed choice data, which are described in the fourth subsection.

3.1 Data and study area

The analyses in this paper are based on telephone interviews held among offices' decision makers and other employees who are involved in (possible) location decisions of offices in the Dutch Randstad area. Offices were selected with at least 20 employees. From every office one respondent was asked questions about the office's current location, the activities performed by the office, and the travel frequencies and distances of its employees and visitors. For the location choice model this yielded 297 valid observations of the current location of offices along with a number of characteristics of these offices. Following the telephone interviews a stated choice experiment was conducted by e-mail and post among the respondents of the telephone interviews. This resulted in a data set of 167 valid stated choice observations

A preceding short series of explorative in-depth interviews raised the suggestion that there is a large variety among offices not only in the importance of railway accessibility for their location choices but also in the factors that determine this importance. A more traditional segmentation solely based on economic sectors does not seem appropriate for this application. Therefore several additional office characteristics influencing taste heterogeneity towards accessibility are considered. A priori expectations existed that the importance of accessibility by the different transport modes depends on the distance of incoming and outgoing business trips and the number of business trips made per employee. On the basis of these expectations and on initial model results a set of characteristics is selected that has most explanatory power in the combined stated/revealed choice model within the capabilities of the data set. Table 1 below gives an overview of these characteristics and how often the different values occur in the telephone interview data set. The table also indicates the dummy values that were assigned to be used in the discrete choice models. Non-response led to missing values (indicated as “unknown” values) for the characteristics that were derived from the telephone interviews (all but the branch of industry).

The offices in the data set are located in the provinces of North-Holland, South-Holland and Utrecht, which make up the Randstad area. Two high-speed rail connections are relevant for this region: an Amsterdam-Brussels-Paris connection for which new dedicated infrastructure is under construction, and an Amsterdam-Cologne-Frankfurt connection using conventional railway infrastructure within the Netherlands. These lines are part of the PBKAL-network that connects the major agglomerations in Western Europe. The high-speed railway infrastructure in the Netherlands can lead to a significant travel time gain for both international and domestic train services. For example, a train trip between Amsterdam and Rotterdam (currently taking over an hour) will be reduced to about 35 minutes. However, its effect on accessibility is limited by the higher train fares of high-speed rail compared to conventional train services. Figure 1 below shows the alignment of the study area within the schematically drawn PBKAL network. The figure also shows the demarcation of urban regions (COROP-regions, a classification commonly used for spatial statistics in the Netherlands.) This regional demarcation is used in the revealed choice parts of the discrete choice models.

Table 1: Respondent characteristics used to represent taste heterogeneity.

Office characteristic	Value	Dummy	Sample[1]
Branch of industry	Business/financial services	1	62,6 %
	Other	0	37,4 %
Employees visiting customers at least once a week	Less than 50% of employees	0	55,2 %
	50% or more	1	35,4 %
	Unknown	0	9,4 %
Spatial orientation customers	Local/regional	0	23,9 %
	National/international	1	40,7 %
	Not relevant	0	29,6 %
	Unknown	0	5,7 %
Has employees making international business trips at least once a month	Yes	1	39,4 %
	No	0	57,9 %
	Unknown	0	3,7 %

