How to obtain representative spatial choice sets?
Dominance and centrality analysed for firm location choices

Michiel de Bok
Significance
Koninginnegracht 23
The Hague, 2514AB
The Netherlands
Tel: (+31)-70-3121541
e-mail: debok@significance.nl

Francesca Pagliara
Assistant professor in Transportation
Department of Transportation Engineering
University of Naples Federico II
Via Claudio 21
Napoli, 80123
Italy
Tel: (+39)-0817683932
e-mail: fpagliar@unina.it

Abstract: Firm location choice is a typical example of spatial choices with corresponding problems arising from the choice context: large numbers of alternatives and complex spatial interdependencies among them. This is in particular the case in spatial detailed urban simulation models. The task of the analyst is to try to capture at best the extent of the true choice set. However, empirical studies on firm location choice typically apply conventional multinomial logit models with randomly sampled choice sets, which are very likely to be biased and therefore non representative.

The first hypothesis is that in firm location choice, it may happen that some alternatives are not taken into account since they are dominated by others. In the context of destination choice, this concept has been translated into the introduction of dominance variables in the utility function, describing the better perception of an alternative. Following this concept, dominance variables are specified for the context of firm location choice. The second hypothesis regarding choice set composition, is that clustered location alternatives are more likely to be each other's substitute. The competing destinations model formulates centrality variables that capture the clustering of alternatives, and that can be used as a proxy measure for competition patterns among clustered alternatives.

This paper aims to test these two choice set hypothesis for firm location choices. The choice set composition is systematically included in model estimations through parameters for dominance and centrality to test their influence on choice set membership. The proposed methodology and attributes are tested on a dataset with observed firm relocations in The Netherlands, and various specifications of the dominance variables are presented in the paper. The results confirm a significant influence of both dominance and centrality on the choice set formation in firm location choices.

Keywords: Firm location, Competing Destinations, Choice set formation, Dominance, Centrality

Abstract Reference Number - 168

1. INTRODUCTION

Choice behaviour can be considered as the process of selecting a certain alternative from a limited set of discrete opportunities in accordance with some decision rule (Golledge and Timmermans, 1990; Thill, 1992).

Modelling such behaviour is a task on the scale of complexity of individual decision processes. The apparent lack of concern for the individual choice set is all the more surprising because it has been recognised that many of the differences between choice among spatial alternatives and other types of choices appear in connection to choice sets (Fotheringham, 1983). Most spatial choices are made from large sets of possible alternatives. For example, the number of places in a medium-sized city where a household might choose to live can number in the thousands (Thill, 1992). The number of the elements in the universe of alternatives makes it hard to assume beforehand that the individual is able to evaluate each and every one of them and then, make an educated decision. A portion of the universe is considered instead.

The task of the analyst is to try to capture at best the extent of the true choice set. Traditional approaches consist of delineation based on a restricted list of deterministic criteria selected by the analyst. "*The validity of a choice set defined according to one of these approaches relies heavily on the analyst's expertise of the behavioural problem under study and the study area. This opens the door to likely mis-specification of choice sets"* (Thill, 1992).

The implications of an invalid measurement of choice sets depend on the situation. Most spatial choice approaches assume that the odds of choosing a particular alternative are independent of the composition of the choice (IIA property). If indeed this assumption is valid and choice set composition has no effect on pairwise choice probabilities, McFadden's (1978) proof required using a choice set by taking a random sample of the universal set. If the IIA property does not hold, choice behaviour does depend on choice set composition (which seems realistic for many spatial choice problems). In that case, unbiased estimates can only be obtained if choice set composition is systematically included in the estimation, but such ideal is very difficult if not impossible to establish when using observed choice behaviour for model estimation. Regardless of this situation, however, the *choice probabilities* will always depend on the right operational definitions of individual choice sets. If the choice set defined by the analyst includes alternatives actually never evaluated by the decision-maker, the choice model assigns non-negative probabilities to all alternatives in the choice set, including those that are not in the true choice set. Choice set composition will affect predicted market shares as the latent demand is allocated to the alternatives belonging to an individuals' choice set. Thus, predictions of market shares of choice alternatives will be wrong if individual choice sets are misspecified. If the composition of the choice set also influences individual preferences and choice behaviour, the parameters of the estimated utility or preference function will also be biased.

Firm location choices are a typical example of choice behaviour in spatial context. In particular in geographic detailed urban studies, the context is continuous with large numbers of alternatives and complex interdependencies exist between spatial alternatives. From the different approaches that exist in spatial choice modelling, only few have been applied in the empirical literature on firm location choice.

It has been proposed in the literature, in the context of destination choice, that some alternatives are not taken into account since they are dominated by others. This concept has been translated into the introduction in the utility function of dominance variables describing the better perception of an alternative.

The objective of this paper is twofold. The first is to give a general contribution to choice set modelling by extending and applying the concept of dominance among alternatives to the framework of sampling alternatives. The main result is the definition of a methodology for the generation of ad hoc dominance attributes used as weights for sampling alternatives, which can be used in choice set modelling. The second aim is to make a specific contribution to firm location choice modelling. Specifically, dominance variables are defined from the above methodology and introduced for the first time into this choice context. The proposed methodology and attributes are tested on real data and results are compared with the standard Multinomial Logit model and Competing Destinations model. This paper is organised in five sections. Section 2 deals with an overview of the current practice of choice set formation for firm location. In section 3 the concept of dominance in spatial choices is reported and adopted to model the choice set formation for firm location reporting a new methodology. Section 4 analyses the case study, while in section 5 conclusions and further perspectives are presented.

2. BACKGROUND

2.1 Choice set formation modelling approaches

In the literature, three main strands addressing choice set formation for problems with a large number of alternatives can be identified and they are deterministic, probabilistic and sampling of alternatives. An overview of the different approaches of the first two strands is given in Thill (1992), who focused on destination choice and, in a more recent literature review, in Pagliara and Timmermans (2009), who focused on spatial contexts in general.

The models of the first strand are based on a deterministic specification of choice sets, where the choice sets are an exogenous input to the estimation step (Gautschi, 1981; Weisbrod et al., 1984; Adler and Ben-Akiva, 1976; Miller and O'Kelly, 1983; Southworth, 1981; Golledge and Timmermans, 1990 and include also models of the time-geographic approach (Hägerstrand, 1970; Landau et al., 1982a; Thill and Horowitz, 1997; Scott, 2006).

The second strand, which is often called the probabilistic approach, was founded by Manski (1977) and integrates the choice set formation step into the estimation procedure and jointly estimates the selection of a choice set and the choice of a particular alternative of this choice set. Specifically, he proposed to simulate the choice probability of an alternative *d* as the sum extended to all possible choice sets *C* which can occur from a given set of alternatives of the probability that *C* occurs, $p(C)$, multiplied by the probability of choosing *d* within C, $p(d/C)$:

$p(d) = \sum_C p(C)\, p(d/C)$ (1)

The main problem that is associated with Manski's approach is that the number of elements to which the sum is extended increases exponentially when the number of alternatives increase. Therefore, the application of this method becomes very difficult in choice contexts implying a large number of alternatives. To avoid the computational problems connected with the enumeration of all the possible choice sets, an implicit approach has been proposed in the literature. In this case the probability that an alternative d belongs to the decision-maker's choice set C, $p(d \in C)$, is jointly simulated with its probability of being chosen within this set, $p(d/C)$, by adding the logarithm of $p(d \in C)$ in the utility function of alternative d, U_d, the theoretical proof is given in (Cascetta and Papola 2001):

$$U_d = V_d + \ln p(d \in C) + \varepsilon_d \qquad (2)$$

By assuming, for instance, the random residual ε_d identically and independently distributed as Gumbel variables the following formulation is obtained:

$$p[d] = \frac{\exp(V_d) \cdot p\ (d \in C)}{\sum_{d'} \exp(V_{d'}) \cdot p(d' \in C)} \qquad (3)$$

Various approaches have been proposed for the simulation of $p(d \in C)$ (Cascetta et al., 2007).

