Dominance variables and intervening opportunities for choice set generation

Author(s):
Cascetta, Ennio; Pagliara, Francesca; Axhausen, Kay W.

Publication Date:
2007

Permanent Link:
https://doi.org/10.3929/ethz-a-005564736

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
DOMINANCE VARIABLES AND INTERVENING OPPORTUNITIES FOR CHOICE SET GENERATION

Ennio Cascetta*, Francesca PAGLIARA* and Kay W. Axhausen**

* University of Naples Fedrico II - Italy
**IVT, ETH – Zurich

ABSTRACT
Models adopted in the literature to represent spatial choices are generally rather elementary and result in the application of random utility theory to the choice among hundreds of alternatives. The attributes are usually related to spatial attractiveness and to generalised travel cost without any reference to perception/availability attributes. The objective of this paper is twofold: to use perception/availability variables named dominance variables for modelling spatial choices to have a better predictive model and to use dominance criteria as weights for the sampling probabilities to show how weighted sampling of alternatives provide parameters estimates “closer” to the full choice set.

KEYWORDS
Choice set, spatial choices, dominance variables, random utility models, sampling techniques

INTRODUCTION
Spatial decisions and processes are fundamental to the understanding of spatial structure. In the earlier stages of spatial analysis proposed explanations were typically on the aggregate level. Now a new field has evolved and matured which calls for a deeper understanding of spatial structure with a particular emphasis on spatial decisions and processes. The book by Fisher et al. (1990), Spatial Choices and Processes, discusses a wide variety of new modelling approaches, techniques and issues related to spatial decision and processes.

However, the starting point has been the spatial interaction approach, which can be broadly defined as movement or communication over space that results from a decision process. The term thus encompasses such diverse behaviour as migration, shopping, residing, travel-to-work, the choice of health-care services, recreation, the movement of goods, telephone calls, the choice of a university by students, airline passenger traffic, and even attendance at events such as conferences, theatre and football matches (Thill, 1995). In each case, an individual trades off in some manner the benefit of the interaction (the purchase of goods at a store, for example) with the costs that are necessary in overcoming the spatial separation between the individual and his/her possible destination. Olsson (1970) states that “The concept of spatial interaction is central for everyone concerned with theoretical geography and regional science…..Under the umbrella of spatial interaction, it has been possible to accommodate most model work in transportation, migration, commuting, and diffusion, as well as significant aspects of location theory”.

Most spatial interactions result from some sort of spatial choice, whether it be an individual consumer selecting a store, an household selecting a residence or a migrant selecting a city in which to live.
Consequently, there is a strong relationship between modelling spatial interaction and modelling spatial choice (Fotheringham and O’Kelly, 1989).

The discrete choice framework provides an alternative theoretical justification for the form of the two singly-constrained spatial interaction models. The production-constrained interaction model represents the choices of destinations (e.g. shops, cities, etc.), while the attraction-constrained model represents the choices of origins (e.g. residences) (Wilson, 2000). However, the discrete choice framework provides a more behavioural framework for understanding the rationale of these models.

Understanding the behavioural foundation of spatial choice models allows the identification of possible shortcomings in this foundation and can lead to improvements in model formulation. A demonstration of this is provided in terms of modelling hierarchical destination choice which is used to suggest a more general framework in which to model spatial choice and hence spatial interaction. The Competing Destination model by Fotheringham represents an example (1983; 1988).

Essentially, the basic difference between spatial interaction and spatial choice models is one of usage: by consensus, the term spatial interaction is applied to aggregate flows whereas the term spatial choice is applied to an individual selection of a location. Clearly, aggregate flows are the result of a collection of individual decisions so the two are inextricably linked: the variables that explain the spatial choice of an individual tend to be very similar to the variables that explain the spatial choices of a large number of individuals.

The models adopted are generally rather elementary and result in the application of random utility theory to the choice among hundreds or even thousands of physical alternatives. The attributes are usually related to spatial attractiveness (e.g. number of services available) and to generalised travel cost or disutility without any reference to choice set formation process or perception/availability attributes.