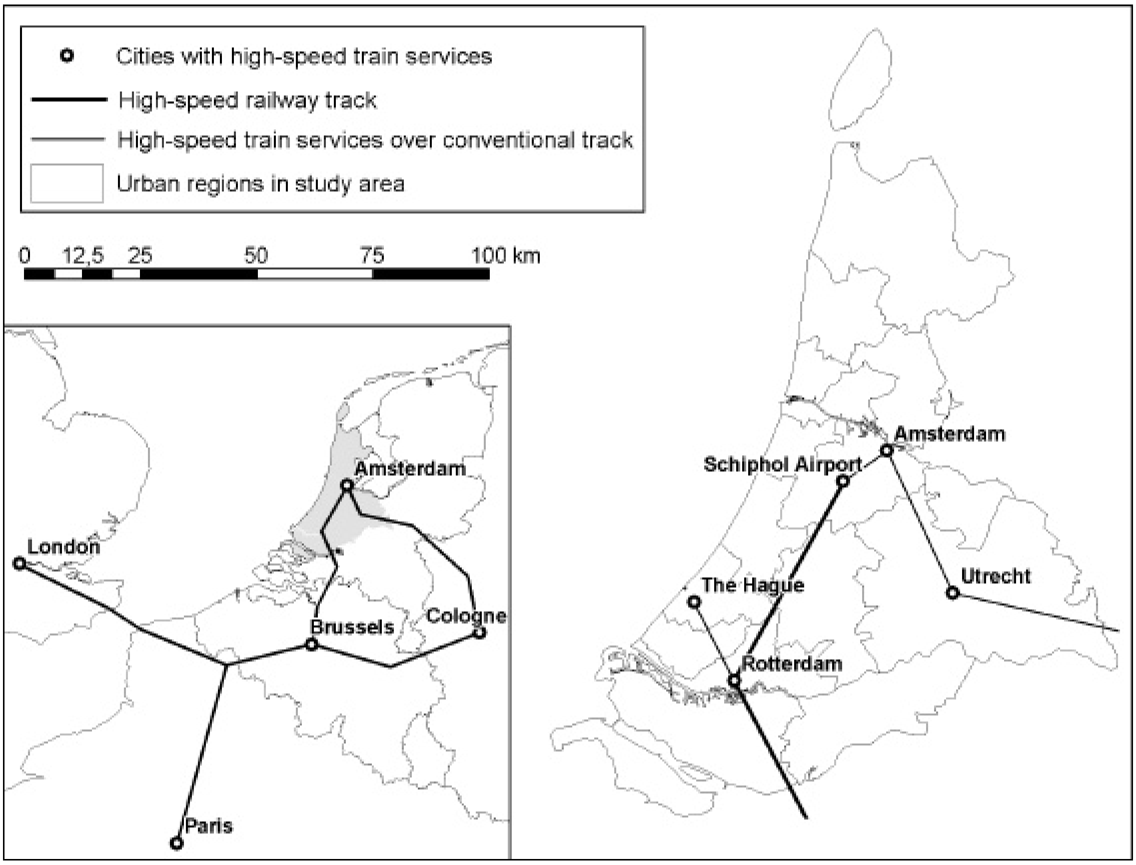

Figure 1: Study area and its position in the PBKAL network.

[1] Percentages are based on the sample for the revealed choice data. Due to rounding the percentages may not always count to 100 % exactly.

3.2 Revealed location choice model

Before combining stated choice and revealed choice data into joined location choice models, separate models for these data sources are estimated. The location choice models in this paper are all discrete choice models based random utility theory. In random utility theory the utility of choosing a certain location is divided into an observed component and a randomly distributed unobserved component. The random component represents all factors that are not explicit in the model. For simplicity in the current paper only additive utility functions are assumed, i.e. no interaction effects are estimated[2]. The models do take account of taste heterogeneity by implementing the dummy values from Table 1 into the taste parameter. The observed utility component V_{zq} of option z for office q is expressed as:

$$V_{zq} = \sum_{k \in K} b_{kq} X_{zk} = \sum_{k \in K} \left(b_{k,0} + \sum_{m \in M} b_{k,m} W_{mq} \right) X_{zk} \quad (1)$$

Hereby is:

b_{kq} : the taste parameter of the k^{th} attribute for office q,

X_{zk} : the value of the k^{th} attribute for location option z, as discussed below,

$ß_k$: components of the taste parameter to be estimated

W_{mq} : the m^{th} heterogeneity dummy code from Table 1 for office q.

In the revealed choice model offices are modelled to choose a location among a set of alternative locations, represented in a GIS by grid cells with a width of 250 meter. Since the number of grid cells within the study area is very large, the model is based on a sample of locations following the “independence of irrelevant alternatives” property of the model within the 14 urban regions in Figure 1 above. Choice sets are generated following a methodology described by McFadden (1978). Each of the choice sets consists of the current location of the office (the chosen option), 6 randomly sampled alternatives from the same urban region and 7 randomly sampled alternatives out of each of the other regions, to make a total of 98 choice alternatives.

In many applications of random utility theory it is assumed that the unobserved component is independently and identically distributed with a type I extreme value distribution, and that the utility is independent of irrelevant alternatives (so the observed component is a function only

---

[2] The two approaches to combine stated and revealed choice data would require different ways to take account of interaction effects, making comparison more difficult.

of attributes of the choice option itself). Under these assumptions the standard multinomial logit model can be derived. However, as described in section 2, in location choice models these assumptions can be relaxed by holding locations within the same city or region to be more similar than locations in different cities or regions. In this paper for the revealed choice model a nested logit form is assumed, whereby locations in the same urban region are placed within the same nest. The inclusive-value parameters that determine the degree of similarity can be different across regions, since the regions are different in size and other characteristics. The probability of an office q choosing a location z in region r is then expressed as:

$$P_q(z) = \frac{N_r \exp\left[V_{zq} + (t_r - 1)\ln\left[\sum_{z \in Z_r} \exp[V_{zq}]\right]\right]}{\sum_r N_r \exp\left[t_r \ln\left[\sum_{z \in Z_r} \exp[V_{zq}]\right]\right]} \qquad (2)$$

Hereby is:

N_r : The total number of zones available in region r,

t_r : A set of (region-specific) inclusive value parameters to be estimated for the nested logit model, theoretically restricted as $0<t=1$ (Daly, 1987),

Z_r : The collection of sampled zones z in region r,

In this model the urban regions are assumed to be static, as the immediate ability of high-speed rail to change patterns of commuting or business travel is seen to be rather limited. If in the long term indirect effects may cause location and/or transport effects to be more substantial then linking the location choice model to an analysis of functional regions is possible.

The utility function is built upon several accessibility and non-accessibility location attributes. For accessibility distinction is made between centrality indicators and connectivity indicators. Centrality refers to the position (in terms of travel time, travel cost, etc.) of a location relative to possible origins and destinations of trips. Centrality indicators are for example the potential accessibility indicators derived from spatial interaction models, which can be interpreted as the size of the labour market or product market. Following an origin-constrained spatial interaction model with mode choice, this can be expressed as:

$$A_i = \sum_m \sum_j D_j \exp[V_{ijm}] \qquad (3)$$

Hereby is:

A_i : The accessibility of an origin zone i,

D_j : The attraction of a destionation zone *j*,

V_{ijm} : The observed utility component of travelling from origin *i* to destination *j* by transport mode *m*,