The third strand of choice set formation techniques are sampling of alternatives. This technique is commonly applied to avoid the computational burden involved in estimating choice models with large number of alternatives (Bierlaire et al., 2006). Chapters 8 and 9 of the book by Ben-Akiva and Lerman (1985) are prototypal since they provide all the different methods of sampling alternatives and corresponding estimators for the choice set. Specifically, four conventional sampling approaches can be identified: (1) simple random sampling, (2) stratified sampling (special cases are the exogenous sampling and the choice-based sampling), (3) cluster sampling, (4) double sampling, and multistage extensions. In this respect the most important conclusions regarding sampling strategies for discrete choice analysis. The estimation procedure used for simple random samples is appropriate to exogenous stratified samples or to any general stratified sample in which the fraction of each stratum in the sample equals the corresponding population share. Choice-based samples in general require different estimation procedures than exogenous stratified samples. Mixture of exogenous and choice-based samples, called enriched samples, can be used to estimate a choice model's parameters.

2.2 Choice set formation in firm location studies

The first choice studies on firm location were usually based on geographical rough data, applying a nested choice structure to deal with spatial interdependencies of alternatives (Hansen, 1987). These models were estimated on data for rough geographic zones (cities) and the aggregate nature of the data made it less problematic to define a priori segmentation of choice alternatives, in this case cities. However, a nested model offers too little flexibility to deal with studies that apply a more detailed or continuous urban environment.

With the introduction of microscopic simulation models more attention was given to geographic detail; making the problem of continous space and complex interdependencies between spatial clustered alternatives more apparent. However, in the empirical literature on firm location choice very little work has been done in trying to make these spatial interdependencies explicit in the choice models. Most approaches ignore the spatial interdependencies of alternatives and apply a multinomial logit model and a random unweighted sample of alternatives as choice sets (Anderstig and Mattson, 1991; Shukla and Waddell, 1991; Waddell and Ulfarsson, 2003; de Bok and Sanders, 2005; Maoh and Kanaroglou, 2007).

In a recent contribution Elgar (2011) applied a form a stratified sampling procedure, the third type of choice set formation techniques. In this study explicit search areas are introduced into the estimation of firm location models around two anchor points: the current location and the owner location of the firm. Stratified sampling was applied by sampling fixed numbers of alternatives inside and outside the search area.

Only few examples exists, in which the spatial clustering of alternatives was made explicit in the choice model. Kim et al. (2008) is an example where count models are used to predict job locations in microscopic urban simulation model. These count models are free from the IIA assumption underlying the MNL model, but they have less capacity to predict large count situations, often found in employment location. The competing destinations model has been applied in de Bok (2007) and de Bok and Van Oort (2011) to model the location choice of relocating firms. The proximity measure that predicts the choice set probability is calculated from the spatial clustering of available firm locations. The estimated parameters for the proximity measures were significant in the models for most industry sectors, confirming the existence of spatial interdependecies between alternatives. However, the competing destinations model is primarily based on a proxy measure for choice set probabilities.

2.3 The concept of dominance in spatial choices

In many choice contexts, some alternatives are not taken into account by the decision maker since they are ‘‘dominated” by other alternatives. In general, an alternative *i* is dominated by another alternative *j* if i is “worse” than *j*, with respect to one or more characteristics, without being better with respect to any characteristic. The concept of dominance among alternatives has been used within Random Utility (RU) theory only in destination choice modelling (Cascetta and Papola, 2009) and in residential location choice (Cascetta et al, 2007). In this paper a general approach to extend and apply the concept of dominance among alternatives to RU theory is proposed for the first time to tacke with problem of choice set formation in the context of firm location but with a different meaning. A weighted selection of the alternatives, using dominance criteria as weights for the sampling probabilities, is here proposed. Before describing the methodology adopted, the concept of dominance is briefly reported in the following.

In general, Cascetta and Papola stated (2009) it is necessary to define two steps: a rule for defining (a) when an alternative *i* is dominated by another alternative *j* and (b) various possible ways of exploiting the dominance information about pairs of alternatives in the choice set simulation process.

Regarding (a), a simple rule is to define *i* dominating *j,* for decision maker *n*, if all ‘‘utility” or ‘‘quality” attributes Q_n (i.e. attributes with positive coefficients) are larger (not smaller) in *i* than in *j* and all cost attributes C_n (i.e. attributes with negative coefficients) are smaller (not larger) in *i* than in *j*, with at least one inequality strictly satisfied. Formally, a binary dominance variable y^n_{ij} can be defined equal to one if *i* is dominated by *j*:

$$y^n_{ij} = \left\{ 1 \quad if\ Q^n_{ki} \leqslant Q^n_{kj}\ and\ C^n_{hi} \geqslant C^n_{hj}\ \forall\ k,h \right\} \qquad (4)$$

and zero otherwise.

Stronger dominance rules can also be generated, for instance by introducing some ‘‘perception” thresholds (see for example Cantillo and Ortuzar, 2006) in the previously described comparisons of attribute values:

$$y^n_{ij} = \left\{ 1 \quad if\ Q^n_{ki} \leqslant \frac{1}{\alpha_k} Q^n_{kj}\ and\ C^n_{hi} \geqslant \alpha_h C^n_{hj}\ \forall\ k,h \right\} \qquad (5)$$

with α_h, α_k greater than 1 for any *h, k*.

Regarding b), dominance information about pairs of alternatives can be exploited both in deterministic and probabilistic choice set simulation approaches through the following steps. First, all the variables y^n_{ij} should be identified for any pair *ij* of alternatives and for any decision maker *n*. Then a global dominance degree dom^n_j of each alternative *j* and decision maker *n*, that is the number of alternatives within the choice set of all feasible alternatives for decion-maker *n* dominating *j*, can be defined as:

$$dom^n{}_j = \sum_{i \in CS^n} y^n{}_{ji} \qquad (6)$$

Note that different global dominance degrees can be identified according to the different possible ways of defining y^n_{ij} (see for example (1) and (2)). Subsequently, all alternatives dominated according to a certain y^n_{ij} definition (or the alternatives exhibiting a dominance degree greater than a certain threshold) can be deterministically excluded from the choice set of decision maker n or, alternatively, dominance attributes (like dom^n_j) can be introduced as perception attributes *Y* in any probabilistic choice set simulation approach. Obviously, the two approaches can also be combined, for instance defining an initial deterministic selection of perceived alternatives and then applying some probabilistic choice set simulation approach to this selection. However, stochastic application seems the most natural way of applying dominance information within RU theory.

3. A MODEL FOR PROBABILISTIC CHOICE SETS IN FIRM LOCATION MODELS

This study applies discrete choice models with probabilistic choice sets using the general form of equation (3) in the case of firm location choices. We will analyse different specifications of $p(d \in C)$. The concept of dominance is explored as an estimator for choice set membership and different specifications of this concept are tested. In addition to that, centrality is taken into account, following the notion that the spatial clustering affects the substitution pattern among alternatives (Fotheringham, 1983). First an overview is given of the whole choice set formation procedure.