In this paper, the objective is twofold: to use dominance variables, i.e. variables reflecting the spatial position and hierarchies of alternatives, for modelling spatial choices in order to have a better predictive model and to use dominance criteria as weights for the sampling probabilities to show how weighted sampling of alternatives provide parameters estimates “closer” to the full choice set.

The next section defines the use of dominance variables to model spatial choices within random utility theory. The third section provides the application of dominance variables in the residential location choice as an example of spatial choices. The fourth section discusses the use of dominance criteria in sampling techniques. Conclusions and further perspectives are reported in the last section.

THE CONCEPT OF DOMINANCE WITHIN RANDOM UTILITY MODELS

Random utility models are widely used to analyze choice behaviour and predict choices among discrete sets of alternatives. These models are based on the assumption that an individual’s preference among the available alternatives can be described with a utility function and that the individual selects the alternative with the highest utility (Ben-Akiva and Lerman, 1985; Cascetta, 2001; Arentze and Timmermans, 2003).

In many choice contexts, it may happen that some alternatives are not taken into account by the decision maker since they are dominated by other alternatives. In general, an alternative \( d \) is dominated by another alternative \( d^* \) if \( d \) is “worse” than \( d^* \), with respect to one or more characteristics, without being better with respect to any characteristic. The concept of dominance among alternatives is widely recognized in project evaluation, e.g. through Multi-Criteria Decision-Making (MCDM) (Haimes and Chankong, 1985) where
dominated projects are excluded from the choice set on the basis of the principle of rationality (transferability of preferences). Instead, it has never been used within random utility theory.

In this paper a general approach to extend and apply the concept of dominance among alternatives to RU theory is proposed, defining (a) when an alternative \( d \) is dominated by another alternative \( d^* \) (b) possible ways of exploiting the dominance information about pairs of alternatives in the choice set generation process and (c) the use of dominance criteria as weights for the sampling probabilities.

Concerning (a) some dominance rules have been proposed. Specifically, it is assumed that an alternative \( d \) dominates an alternative \( d^* \) (for a decision-maker moving from origin zone \( o \)) if the attractiveness of \( d \) is greater than that of \( d^* \) and at the same time the generalised costs \( c_{od} \) are smaller than \( c_{od^*} \) (global dominance rule). Moreover, a spatial domination can be constructed which recalls the concept of intervening opportunities (Stouffer, 1960; Goncalves and Ulyssea Neto, 1993; Almeida and Goncalves, 2001). It is assumed that \( d \) spatially dominates \( d^* \) if it dominates \( d^* \) in relation to the above conditions and \( d \) is along the path to reach \( d^* \) from the individual origin \( o \) (i.e. if the length of the shortest path \( od^* \) is close to that of the shortest path \( od^* \)) (spatial dominance rule). In this case, \( d \) represents an intervening opportunity along the path, or bundle of paths, towards \( d^* \). In Fig. 1 an example of spatial domination is reported:

![Fig. 1 – Example of spatial domination](image)

Dominance variables are obtained through combinations of the previous rules. Therefore, a dominance value will be assigned to each alternative. The new variable can be defined in several ways: it can be a Boolean variable, it can be a variable taking values between 0 and 1; it can be the number of times an alternative is dominated by the others as it will be represented in this paper.

Dominance variables can affect the choice set formation process (Cascetta and Papola, 2005). The importance of properly specifying the choice set to avoid biasing model parameters has been well recognized in the literature (Pellegrini et al., 1997; Dai, 1998). This topic is particularly relevant for spatial choice models where alternatives are generally numerous and somewhat artificially defined (i.e. traffic zones). The choice set refers to the set of discrete alternatives considered by an individual in the decision-making process which is a subset of the universal choice set that consists of all alternatives available to the decision-maker (Nedugadi, 1987). The traditional approaches to choice sets generation typically are based on a restricted list of deterministic criteria selected by the analyst, that reflect available information as well as a priori beliefs about behaviour.
Other approaches to delineate choice sets are behavioural, which can be explicit or implicit. The first one explicitly simulates the choice set with a specific model, while the second one simulates the availability/perception of an alternative within the choice model of the alternative itself.