In the current application four different transport modes are distinguished: car (both driver and passenger), train, other public transport and a leftover category with mainly cycling and walking. The observed utility functions for the different transport modes consist of one or more transport impedance attributes and an alternative-specific constant (except for the train, which is used as a reference). For potential accessibility indicators the shape of the impedance function is important, since it determines how the indicators react on a change in the impedance attributes. A linear utility function often performs well for urban transport models, but for larger study areas logarithmic conversions are commonly seen as superior. Also more generalised impedance functions can be used. In the current paper a Box-Cox conversion (e.g. Mandel *et al.*, 1997; Tiefelsdorf, 2003) is applied to determine the shape of the impedance parameters. It introduces per attribute an extra shape-parameter *?* in the model. The Box-Cox conversion has a linear function (*?* = 1) and a logarithmic conversion (*?* = 0) as special cases The utility function (except for the rare case where *?* exactly equals zero) is then specified as:

$$V_{ijm} = a_m + \sum_{k \in K_m} b_{mk} \left(X_{ijm}^{l_{mk}} - 1 \right) \Big/ l_{mk} \qquad (4)$$

Hereby is:

a_m : A mode-specific constant to be estimated,

$ß_{mk}$: A set of K_m impedance parameters for mode m to be estimated,

X_{ijmk} : A set of K_m impedance attributes for travel by mode *m* from origin *i* to destination *j*,

$?_{mk}$: A set of K_m Box-Cox parameters for mode m to be estimated.

Connectivity is defined as how well a location is connected to a transport network. Two aspects of connectivity can be distinguished: the access to a transport node and the level-of-service of that transport node. For train connectivity the distance to the nearest railway station is taken into account, as well as the frequency of trains at this station and the availability of intercity services. Car connectivity is expressed by the travel time to a motorway ramp. Analogous to the centrality indicators for the node access attributes Box-Cox conversions are used to determine the shape of the utility function.

Besides the accessibility attributes also a land-use typology is used as a location attribute. The typology used (Maat *et al.*, forthcoming) depicts both the function of the zone (residential areas, industrial sites, city centres, other) and the density of land-use in the zone. A total of ten types are distinguished. These types are quantified as –1/0/1 effect codes, using the left-over category as a reference.

The price of real estate is not modelled explicitly in the revealed choice model. Taking account of real estate prices is complex in discrete choice models that are based on revealed choice data, because of a mutual dependency of real estate price and location attractiveness. Therefore, applications could result in parameter estimations being insignificant or having a theoretically unexpected sign. However, under the assumption that real estate price is linearly dependant on the other attributes in the model, the price can be taken into account implicitly, incorporated into the other attributes.

3.3 Stated location choice experiment

In a stated choice experiment respondents are asked to choose an option out of a hypothetical choice set. A choice option can exist of any attributes with values that seem feasible to the respondent, including (combinations of) attributes and attribute levels that do not exist (yet) in reality. The choice sets can be composed in such a way that attributes are uncorrelated to each other. Stated choice experiments can thus be used to study technological innovations, such as high-speed railway implementation, and to better isolate the influence of attributes that are correlated in reality. (However, as a main drawback stated choice models are typically not good at replicating actual market shares.) For an overview of the advantages and disadvantages of stated choice research compared to revealed choice models see Louviere et al. (2000).

Attributes taken into account are indicators for train connectivity, car connectivity and three non-accessibility factors; see Table 2. Centrality was not included, since its indicators are often difficult to interpret and because differences in centrality are normally small at the urban spatial scale that was supposed. To minimize correlations between key attributes in the experiment an orthogonal design is applied to compose choice sets. A bridging design is used, where the total design of nine attributes consists of two sub-designs of six attributes each, thus having three attributes in common. Every respondent was sent a questionnaire consisting of eight choice situations, four out of each sub-design. Respondents where thereby instructed to regard all attributes not explicitly described in a choice situation to be equal across the alternatives and to be at an acceptable level.

Table 2: Attributes and levels in the stated choice experiment

Attributes	Levels	Type	Sub-design
Train connectivity			
Travel time to a station	5, 10, 15 or 20 minutes	Absolute	Both
Transport mode to this station	Walk or bus	-1/1 Dummy	1
Total frequency of trains departing from this station	4, 16, 28 or 40 trains per hour	Absolute	Both
Type of train services departing from this station	Only regional trains, also intercity trains, also domestic HST, also international HST	-1/0/1 Effect code	Both
Car connectivity			
Travel time to a motorway access	5 or 15 minutes	Absolute	1
Number of parking places per 100 employees	75 or 100	Absolute	2
Non-accessibility factors			
Type of building	"Nice but not extraordinary" or "architectonic remarkable"	-1/1 Dummy	2
Type of environment	"In a city centre" or "in a city-rim office park"	-1/1 Dummy	1
Price of real estate	€150 or €200 per m^2 per year	Absolute	2

To analyse the stated choice data a multinomial logit model with an observed utility as in Equation 1 is estimated. Although a random parameter logit model (a version of the mixed logit class of models, see Hensher and Greene, 2003) would be better to take account of multiple observations per respondent and techniques exist to use mixed logit for combination of data sources (Brownstone *et al.*, 2000), it is practically not feasible to use this type of model in the current application because of the large number of revealed choice options.

In the stated choice model the chance of a respondent q choosing option z is expressed as the simple logit form:

$$P(z) = \exp\left[V_{zq}\right] \Big/ \sum_{z \in Z} \exp\left[V_{zq}\right] \qquad (6)$$

Two attributes have four levels, which allows estimation of non-linear effects. For the travel time to a station a third-order polynomial function is used, whereas a logarithmic conversion is applied to the train frequency at a station. The current paper will not go deeper into the meaning of these non-linear functions. In this application the functions are included solely to improve the stated choice models.