3.1 Choice set formation

The firm location models are estimated on a dataset with choice sets of location alternatives that are considered to be a representative for the choice context of each observed firm relocation. The derivation of these choice sets is illustrated in the following figure. First of all, the universal set of firm locations comprises of all firm location that exist in the research area. This set is reduced to the set of locations that are actually available by reducing the universal set with the locations that are occupied by the existing firm population. In case of a relocating firm, only those alternatives are assumed to be relevant that meet a minimal size constraint following from the size of the decision maker, the relocating firm. This set of feasible alternatives is still a large set of alternatives that have to be brought down to a choice set with a limited but representative number of alternatives (20, including the chosen location). In many cases a random sampling approach is applied if the alternatives have equal probabilities of being in a choice set. In this case a probabilistic approach will be applied, in which the selection is based on a measure for choice set membership probability, such as dominance. The significance of this choice set membership is tested by estimating a parameter for the choice set membership proxies that are specified in the following paragraph.

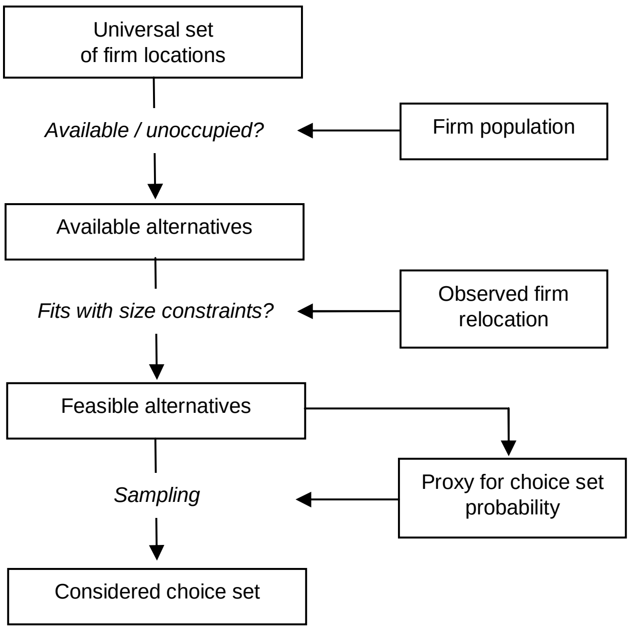


Figure 3-1: Choice set formation procedure

3.2 Choice set parameters

Centrality

The first proposition to measure competition between spatial clustered alternatives through a centrality measure was introduced by Fotheringham (1983). Based on the location of available alternatives, a centrality measure has been computed that measured the clustering of alternatives in each other's proximity. The closer two alternatives are located, the more likely they are to be substitutes to each other. This affects the choice probability of each individual alternative. The similarity between spatial alternatives is measured with a centrality measure that is a proxy for the spatial cluster membership. Following Fotheringham (1983), this centrality measure c_i is based on geographic space. The closer the alternatives are in space, the more likely they are to be substitutes for one another:

$$c_i = 1/(K-1)\sum_{k \neq i} w_k / d_{ik} \qquad (7)$$

where K is the number of available firm locations, d_{ik} is the distance between alternative k and i and w_k as the size of alternative k. The size of an alternative is specified as the available (unoccupied) floor space or industrial area at a firm location. So, for each alternative that is selected in the consideration set, the centrality relative to all other available location alternatives is computed. It is important to stress that c_i measures the clustering of *available* locations. In the literature, centrality is often measured relative to current activities instead of available alternatives (Pelligrini and Fortheringham, 2002). In those cases, centrality is similar to agglomeration. In this case, the model measures centrality relative to available firm locations. The influence of agglomeration economies is measured with the presented accessibility and agglomeration attributes.

The probability for choice set membership $p(i \in C)$ becomes:

$$p(i \in C) = (c_i)^{\theta} \qquad (8)$$

Parameter θ is estimated. Values < 0 indicate that alternatives that have many substitutes in proximity, have a high value for c_i, and have a smaller probability of being selected in a choice set.

Dominance

The dominance of location alternatives is determined at first a pairwise comparison of all alternatives in the universal choice set, see. The next step is to calculate the global degree of dominance. In the pairwise comparison different domination rules can be defined to determine if *i* is dominating *j*, given the available location attributes. First, a spatial dominance rule is defined considering the distance between two locations. Specifically, for relocating firm *n*, a binary dominance variable y^n_{ij} is defined equal to one if alternative *i* is dominated by *j*, if *i* is less close to *n*, compared to *j*:

$$y^n{}_{ij} = \begin{cases} 1 & if\ Dist_{ni} \geq Dist_{nj} \\ 0 & otherwise \end{cases} \tag{9}$$

The second dominance rule is based on the accessibility of different alternatives. An alternative is dominated by another alternative if it has a worse proximity to train and highway onramp and lower logsum accessibility values. Thus, in our case, *i* is dominated by *j* if the following inequities are valid and at least one is strictly satisfied:

$$y^n{}_{ij} = \begin{cases} 1 & if\ Q^n{}_{ki} \leq Q^n{}_{kj} \ \ \forall \ \ k \\ 0 & otherwise \end{cases} \tag{10}$$

Finally, the global dominance is defined on both the spatial location and accessibility. In this case i is dominated by j if the following conditions hold:

$$y^n{}_{ij} = \begin{cases} 1 & if\ Dist_{ni} \geq Dist_{nj} \quad and \quad Q^n{}_{ki} \leq Q^n{}_{kj} \ \ \forall \ \ k \\ 0 & otherwise \end{cases} \tag{11}$$

For each of these dominance rules, the dominance degree of each alternative *i* is computed by summing the dominance of *i* compared to all alternatives *j*. This measure gives the number of alternatives within the feasible choice set of decision-maker *n* that dominate *i*. Thus, a low value for dominance means a high dominant ranking:

$$dom^n{}_i = \sum_{j \in CS^n} y^n{}_{ij} \tag{12}$$

The probability for choice set membership $p(i \in C)$ becomes:

$$p(i \in C) = \left(dom^n{}_i\right)^{\theta} \tag{13}$$

Parameter θ is estimated. Values < 0 indicate that alternatives that are dominated by other alternatives, and thus have a high value for $dom^n{}_i$, have a smaller probability of being selected in a choice set.

4. APPLICATION TO CASE STUDY IN THE NETHERLANDS

4.1 The study area

The study area for which the empirical models in this article are estimated, is the Dutch province of South-Holland at the south-western edge of the Dutch Randstad region, which has a high population density (about 1,190 persons/km^2). See Figure 4-1; top left. It includes the second- and third-largest cities in the Netherlands (Rotterdam and The Hague) as well as numerous medium sized cities, such as Leiden, Delft, Schiedam, and Dordrecht. The longitudinal data cover the period 1988-1997, and after derivation of firm transition variables (growth, relocation) the dataset available for estimation covers the period 1990-1996. Henderson (1997) and Combes (2000) find that effects of agglomeration economies on economic growth peak after about 5 years and die out after 6-7 years. Thus, the interval over which relocation was measured appears to be sufficiently long to allow measurable differences over regions and locations to emerge.