3.4 Combined stated/revealed location choice models

In the combined stated/revealed location choice models single taste parameters are estimated for attributes that are similar in the two data sets. However, a known problem in combining data sets in a logit model is that the scale parameters of the models to be combined are unknown (Swait and Louviere, 1993). The scale parameter is inversely related to the standard deviation of the unobserved component of the utility function and thus expresses how well the observed utility component explains the choices being made by the respondents. The scale parameter of a data set cannot be directly estimated in a logit model, but is instead incorporated into the taste parameters. Therefore, taste parameters can only be hold equal when correcting for the difference in scale parameter.

Although the absolute value of the scale parameters is unknown, the ratio between the scale parameters can be determined when combining multiple data sets through data fusion. In literature, several methods of estimating this relative scale parameter have been developed: methods using different types of logit models (for different approaches, see e.g. Hensher *et al.*, 1999; Brownstone *et al.*, 2000; Louviere *et al.*, 2000) or requiring 'manual' calculations (e.g. Swait and Louviere, 1993; Swait *et al.*, 1994). In this paper two different methods are used: one using a nested logit structure with full-information maximum likelihood (FIML) estimation (following Louviere *et al.*, 2000) and one using a sequential estimation procedure (following Swait *et al.*, 1994).

In the FIML approach the stated choice options are treated as if it were additional options in the revealed choice model. The revealed choice options are therefore replicated eight times to match the number of stated choice observations per respondent. In a nested logit structure the stated choice options are then combined into separate branches with equal inclusive value parameters; the revealed choice urban regions are retained in the root of the nesting (see Figure 2). The full data set is then duplicated, whereby once the revealed choice option is given as chosen and once the stated choice option. In this model set-up the relative scale parameter of the stated choice data with respect to the revealed choice data then equals the inclusive-value parameter of the stated choice branch. The taste parameters for attributes similar to the two data sets are then estimated to maximise model fit for stated and revealed choice data together.

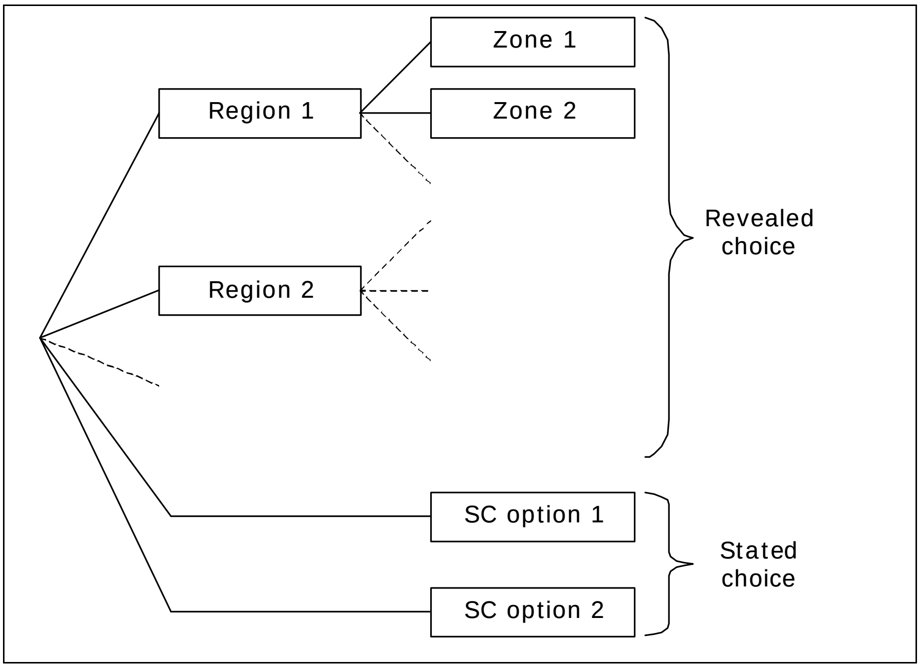

Figure 2: Nesting structure of the full information maximum likelihood method.

In the sequential estimation procedure first a multinomial logit model for the stated choice data is estimated. For the attributes similar to the two models the stated choice estimates are then assumed to be the right values for the taste parameters. By using these taste parameters in a logit model to the revealed choice data the relative scale parameter can be estimated. With this method the stated choice data are used to determine how the model reacts to an attribute change and the revealed choice data to establish a realistic magnitude of this reaction.

The two approaches are thus not only different in the complexity and type of the estimation method, but are also grounded in different assumptions with regard to the characteristics of the stated and revealed choice data. In the full-information maximum likelihood approach stated choice data is seen as a way to enrich the revealed choice data and thus to decrease correlations among attributes and to identify attributes that could not be distinguished on the basis of revealed choice data only (Louviere *et al.*, 2000). The sequential approach goes a step further by stating that parameter estimations for attributes should only be based on stated choice data and that the purpose of the revealed choice data is to ground these parameters in reality by estimating the relative scale parameter and alternative specific constants (Swait *et al.*, 1994). The current application differs from this view as no alternative specific constants are estimated; instead additional attributes are taken into account in the revealed choice data.

For application of the two methods described above it is important on which attributes data fusion is to take place. In order to use an attribute for combining data sources it is important that the attribute in both data sets is measuring the same feature and is expressed in the same unit. Furthermore, several aspects may cause two similar attributes not to be suitable for grounding data fusion, most notably measurement errors in observed data and respondent's perceptions differing from real values. Since the main purpose of the data fusion is to determine the effect of station level-of-service characteristics on location choices in the current paper solely the train frequency and the type of train services are used as merging attributes. The rather limited number of merging attributes may make the scale parameter less accurate for other attributes than station level-of-service, but seems most suitable for the purpose of this application.

4 Parameter estimation results

In the appendix the parameter estimation results are presented for all five models. All models show a satisfactory ?^{2} (pseudo-r^2) for model fit.

4.1 Separate revealed choice and stated choice models

In the revealed choice model potential accessibility to employees is an important attribute for the location choices; accessibility to business partners is also a relevant factor. Besides these centrality indicators connectivity indicators – especially access to a motorway and access to a station – are important, indicating that not only the actual travel effort is relevant but also factors such as the opportunity to make use of a transport mode and the ease for visitors to find an office are important. Furthermore, several elements of the land-use typology are significant. Among the heterogeneity parameters estimated the model shows estimations only for the branch of industry to be significant at a ten percent uncertainty level.

A nested logit structure appears to be considerably better than a multinomial logit model. The inclusive value parameters are below or around the theoretical upper boundary of one (see Daly, 1987). For Utrecht and The Hague these parameters are significantly higher than for the regions; this can be interpreted as that locations within the urban regions of Utrecht and The Hague are less similar to locations in the same region than is the case for other regions.

In the stated choice model the price of real estate is the most important attribute. (Recall that prices are implicit in the revealed choice model.) Also many of the other parameters are significant at a one percent uncertainty level. This includes some of the office heterogeneity parameters. Firstly, having employees regularly visiting customers significantly increases the importance of motorway access time. Furthermore, with regard to train connectivity, having a spatially larger target market makes the presence of intercity services more important to an office. As expected international high-speed railway services are only significant to offices whose employees frequently make international business trips. In contrast to the revealed choice model, the stated choice model does not show high significance for the branch of industry as an office characteristic.

4.2 Combined stated/revealed choice models

Both combined stated/revealed choice models show satisfactory results in terms of model fit and parameter significance. Especially the FIML approach shows high parameter significance for both main effects and office heterogeneity parameters. Furthermore, a likelihood ratio test shows that merging of the attributes in the full-information approach is superior to separate stated and revealed choice parameters at 98% certainty. For the sequential approach the revealed choice sub-model has a goodness-of-fit comparable to the revealed choice model. Since the models are based on different data set-ups the maximum-likelihoods and corresponding $?^2$ values cannot be directly used for comparing the two combined models.

Substantial differences do exist between the results for station level-of-service of the two combined models. For the sequential approach estimated parameters are generally lower and statistically less significant than for the FIML approach, both for the main effects and for the office heterogeneity effect. This is accompanied by a higher relative scale parameter for the sequential approach, which was expected since in this approach parameters are optimised for stated choice data and would therefore show less declarative power on revealed choice data. For other attributes results of the combined models are generally consistent to each other. On the basis of these results a preference exists for the FIML approach.

5 Spatial-economic effects of high-speed rail in the Netherlands

The results above show that both centrality and connectivity are important factors to the location choices of offices. Domestic high-speed rail services can thus influence location choices by decreasing domestic travel times and thereby increasing potential accessibility for location in or near cities with a high-speed railway station. Furthermore, a station that has domestic high-speed railway services has an extra attractiveness on its immediate surroundings by its connectivity effect, which in the current model could be partly due to the possibility to use the high-speed train services with little access/egress effort and partly to an image effect of the high-speed train services.

In the Netherlands, the high-speed railway line Amsterdam-Brussels-Paris – of which the Dutch section is to be opened in 2007 – not only decreases travel times on international connections but also considerable on domestic connections. Especially the Amsterdam-Rotterdam-Breda service will improve potential accessibility of places near Amsterdam and Rotterdam. Figure 3 below shows the effect of domestic high-speed train services on potential accessibility (following Equation 3) for both commuting and business travel. For commuting potential accessibility represents the size of the labour market. As can be seen from the figure, the impact of high-speed rail is rather small and limited to the city centres of Amsterdam and Rotterdam. Potential accessibility for business travel is more affected by the high-speed railway services. The pool for business contacts increases for a wider area around Amsterdam and Rotterdam.