The longitudinal firm data include variables for individual establishments from the Firm Register South-Holland, see fragment of research area on top right of Figure 4-1. These data are of interest for several reasons. First, the data include all establishments present in South-Holland in each year of the sample period. The dataset gives annual information on all establishments in the region (approximately 90,000). Establishments are enumerated based on information furnished by the Chamber of Commerce, insurance companies, and industrial sector associations and an annual questionnaire is sent to each. The average annual response rate to the questionnaire is 96%. Second, the data are available at a very fine scale. Questionnaire results identify each firm's 6-digit zip code (a small area containing about 100 different mailing addresses), and 5-digit activity code. These features are an advantage when testing for spatial externalities. The spatial scale at which the firm dynamics can be studied is very small, particularly when compared to U.S. counties or cities, which in some cases are defined as two or more contiguous counties. The entire area of South Holland measures 2,350 km^2.

In this article we focus on 7 sectors: manufacturing, construction, transport and distribution, finance, business services, government and general services. These industries are expected to have a different preference in terms of accessibility and agglomeration economies. In the estimation of location choice models very small firms (1 or 2 employees) we filtered out to avoid disturbances from 'empty' (purely administrative) firms.

Table 4-1: Number of relocating firms in the data set

	1990	1991	1992	1993	1994	1995	1996	Total
Observed relocations (firms > 3 employees)								
Manufacturing	119	95	111	95	106	106	121	753
Construction	153	112	149	141	144	165	169	1033
Transport, Warehousing and Comm.	143	122	131	85	142	113	160	896
Financial services	67	56	56	53	77	50	69	428
Business services	304	250	265	243	325	286	319	1992
Government	29	22	14	28	29	28	35	185
Consumer Services	58	58	66	58	62	73	66	441
Total	873	715	792	703	885	821	939	5728

To illustrate the size of the choice set formation problem, Table 4-2 gives some descriptive statistics on the size of feasible choice sets for each observed firm relocation, broken down across the firm categories. Where in previous studies we used a random sampling of alternatives to specify the choice sets (De Bok, 2007), in this study we will test if we can obtain more representative choice sets by applying a probabilistic approach, using dominance and centrality to estimate the probability an alternative belongs to a choice set.

Table 4-2: Descriptive statistics of choice sets: number of observations, average number of alternatives in feasible choice set and average size of sampled consideration sets

	# observations	average size feasible choice set	average size consideration set
Manufacturing	753	2465	20
Construction	1033	3408	20
Transport, Warehousing and Communication	896	2869	20
Financial services	428	4926	20
Business services	1992	4693	20
Government	185	2083	20
Consumer Services	441	4464	20
Total	5728	3798	20

4.2 Accessibility attributes

This subsection presents all the spatial attributes that are tested in the utility function of the choice models. The accessibility related attributes fall into three dinstinctive categories, as applied in De Bok van Oort (2011): infrastructure proximity, urbanization economies and localization economies.

Proximity measures

The proximity measures to transport infrastructure access points are calculated from the location of each firm (6 digit zip code) and a GIS analysis. In this analysis the distance is computed between the firm and the nearest highway onramp and the nearest train station. The distances to the nearest highway onramp and nearest train station are included as spatial attribute that were re-coded into a categorical variable describing the orientation of a location to the transport infrastructures. An α-location is a typical train stations location: within 800 meters of a train station and not too close to a highway onramp. Locations near highway onramps (within 2000 meters) are labeled as γ-locations. If a location is close to a train station as well as a highway onramp (within 800 meters and 2000 meters, respectively) it is labeled a β-location. If a location has a considerable distance to both the nearest train station and highway onramp, it is labeled a ρ-location. The models location type dummies are implemented with an effect-coding scheme so the parameter for ρ-locations can be derived from the estimated parameters for αβγ-locations. This typology resembles the policy-induced accessibility categorization of locations of the Dutch government (Schwanen et al, 2004).

Urbanisation economies

The logsum accessibility attributes (Figure 1, bottom left) and the travel time between zones in the study area are derived from the National Modeling System (NMS), the national transport model for the Netherlands (Hague Consulting Group, 2000). The model is developed by the Transport Research Centre, and is applied in a back casting study that made the travel times and logsums available back to 1985. The NTM is based on disaggregate discrete choice models, and provides the logsums for (non-home based) business trips and the (reflected) logsum for commuting trips. The logsum for business trips is assumed to be a representative measure for customer and supplier accessibility. This logsum is calculated as the sum of the trip utilities to all destination zones *d* for all person types *p* for all trips:

$$A_{om} = \log \sum_{d} \sum_{p} \exp(\mu_m \cdot V_{odpm}) \tag{5}$$

with purpose *m* = 'business trips', V_{odpm} as the expected utility for person type *p* to make a business trip from origin *o* to destination *d* and with the purpose specific scale parameter μ_m.

Labor market accessibility is derived from the utility of commuting trips in the transport model, from the perspective of the employer. For this reason, we analyze the commuting trips with a reflected logsum that measures the accessibility at the destination side of all commuting trips. The reflected logsum for commuting trips at firm location d, is specified as:

$$A_{dm} = \log \sum_{o} \sum_{p} \exp(\mu_m \cdot V_{odpm}) \tag{6}$$

with trip purpose *m* 'commuting' and V_{odpm} as the expected utility for person type *p* to commute from origin *o* to destination *d*.