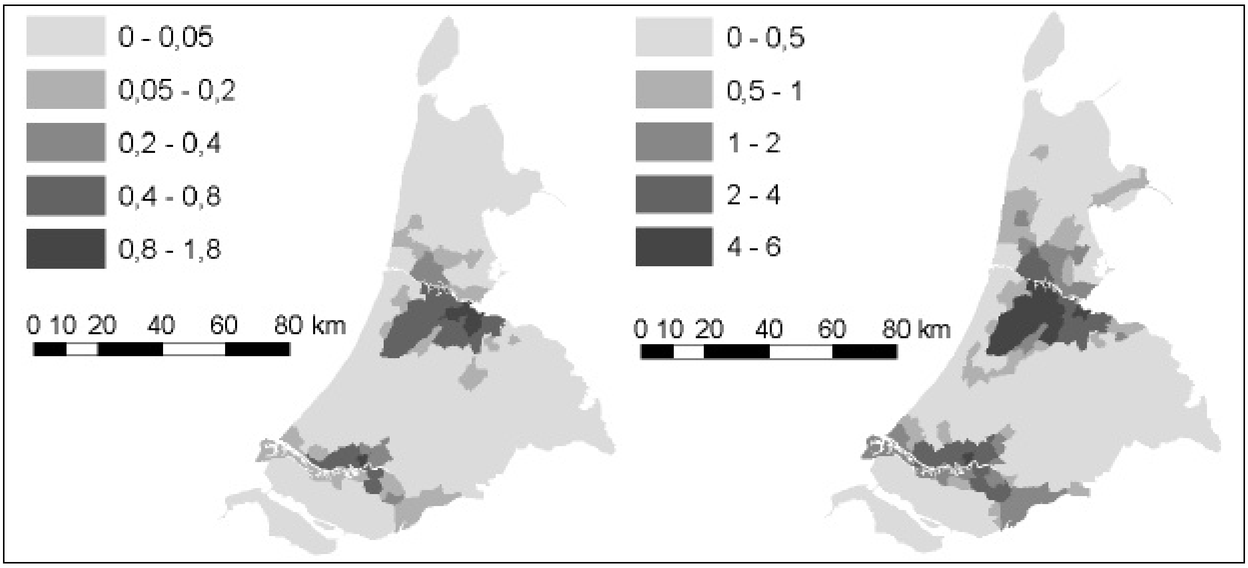

Figure 3: High-speed rail's effect (ceteris paribus) on potential accessibility towards employees (left) and business partners (right) based on year 2000 data (the maximum accessibility by train in 2000 is 100).

For international high-speed railway services the effect of connectivity is large in both combined models, however only for a distinct group of offices of which part of the employees frequently make international business trips. This group accounts for just below 40 percent of all offices in the sample. The image effect of high-speed rail is assumed to be important here, because the share of international trips in the total amount of business trips is small for most offices.

The increased accessibility of railway station areas – especially the high-speed railway stations due to the connectivity effect – leads to a higher attractiveness of railway station locations and therefore to a higher concentration of offices around stations. The results of the nested logit model show that relocations can be expected to be largely distributive within the urban regions. For the urban regions of both Amsterdam and Rotterdam inclusive value parameters are well below one, indicating that locations within these regions are more competing to each other than to locations in other regions. However, for locations within Utrecht and The Hague similarity is rather low. This supports our prior expectation that the degree of similarity can be different across regions. In our application this approach also performed better than alternative nesting structures, such as a model with regional specific constants in combination with identical inclusive value parameters across regions.

For the eventual effect of high-speed rail on the location choices of office employment several other factors are important. Firstly, if more office employment clusters at railway station areas then this leads to a further increase in accessibility of railway station areas due to shorter access and egress distances. Additionally, these station locations can attract more employment because of clustering effects. Finally, although high-speed rail is not expected to have a considerable impact on functional regions because of the small share of the train in trips within the Netherlands, the extent to which competition effects can have a divertive effect in this case is still unclear.

6 Conclusions

This paper focuses on (possible) intraregional effects of high-speed railway infrastructure. To do so, this paper presents the results of a discrete choice model for office locations in the Dutch Randstad area. The model combines both stated choice and revealed choice data in a single model to take account of the potential accessibility effect by reducing domestic travel times as well as the connectivity effect of high-speed rail being a new transport mode within the Netherlands. Furthermore, a nested logit structure is applied to the model to study how high-speed rail may affect location choices within and between urban regions.

Two approaches for combining stated and revealed choice data have been applied: one approach that estimates the parameters for station based on stated and revealed choice data together and one approach that uses only stated choice data for the level-of-service parameters and revealed choice data to ground these parameters in a ‘more realistic’ context (i.e. a model that better takes account of the variation and market shares that are observed in reality). Although the overall model statistics for these two approaches are quite similar, the first approach is preferred over the second because it shows clearer relationships for the station level-of-service relationships.

The parameter estimations and the changes in potential accessibility show that both centrality and connectivity effects of high-speed rail are important. By reducing travel times domestic high-speed railway services increases the potential accessibility for business travel in the areas around Amsterdam and Rotterdam. For commuting the effect is much smaller and limited to the Amsterdam and Rotterdam city centres. Connectivity is most important for international high-speed railway services, although this effect is only relevant for offices with employees who regularly make international business trips. For most urban regions the nested logit model found relocation effects to be stronger within the region than between urban regions. The model thus gives an explanation of the intraregional distributive effects that are reported in literature.

Further research should give more details on how high-speed railway services have an impact on competition between offices in the cities that are connected. Also, several feedback mechanisms can be important. For example, an increased concentration of offices around railway stations has a further effect on potential accessibility. Finally, the model that has been developed can be used to study different scenarios, such as the location of high-speed railway stations and the height of the additional train fare for high-speed train services.

References

Blum, U., K.E. Haynes and C. Karlsson (1997). "Introduction to the special issue: the regional and urban effects of high-speed trains." *The Annals of Regional Science* **31**(1): 1-20.

Bonnafous, A. (1987). "The regional impact of the TGV." *Transportation* **14**: 127-137.

Brotchie, J. (1991). Fast rail networks and socio-economic impacts. In: *Cities of the 21th century: new technologies and spatial systems*. J. Brotchie, M. Batty, P. Hall and J. Newton, Eds. Melbourne, Longman Cheshire: 25-37.

Brownstone, D., D.S. Bunch and K. Train (2000). "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles." *Transportation Research Part B: Methodological* **34**: 315-338.

Daly, A. (1987). "Estimating "tree" logit models." *Transportation Research part B* **21**(4): 251-267.

Geurs, K.T. and B. van Wee (2004). "Accessibility evaluation of land-use and transport strategies: review and research directions." *Journal of Transport Geography* **12**: 127-140.

Haynes, K.E. (1997). "Labor markets and regional transportation improvements: the case of high speed trains, an introduction and review." *The Annals of Regional Science* **31**: 57-76.

Hensher, D., J. Louviere and J. Swait (1999). "Combining sources of preference data." *Journal of Econometrics* **89**: 197-221.

Hensher, D.A. and W.H. Greene (2003). "The mixed logit model: the state of practice." *Transportation* **30**: 133-176.

Hirota, R. (1984). *Present situation and effects of the Shinkansen*. International seminar on high-speed trains, Paris.
Hotelling, H. (1929). "Stability in competition." *The Economic Journal* **39**: 41-57.
Krugman, P. (1991). "Increasing returns and economic geography." *Journal of Political Economy* **99**(3): 483-499.
Louviere, J.J., D.A. Hensher and J.D. Swait (2000). *Stated choice methods: analysis and applications*. Cambridge, Cambridge university press.
Maat, K., J.J. Harts, M. Zeijlmans van Emmichoven and R. Goetgeluk (forthcoming). *Dynamiek van stedelijke milieus, 1996-2002*. Den Haag, Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieu.
Mandel, B., M. Gaudry and W. Rothengatter (1997). "A disaggregate Box-Cox Logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand forecasts." *The Annals of Regional Science* **31**: 99-120.
Mannone, V. (1997). "Gares TGV et nouvelles dynamiques urbaines en centre ville: Le cas des villes desservies par le TGV Sud-Est." *Les Cahiers Scientifiques du Transport* **31**: 71-97.
McFadden, D. (1978). Modelling the choice of residential location. In: *Spatial interaction theory and planning models*. A. Karlqvist, L. Lundqvist, F. Snickars and J.W. Weibull, Eds. Amsterdam, North-Holland: 75-96.
Nakamura, H. and T. Ueda (1989). *The impacts of the Shinkansen on regional development*. 5th World Conference on Transport Research, Yokohama, Western Periodicals, Ventura, California.
Pellenbarg, P.H. (1985). *Bedrijfslokatie en ruimtelijke kognitie*. Geografisch Instituut. Groningen, Rijksuniversiteit Groningen.
Rietveld, P (1994). "Spatial economic impacts of transport infrastructure supply." *Transportation Research Part A: Policy and Practice* **28**(4): 329-341.
Sands, B. (1993). "The development effects of high-speed rail stations and implications for California." *Built Environment* **19**(3/4): 257-284.
Swait, J. and J. Louviere (1993). "The role of the scale parameter in the estimation and comparison of multinomial logit models." *Journal of Marketing Research* **30**(3): 305-314.
Swait, J., J.J. Louviere and M. Williams (1994). "A sequential approach to exploiting the combined strenghts of SP and RP data: application to freight shipper choice." *Transportation* **21**: 135-152.
Tiefelsdorf, M. (2003). "Misspecifications in interaction model distance decay relations: A spatial structure effect." *Journal of Geographical Systems* **2003**(5): 25-50.
Train, K.E (2002). *Discrete choice methods with simulation*. Cambridge, Cambridge University Press.
van den Berg, L. and P.M.J. Pol (1998). "The urban implications of the developing European high-speed-train network." *Environment and Planning C: Government and Policy* **16**: 483-497.
Vickerman, R. (1997). "High-speed rail in Europe: experience and issues for future development." *The Annals of Regional Science* **31**: 21-38.
Vickerman, R.W. (1974). "Accessibility, attraction and potential: a review of some concepts and their use in determining mobility." *Environment and Planning A* **6**: 675-691.
Waddell, P. and G. Ulfarsson (2003). *Accessibility and agglomeration: discrete-choice models of employment location by industry sector*. 82nd Annual meeting of the Transportation Research Board, Washington D.C.
Wilson, A.G. (1971). "A family of spatial interaction models, and associated developments." *Environment and Planning* **3**: 1-32.