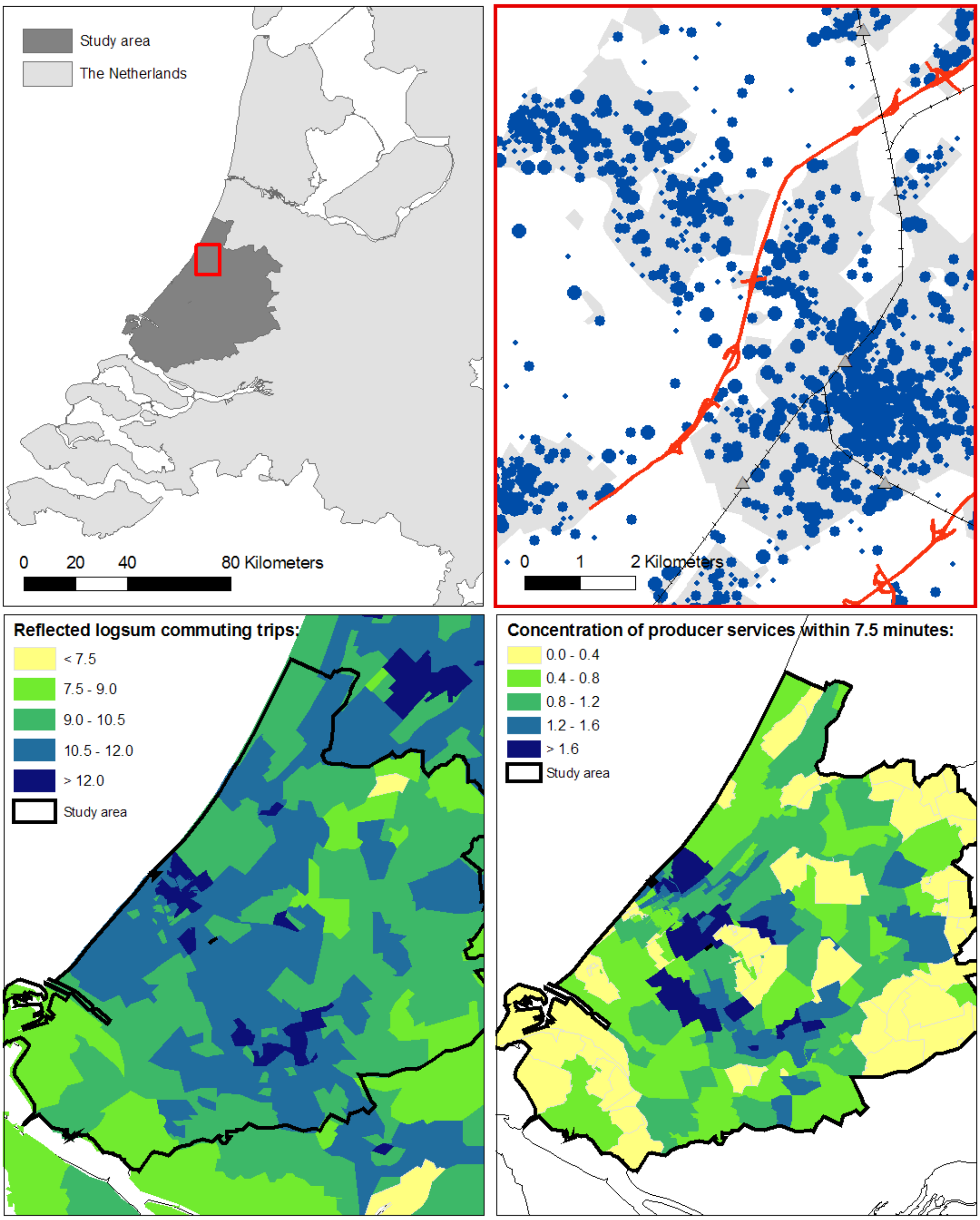
Localisation economies

The congested travel time matrices are calculated in the peak hour, and are used to determine a range band R_{jb} for each zone and to measure the composition of the firm population in this area. The composition is measured as the level of diversification or specialization within the range band from each location.

Concentration is measured as the representation of an industry within a specific travel range of a location relative to that industry's share in the region. We applied a range band of 7.5 minutes, as research by Van Soest et al (2006) reveals that the spatial reach and impacts of agglomeration externalities in South-Holland are limited in nature. For location *j,* the share of the employment in sector *s* in a range band R_{jb} from *j* is measured relative to the share of employment in that industry in the whole region. The production specialization index for location *j* and range band R_{jb} becomes:

$$PS_{jsb} = \frac{E_{sR_{jb}} \Big/ \sum_{s} E_{sR_{jb}}}{\sum_{j} E_{sR_{jb}} \Big/ \sum_{j} \sum_{s} E_{sR_{jb}}} \tag{7}$$

Figure 4-1: Clockwise from the top left: the research area, detail of the firm data and their proximity to transport infrastructures, specialisation index, and the logsum for commuting trips



with E_{sRjb} as the employment in industry within range band R_{jb}. The specialization of firms in business services is visualized in Figure 4-1, bottom right.

Diversity of the firm population within a range band is measured with the productivity diversity index (Paci and Usai, 1999). If S defines the number of industries and all industries are sorted in increasing order, then the production diversity index PD_{jb} for location j and range band R_{jb} is defined as:

$$PD_{jb} = \frac{1}{(S-1)E_{SR_{jb}}} \sum_{s=1}^{S-1} E_{sR_{jb}} \qquad (8)$$

with E_{sRjb} as the employment in the largest industry within range band R_{jb}.

5. RESULTS

Table 5-1 to Table 5-3 present the estimated location choice models by industry sector. For each sector, six models are presented:

- Model 0: Basic MNL model
- Model 1: Competing destinations model with probabilistic choice sets
- Model 2: Spatial dominance model with probabilistic choice sets
- Model 3: Accessibility dominance model with probabilistic choice sets
- Model 4: Global dominance model with probabilistic choice sets
- Model 8: Spatial dominance & competing locations model with probabilistic choice sets

The different model specifications are used to identify the probabilistic approach that yields the location choice model with the best statistical fit. First the general location preferences of firms is discussed. These results have been discussed and published in recent articles (De Bok, 2009, De Bok and Van Oort, 2011) so we will only highlight the main findings.

Location preferences

The estimated parameters for the four infrastructure proximity dummies reveal a significant and an industry-specific preference for infrastructure location types. Firms in manufacturing, construction and transport and distribution have a significant preference for highway locations (γ-locations). Proximity to the highway infrastructure is relevant because firms in these sectors have relatively large input and output flows. Firms in business services and financial services, also have positive parameters for highway proximity, but in particular when highway proximity is combined with proximity to train stations (β-locations). Proximity to train stations provides good accessibility to public transport for the commuters in this sector. Highway proximity allows convenient access for business trips to customers or for commuting.

Good accessibility, following a definition for urbanization economies, is tested with the estimated parameters for the logsum for commuting and business trips. The models do not reveal a preference for good accessibility for all industry sectors, but mainly for the industry sectors that are expected to be more dependent on a relatively high skilled labor force and more oriented on clients: business services and finance. For some other sectors, a higher accessibility might be a disadvantage, because in many cases this can lead to higher land prices. Previous research on location dynamics of firms in the Netherlands shows that firms in construction and manufacturing are likely to prefer less expensive locations, and a lower level of accessibility is then accepted (Van Oort, 2004).

Almost all relocating firms prefer specialized locations, with companies from their own industry sector. This result is interpreted as evidence for the existence of Marshallian externalities. Diversity has less significance as a location factor, and is perceived as a disutility for a location.

Finally, the estimated parameters for relocation distance are negative and significant in all model specifications. Thus, firms in all sectors prefer locations in the proximity of the original location attaching a significant disutility to the relocation distance. Similar results are reported in Pellenbarg et al. (2002) and Maoh and Kanaroglou (2007) for SME's in Canada. This result confirms the existence of keep-factors: relocating firms strive to maintain their existing spatial relations with employees, customers and suppliers.

Choice set composition

The improved final log-likellihood of the models, with parameters for probabilistic choice sets, prove that the model fit of location choice models is improved by the application of probabilistic choice sets. The specification of the probability of choice set membership is far from straightforward but the different approaches that were tested lead to some empirical conclusions on the identification of the optimal approach in this choice context.

In particular spatial dominance (equation (10)) proves to be a good predictor for choice set membership. Without exception the theta parameter for spatial dominance is significant and has a negative value. This means the less dominant an alternative is (high rank value) the lower the probability it is part of the choice set. To translate this to the perspective of the decision maker, from the available alternatives, it is more likely that the decision maker (the firm) takes alternatives closer by into consideration, possible because he is more likely to be aware the alternative is available. This is in line with the intervening opportunities concept formulated by Cascetta et al. (2007). The spatial dominance attribute is derived from the distance to the original location, which is also an attribute in the deterministic utility. However, both measures represent a different aspect of the choice process, and how they are formulated don't lead to multicollinearity. This can be checked by the signs of the estimated parameters.

The other two dominance parameters, where dominance was derived from the accessibility attributes (equation (10)) or the combination of accessibility and distance (equation (11)), do not improve models as much as the spatial dominance attribute. This result is likely to be caused by multicolinearity between the derivation from accessibility attributes that are also part of the deterministic utility, leading to inflation and the wrong sign for correlated parameters. This problem is for instance noticeable in Model 3 for the construction sector: accessibility dominance is significant but with the wrong sign, while the significance of correlated proximity attributes in the deterministic utility improve.

Clustermembership (centrality of alternatives) is another important estimator for choice set membership. This parameter was already present in previous studies (De Bok, 2009) and the results indicate the same: significant and negative sign of theta. This implies that clustered alternatives have more competition and a lower probability of choice set membership for the individual alternative.

Finally models are estimated including both the parameter for spatial dominance and that for clustermembership. These models prove to have the highest rho-square or best model fit for all sectors, except government. These results lead to the general conclusion that for firm relocation decisions more representative choice sets can be obtained by taking into account the spatial dominance of an alternative (is it close by relative to the other alternatives in the choice set) and spatial competition between alternatives (are many other alternatives available as equal substitutes).

Table 5-1: Estimation results Manufacturing, Construction, Transport & distribution

Industry sector:	**Manufactering**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dom.		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[km$^{1/2}$]	-1.67	-28.91 **	-1.69	-29.20 **	-0.94	-7.04 **	-1.67	-28.93 **	-1.60	-23.52 **	-0.97	-7.34 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	-0.15	-0.48	-0.12	-0.38	-0.19	-0.56	0.02	0.05	-0.32	-1.01	-0.16	-0.47
β-location; near trainstation & highway onramp [-]	-0.34	-1.65	-0.29	-1.36	-0.36	-1.65	-0.16	-0.46	-0.53	-2.42 *	-0.31	-1.39
γ-location; near highway onramp [-]	0.17	1.67	0.18	1.82	0.17	1.69	0.35	1.17	-0.04	-0.31	0.19	1.85
Urbanisation economies												
Logsum business and commuting trips [-]	-0.03	-0.64	0.06	1.00	-0.02	-0.45	-0.01	-0.22	-0.06	-1.23	0.06	1.07
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-0.76	-3.31 **	-0.56	-2.19 *	-0.70	-2.89 **	-0.76	-3.27 **	-0.77	-3.32 **	-0.51	-1.92
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	0.36	4.53 **	0.38	4.66 **	0.37	4.47 **	0.36	4.43 **	0.37	4.58 **	0.38	4.60 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-0.87	-2.69 **							-0.78	-2.49 *
Spatial dominance [-]					-0.42	-5.79 **					-0.42	-5.69 **
Accessibility dominance [-]							0.04	0.66				
Global dominance [-]									-0.08	-2.06 *		
Number of observations	754		754		754		754		754		754	
Null log-likelihood	-2,259		-2,259		-2,259		-2,259		-2,259		-2,259	
Init log-likelihood	-2,259		-2,259		-2,259		-2,259		-2,259		-2,259	
Final log-likelihood	-1,356		-1,351		-1,336		-1,355		-1,353		-1,332	
Rho-square	0.400		0.402		0.409		0.400		0.401		0.410	

Industry sector:	**Construction**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dom.		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[km$^{1/2}$]	-1.90	-34.73 **	-1.92	-34.25 **	-1.40	-10.35 **	-1.90	-34.73 **	-1.84	-29.20 **	-1.44	-10.67 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	-0.15	-0.51	-0.09	-0.31	-0.16	-0.54	0.68	1.71	-0.27	-0.94	-0.10	-0.36
β-location; near trainstation & highway onramp [-]	0.14	0.82	0.23	1.36	0.17	0.98	0.99	3.08 **	-0.02	-0.10	0.26	1.47
γ-location; near highway onramp [-]	0.20	2.36 *	0.24	2.69 **	0.21	2.39 *	1.09	3.64 **	0.04	0.30	0.24	2.72 **
Urbanisation economies												
Logsum business and commuting trips [-]	-0.08	-2.05 *	0.01	0.18	-0.08	-1.88	-0.02	-0.42	-0.11	-2.50 *	0.01	0.31
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-0.54	-2.48 *	-0.26	-1.13	-0.51	-2.31 *	-0.50	-2.27 *	-0.55	-2.53 *	-0.24	-1.03
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	0.48	4.94 **	0.40	3.94 **	0.45	4.74 **	0.44	4.52 **	0.49	5.06 **	0.38	3.79 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-1.12	-3.70 **							-1.05	-3.57 **
Spatial dominance [-]					-0.27	-3.97 **					-0.26	-3.83 **
Accessibility dominance [-]							0.16	3.14 **				
Global dominance [-]									-0.06	-1.90		
Number of observations	1,032		1,032		1,032		1,032		1,032		1,032	
Null log-likelihood	-3,092		-3,092		-3,092		-3,092		-3,092		-3,092	
Init log-likelihood	-3,092		-3,092		-3,092		-3,092		-3,092		-3,092	
Final log-likelihood	-1,682		-1,673		-1,672		-1,677		-1,680		-1,664	
Rho-square	0.456		0.459		0.459		0.458		0.457		0.462	

Industry sector:	**Transport and distribution**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dom.		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[km$^{1/2}$]	-1.54	-28.89 **	-1.62	-28.63 **	-0.62	-5.70 **	-1.54	-28.98 **	-1.45	-24.68 **	-0.78	-7.32 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	0.17	0.70	0.36	1.41	0.06	0.24	-0.42	-1.40	-0.06	-0.22	0.22	0.85
β-location; near trainstation & highway onramp [-]	0.16	0.95	0.39	2.26 *	0.15	0.86	-0.52	-2.17 *	-0.13	-0.72	0.38	2.12 *
γ-location; near highway onramp [-]	0.32	3.93 **	0.32	3.84 **	0.33	3.90 **	-0.40	-2.06 *	0.02	0.15	0.35	4.05 **
Urbanisation economies												
Logsum business and commuting trips [-]	-0.13	-3.97 **	0.05	1.22	-0.11	-4.03 **	-0.20	-6.45 **	-0.17	-5.23 **	0.05	1.34
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-0.29	-1.53	0.43	1.95	-0.38	-1.92	-0.45	-2.34 *	-0.34	-1.77	0.29	1.30
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	0.31	9.46 **	0.39	10.66 **	0.30	8.48 **	0.31	9.28 **	0.32	9.44 **	0.37	9.83 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-2.04	-7.81 **							-1.79	-7.41 **
Spatial dominance [-]					-0.54	-8.68 **					-0.49	-8.11 **
Accessibility dominance [-]							-0.14	-4.01 **				
Global dominance [-]									-0.11	-3.38 **		
Number of observations	897		897		897		897		897		897	
Null log-likelihood	-2,687		-2,687		-2,687		-2,687		-2,687		-2,687	
Init log-likelihood	-2,687		-2,687		-2,687		-2,687		-2,687		-2,687	
Final log-likelihood	-1,782		-1,746		-1,740		-1,775		-1,776		-1,711	
Rho-square	0.337		0.350		0.352		0.339		0.339		0.363	