Appendix: Parameter estimation results

	RC model		SC model		FIML approach		Sequential appr.	
	Parameter	t-value	Parameter	t-value	Parameter	t-value	Parameter	t-value
Accessibility								
Accessibility employees	**0,533**	2,90	x		**0,387**	5,84	**0,468**	2,62
Accessibility business contacts	0,198	1,74	x		**0,208**	4,92	0,220	2,00
Train frequency	0,312	2,05	**0,328**	4,13	**0,283**	5,83	**0,234**	2,96
* Branch of industry	0,351	2,44	0,109	1,05	**0,428**	9,04	$7{,}81\ 10^{-2}$	1,02
Intercity services	-0,119	-1,93	$-6{,}79\ 10^{-2}$	-0,72	**-0,146**	-6,98	$-4{,}85\ 10^{-2}$	-0,71
* International business trips	$-3{,}10\ 10^{-2}$	-0,48	-0,222	-2,52	$-1{,}01\ 10^{-2}$	-0,44	-0,158	-2,17
* Spatial orientation customers	$9{,}23\ 10^{-2}$	1,42	0,138	2,53	**0,140**	6,19	$9{,}86\ 10^{-2}$	2,17
Domestic high-speed rail	x		0,126	1,27	0,121	1,10	$8{,}97\ 10^{-2}$	1,22
* International business trips	x		-0,222	-2,52	$-1{,}01\ 10^{-2}$	-0,44	-0,158	-2,17
* Spatial orientation customers	x		0,138	2,53	**0,140**	6,19	$9{,}86\ 10^{-2}$	2,17
International high-speed rail	x		$3{,}76\ 10^{-3}$	0,04	-0,107	-0,75	$2{,}68\ 10^{-3}$	0,04
* International business trips	x		**0,565**	3,74	**0,568**	2,88	**0,403**	2,80
* Spatial orientation customers	x		0,138	2,53	**0,140**	6,19	$9{,}86\ 10^{-2}$	2,17
Station access distance	**-0,368**	-2,61	x		**-0,403**	-10,09	-0,205	-1,74
Box-Cox parameter	-0,103		x		-0,103		-0,103	
Station access time (linear)	x		0,584	2,28	**4,44**	8,39	0,417	2,01
(quadratic)	x		$\mathbf{-6{,}52\ 10^{-2}}$	-2,85	**-0,399**	-8,35	$-4{,}66\ 10^{-2}$	-2,36
(cubic)	x		$\mathbf{1{,}80\ 10^{-3}}$	2,96	$\mathbf{1{,}05\ 10^{-2}}$	8,27	$1{,}29\ 10^{-3}$	2,43
Station access mode	x		**-0,366**	-5,70	**-0,462**	-4,37	**-0,261**	-3,40
Motorway access time	x		$\mathbf{-9{,}72\ 10^{-2}}$	-5,51	$-5{,}32\ 10^{-2}$	-2,55	$\mathbf{-6{,}94\ 10^{-2}}$	-3,36
* Employees visiting customers	x		**-0,158**	-4,64	**-0,115**	-4,80	**-0,113**	-3,13
Motorway freeflow access time	$\mathbf{-6{,}06\ 10^{-2}}$	-3,73	x		$\mathbf{-5{,}70\ 10^{-2}}$	-9,12	$\mathbf{-5{,}91\ 10^{-2}}$	-3,48
Box-Cox parameter	0,312		x		0,312		0,312	
* Branch of industry	$-5{,}03\ 10^{-2}$	-1,77	x		$\mathbf{-5{,}46\ 10^{-2}}$	-6,01	$-4{,}64\ 10^{-2}$	1,78
Parking places available	x		$\mathbf{1{,}91\ 10^{-2}}$	2,98	$\mathbf{7{,}54\ 10^{-2}}$	7,17	$1{,}36\ 10^{-2}$	2,44
Revealed choices land-use typology								
City centre high density	**0,737**	3,28	x		**0,830**	9,82	**0,741**	3,25

City centre low density	0,120	0,31	x		0,140	0,98	$8,86\ 10^{-2}$	0,23
Town centre	-1,06	-2,13	x		**-1,11**	-6,15	-1,08	-2,14
City high density	**-1,38**	-4,14	x		**-1,42**	-11,60	**-1,37**	-4,13
* Branch of industry	**0,955**	2,77	x		**1,15**	9,38	**1,04**	3,04
City average density	0,180	0,68	x		0,104	1,03	0,191	0,72
City low density	-0,295	-1,18	x		-0,225	-2,54	-0,305	-1,22
Town/village	-1,30	-2,51	x		**-1,64**	-7,87	-1,31	-2,51
Industrial/business in city	**1,18**	7,26	x		**1,22**	20,54	**1,19**	7,46
Other industrial/business	**2,07**	10,51	x		**1,97**	27,05	**2,09**	10,71
* Branch of industry	**-1,01**	-4,03	x		**-0881**	-10,09	**-1,04**	-4,24
Other stated choice attributes								
Type of urban environment	x		**-0,366**	-5,03	**-0,419**	-3,81	**-0,262**	-3,24
Type of building	x		$7,73\ 10^{-2}$	1,12	0,174	1,84	$5,52\ 10^{-2}$	1,09
Price of real estate	x		**$-3,77\ 10^{-2}$**	-12,37	$-8,80\ 10^{-3}$	-2,51	**$2,69\ 10^{-2}$**	-4,01
Inclusive value parameters								
Amsterdam	**0,695**	5,87	x		**0,620**	15,70	**0,727**	5,89
Rotterdam	**0,469**	2,96	x		**0,346**	6,73	**0,542**	3,50
The Hague	**1,16**	6,45	x		**1,05**	17,73	**1,24**	6,75
Utrecht	**1,13**	4,95	x		**1,04**	14,03	**1,13**	5,70
Other	**0,652**	4,90	x		**0,551**	14,03	**0,712**	5,12
Scale parameter (SC relative to RC)	x		x		**0,500**	9,73	**1,40**	5,89
Observations	297		167 x 8		297 x 8		297 [3]	
Maximum log-likelihood	-1035		-687		-11198		-1041	
?^{2} (pseudo-r^2)	0,240		0,259		0,245		0,235	

"x" = not applicable; bold figures are significant at 1% uncertainty level.

[3] Model statistics based on revealed choice sub-model.