Table 5-2: Estimation results, Business services, Finance, Government

Industry sector:	**Business services**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dom.		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[$km^{1/2}$]	-1.85	-44.21 **	-1.89	-43.96 **	-0.98	-12.79 **	-1.85	-44.20 **	-1.79	-38.80 **	-1.03	-13.71 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	-0.01	-0.07	0.16	1.02	0.01	0.06	-0.11	-0.60	-0.16	-0.99	0.17	1.01
β-location; near trainstation & highway onramp [-]	0.20	2.12 *	0.39	3.88 **	0.21	2.05 *	0.09	0.60	0.03	0.29	0.39	3.65 **
γ-location; near highway onramp [-]	0.17	2.62 **	0.15	2.34 *	0.17	2.53 *	0.05	0.37	-0.01	-0.11	0.15	2.31 *
Urbanisation economies												
Logsum business and commuting trips [-]	0.03	1.03	0.18	5.14 **	0.02	0.75	0.02	0.50	0.00	0.10	0.18	4.98 **
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-0.85	-5.63 **	-0.30	-1.80	-0.84	-5.31 **	-0.86	-5.68 **	-0.85	-5.62 **	-0.26	-1.50
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	0.42	6.76 **	0.55	8.23 **	0.45	6.88 **	0.41	6.53 **	0.39	6.12 **	0.59	8.35 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-1.28	-8.73 **							-1.24	-8.78 **
Spatial dominance [-]					-0.46	-12.26 **					-0.45	-12.25 **
Accessibility dominance [-]							-0.02	-0.96				
Global dominance [-]									-0.07	-3.15 **		
Number of observations	1,992		1,992		1,992		1,992		1,992		1,992	
Null log-likelihood	-5,967		-5,967		-5,967		-5,967		-5,967		-5,967	
Init log-likelihood	-5,967		-5,967		-5,967		-5,967		-5,967		-5,967	
Final log-likelihood	-3,628		-3,583		-3,552		-3,627		-3,622		-3,509	
Rho-square	0.392		0.400		0.405		0.392		0.393		0.412	

Industry sector:	**Financial services**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dominar		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[$km^{1/2}$]	-2.13	-21.76 **	-2.26	-20.47 **	-0.98	-6.07 **	-2.14	-21.84 **	-1.99	-18.97 **	-1.15	-7.10 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	0.41	1.41	0.62	1.97 *	0.48	1.46	0.05	0.13	0.18	0.61	0.67	1.95
β-location; near trainstation & highway onramp [-]	0.27	1.23	0.61	2.59 **	0.37	1.59	-0.20	-0.61	-0.06	-0.26	0.71	2.83 **
γ-location; near highway onramp [-]	0.35	2.49 *	0.41	2.77 **	0.38	2.58 **	-0.15	-0.54	-0.02	-0.09	0.46	2.99 **
Urbanisation economies												
Logsum business and commuting trips [-]	0.09	1.28	0.30	3.60 **	0.07	1.02	0.02	0.30	0.04	0.55	0.30	3.50 **
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-1.36	-3.58 **	-0.97	-2.50 *	-1.31	-3.20 **	-1.33	-3.47 **	-1.32	-3.48 **	-0.85	-2.03 *
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	0.16	1.82	0.37	4.11 **	0.14	1.52	0.12	1.31	0.13	1.38	0.34	3.61 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-2.12	-5.74 **							-2.00	-5.82 **
Spatial dominance [-]					-0.54	-7.24 **					-0.52	-7.18 **
Accessibility dominance [-]							-0.10	-1.92				
Global dominance [-]									-0.16	-2.78 **		
Number of observations	428		428		428		428		428		428	
Null log-likelihood	-1,282		-1,282		-1,282		-1,282		-1,282		-1,282	
Init log-likelihood	-1,282		-1,282		-1,282		-1,282		-1,282		-1,282	
Final log-likelihood	-687		-667		-665		-686		-683		-647	
Rho-square	0.464		0.479		0.481		0.465		0.468		0.495	

Industry sector:	**Government**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dominar		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[$km^{1/2}$]	-1.93	-13.05 **	-1.94	-12.94 **	-1.43	-4.95 **	-1.93	-13.13 **	-1.95	-11.93 **	-1.44	-5.05 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	0.61	1.36	0.63	1.38	0.56	1.19	1.14	2.00 *	0.65	1.42	0.58	1.21
β-location; near trainstation & highway onramp [-]	0.87	2.99 **	0.90	2.92 **	0.88	2.98 **	1.43	2.98 **	0.92	2.84 **	0.90	2.92 **
γ-location; near highway onramp [-]	0.44	2.17 *	0.45	2.16 *	0.44	2.13 *	1.01	2.44 *	0.48	1.92	0.44	2.12 *
Urbanisation economies												
Logsum business and commuting trips [-]	0.00	0.03	0.02	0.15	0.01	0.10	0.09	0.69	0.01	0.10	0.03	0.21
Diversity attributes												
Diversity Rb < 7,5 min. [-]	0.05	0.11	0.11	0.25	0.09	0.20	0.02	0.04	0.04	0.10	0.15	0.32
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	-0.13	-0.96	-0.09	-0.65	-0.11	-0.85	-0.09	-0.71	-0.12	-0.95	-0.08	-0.58
Parameters probabilistic choice sets:												
Theta; centrality [-]			-0.18	-0.40							-0.16	-0.37
Spatial dominance [-]					-0.25	-1.98 *					-0.25	-1.98 *
Accessibility dominance [-]							0.13	1.62				
Global dominance [-]									0.02	0.28		
Number of observations	185		185		185		185		185		185	
Null log-likelihood	-554		-554		-554		-554		-554		-554	
Init log-likelihood	-554		-554		-554		-554		-554		-554	
Final log-likelihood	-350		-350		-348		-349		-350		-348	
Rho-square	0.368		0.368		0.371		0.370		0.368		0.371	

Table 5-3: Estimation results general services

Industry sector:	**General services**											
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 8	
Model label:	MNL-Basic		Competing destinatio		Spatial dominance		Accessibility dominar		Global dominance		Spatial dom. & CD	
Choice set:	Random sampling		Probabilistic		Probabilistic		Probabilistic		Probabilistic		Probabilistic	
	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val	β	t-val
Utility attributes:												
Distance to original loc.[km$^{1/2}$]	-2.19	-22.46 **	-2.23	-22.68 **	-1.16	-6.52 **	-2.19	-22.51 **	-2.10	-19.60 **	-1.21	-7.10 **
Spatial dominance [-]												
Infrastructure proximity												
α-location; near trainstation [-]	0.59	1.67	0.72	1.99 *	0.65	1.79	0.40	0.88	0.38	1.01	0.75	2.03 *
β-location; near trainstation & highway onramp [-]	0.46	1.93	0.59	2.27 *	0.48	1.93	0.27	0.73	0.23	0.87	0.58	2.18 *
γ-location; near highway onramp [-]	0.37	2.61 **	0.43	2.91 **	0.36	2.44 *	0.18	0.58	0.13	0.70	0.41	2.69 **
Urbanisation economies												
Logsum business and commuting trips [-]	0.19	2.50 *	0.26	3.08 **	0.19	2.38 *	0.17	1.89	0.14	1.79	0.25	2.84 **
Diversity attributes												
Diversity Rb < 7,5 min. [-]	-0.77	-2.29 *	-0.50	-1.39	-0.71	-1.95	-0.78	-2.32 *	-0.77	-2.27 *	-0.47	-1.21
Specialisation attributes												
Specialisation Rb < 7,5 min. [-]	-0.57	-4.09 **	-0.47	-3.24 **	-0.51	-3.61 **	-0.58	-4.13 **	-0.64	-4.43 **	-0.42	-2.74 **
Parameters probabilistic choice sets:												
Theta; centrality [-]			-0.75	-2.20 *							-0.61	-1.88
Spatial dominance [-]					-0.50	-5.90 **					-0.49	-5.92 **
Accessibility dominance [-]							-0.04	-0.68				
Global dominance [-]									-0.10	-1.93		
Number of observations	441		441		441		441		441		441	
Null log-likelihood	-1,321		-1,321		-1,321		-1,321		-1,321		-1,321	
Init log-likelihood	-1,321		-1,321		-1,321		-1,321		-1,321		-1,321	
Final log-likelihood	-695		-692		-677		-695		-693		-675	
Rho-square	0.474		0.476		0.488		0.474		0.476		0.489	

6. CONCLUSIONS AND FURTHER PERSPECTIVES

The estimation results have showed that the model fit of location choice models is improved by the application of probabilistic choice sets. The specification of the probability of choice set membership is far from straightforward but the different approaches that were tested lead to some empirical conclusions on the identification of the optimal approach in the choice context of disaggregate urban simulation models.

The results have lead to the general conclusion that for firm relocation decisions more representative choice sets can be obtained by taking into account the spatial dominance of an alternative (is it close by relative to the other alternatives in the choice set) and spatial competition between alternatives (are many other alternatives available as equal substitutes).

A general observation is made about the implementation of the probabilistic approach into a microscopic urban simulation model. A probabilistic sampling of alternatives, based on dominance either centrality, requires an additional computation step: a pair wise comparison over all alternatives in the feasible choice set of a relocating firm. Since these feasible choice sets are determined for each relocating firm, and the size of this choice set can be considerable (in this example on average about 4000 alternatives, see Table 4-2) this can introduce significant additional computation time, compared to an unweighted random sampling of alternatives.

REFERENCES

Adler TJ, Ben-Akiva M, 1976 "Joint-choice model for frequency, destination and travel mode for shopping trips" *Transportation Research Record* **569** 136-150

Anderstig C, Mattson LG, 1991 An Integrated Model of Residential and Employment Location in a Metropolitan Region *Papers in Regional Science* **70** 167–184

Bierlaire M, Bolduc D, McFadden D, 2006 "The estimation of Generalized Extreme Value models from choice-based examples." Report TRANSP-OR 060810, Transport and Mobility. Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Federale de Lausanne

Cantillo V, Ortuzar J de D, 2006 "Implications of thresholds in discrete choice modeling" *Transport Reviews* **26** 667-691

Cascetta E, Papola A, 2001 "Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand" *Transportation Research C* **9** 249-263

Cascetta and Papola 2009 "Dominance among alternatives in random utility models" *Transportation Research Part A: Policy and Practice* **43** 170-179

Cascetta E, Pagliara F, Axhausen KW, 2007 "The use of dominance variables in spatial choice set generation" Proceedings of the 11th World Conference on Transport Research, Berkeley, 24-28 June

Combes PP, 2000 "Economic structure and local growth: France 1984-1993" *Journal of Urban Economics* **47** 329-355

De Bok M, 2009 "Estimation and validation of a microscopic model for spatial economic effects of transport infrastructure" *Transportation Research Part A: Policy and Practice* **43** 44-59

De Bok M, Sanders F, 2005 "Firm relocation and the accessibility of locations: empirical results from The Netherlands" *Transportation Research Record: Journal of the Transportation Research Board* **1902** 35-43

De Bok M, van Oort F, 2011 "Agglomeration economies, accessibility and the spatial choice behavior of relocating firms" *Journal of transport and Landuse*. **4(1)** 5-24

Elgar I, 2011 *Modelling office mobility and location decisions in a microsimulation environment* Ph.D. dissertation, University of Toronto (Canada), Canada. Retrieved February 21, 2011, from Dissertations & Theses @ University of Toronto.(Publication No. AAT NR39401)

Fortheringham AS, 1983 "A new set of spatial-interaction models: the theory of competing destinations" *Environment and Planning A* **15** 15-36

Gautschi DA, 1981 “Specification of patronage models for retail center choice” *Journal of Marketing Research* **18** 162-174

Golledge RG, Timmermans H, 1990 “Application of behavioral research on spatial problems I: Cognition” *Progress in Human Geography* **14** 57–99

Hägerstrand T, 1970 “What about people in regional science?” *Papers of the Regional Science Association* **24** 7-21

Hansen ER, 1987 “Industrial location choice in São Paulo, Brazil: a nested logit model” *Regional Science and Urban Economics* **17** 89-108

Henderson JV, 2007 “Understanding knowledge spillovers” *Regional Science and Urban Economics* **37** 497-508

Kim H, Waddell P, Shankar VN, Ulfarsson GF, 2008 ”Modelling Micro-Spatial Employment Location Patterns: A Comparison of Count and Choice Approaches” *Geographical Analysis* **40** 123-151

Landau U, Prashker JN, Alpern B, 1982 “Evaluation of activity constrained choice sets to shopping destination choice modeling” *Transportation Research Part A: Policy and Practice* **16 (3)** 199–207

Manski C, 1977 “The structure of random utility models” *Theory and Decision* **8** 229-254

Maoh M, Kanaroglou P, 2007 “Business establishment mobility behavior in urban areas: a microanalytical model for the City of Hamilton in Ontario, Canada” *Journal of Geographic Systems* **9(3)** 229-252

McFadden D, 1978 “Modeling the choice of residential location” in: Karlqvist A, Lundqvist L, Snickars F, Weibull JW (Eds) *Spatial interaction theory and residential location* Amsterdam: North-Holland 75-96

Miller EJ, O’Kelly ME, 1983 “Estimating shopping destination models from travel diary data” *Professional Geographer* **35** 440-449

Paci R, Usai S, 1999 “Externalities, knowledge spillovers and the spatial distribution of innovation” *GeoJournal* **49** 381-390

Pagliara F, Timmermans H, 2009 “Choice set generation in spatial contexts: a review” *Transportation Letters: The International Journal of Transportation Research* **1** 181-196

Pellenbarg P, van Wissen LJG, van Dijk J, 2002 “Firm migration” in: McCann P (Eds) *Industrial location economics*. Cheltenham: Edward Elgar, pp 110-148

Pelligrini PA, Fortheringham AS, 2002 “Modeling spatial choice: a review and synthesis in a migration context” *Progress in Human Geography* **26** 487-510

Scott DM, 2006 “Constrained destination choice set generation: a comparison of GIS-based approaches” Proceedings of the 85th annual meeting on Transportation Research Board, Washington

Shukla V, Waddell P, 1991 “Firm location and land use in discrete urban space: A study of the spatial structure of Dallas-Fort worth” *Regional Science and Urban Economics* **21** 225-253

Southworth F, 1981 “Calibration of multinomial logit models of mode and destination choice” *Transportation Research Part A* **15** 315-325

Thill JC, 1992 “Choice set formation for destination choice modeling” *Progress in Human Geography* **16** 361-382

Thill JC, Horowitz JL, 1997 “Travel- time constraints on destination choice sets” *Geographical Analysis* **29(2)** 108-123

Van Oort, FG, 2004 *Urban Growth and Innovation; Spatially Bounded Externalities in the Netherlands*. Aldershot: Ashgate

Van Soest DP, Gerking S, Van Oort FG, 2006 “Spatial impacts of agglomeration externalities” *Journal of Regional Science* **46** 881-899

Waddell P, Ulfarsson GF, 2003 “Accessibility and Agglomeration: Discrete-Choice Models of Employment Location by Industry Sector” Conference proceedings of the 82nd Annual Meeting of the Transportation Research Board, Washington DC

Weisbrod G, Parcells RJ, Kern C, 1984 “A disaggregate model for predicting shopping area market attraction” *Journal of Retailing* **60(1)** 65–83