Given a universal choice set $S$ including $M$ alternatives, different individuals may consider different choice sets which include only a subset of $S$. Stochastic simulation of the choice set should provide the probability that each of these possible subsets $C$ actually occurs for an individual $p(C)$. The choice probability of the generic alternative $d$ therefore becomes:

$$
p(d) = \prod_{C,i} p(d / C)[X] \cdot p(C)[Y]
$$

which is Manski's (1977) formulation. In equation (1), $X$ and $Y$ are attributes influencing the utility and availability/perception of the alternatives, respectively. Manski's approach requires explicit enumeration of the $2^M-1$ possible choice subsets given $S$, which is troublesome for a large number of alternatives.

The main issue is to find proper specification for the $p(C)$'s in terms of models and attributes. Some authors propose simulating the $p(C)$'s as a function of the sole "utility" attributes of the underlying choice alternatives, i.e. not taking into account specific availability/perception attributes. Relevant examples are the works by Ben Akiva and Boccara (1995), Horowitz and Louviere (1995) and by Swait (2001).

The contribution of this paper consists in proposing new perception attributes (dominance variables), which can be used in any spatial choice context in general and in the residential location choice context in particular. These attributes, which may obviously be introduced whatever the choice set simulation model, have been tested on the IAP (Implicit Availability-Perception) RU model (Cascetta and Papola, 2001). Indeed, in this model, choice set enumeration is avoided by simulating the probability of an alternative $d$ belonging to the choice set, $p(d \in C)$, and by introducing the logarithm of $p(d \in C)$ in the utility of that alternative

$$U_d = U_x + \ln p(d \in C) = \sum \beta X + \sum \gamma Y + \varepsilon_d
$$

The rationale of model (2) is that a lower probability of an alternative $d$ belonging to the choice set reduces the $U_d$ and hence the $p(d)$.

The proposed perception attributes have been also introduced directly in the alternatives’ utilities of an MNL model in order to test the model predictive ability, it follows that $p(d \in C)[Y_d]$ is expressed as:

$$p(d \in C)[Y_d] = \frac{\exp(\gamma Y_d)}{\sum \exp(\gamma Y_d)}$$

then it results that $\ln p(d \in C)[Y_d] = \gamma Y_d + \cos t$ and then the utility becomes:

$$U_d = \sum \beta X + \sum \gamma Y + \varepsilon_d
$$

where $\beta$ and $\gamma$ are coefficients of utility and availability/perception attributes respectively.

As the IAP model and the MNL model have provided the same parameters estimation results (Cascetta and Papola, 2005), in this paper a MNL models has been considered (see third section).
Concerning (c) dominance criteria have been further used as weights for the sampling probabilities. Sampling techniques are applied techniques used to avoid the computational burden involved in estimating choice models with a large number of alternatives (Bierlaire et al., 2006). The objective is to show that weighted sampling gives parameters’ estimates “closer” to those obtained with full choice set (see the fourth section).

APPLICATION TO SPATIAL CHOICE MODELS

In this paper this methodology is applied to the context of residential location choice.

Models of residential location choice are important tools for analyzing urban housing policies, transportation planning policies, and urban social spatial structure and are represented in the transportation planning, urban economic, sociology, and urban geography literature. For transportation planning, residential location choice models are useful for evaluating how households are likely to alter the location of their residences in response to changes in regional demographics, housing policy, transportation service and policy, and location of employment opportunities. Household residential location choices are a function of a wide range of spatial attributes, the taste for which is differentiated by a variety of household characteristics. The differentiation identifies and characterizes the relative importance of different attributes to various types of households and the desire to reside in areas with others similar social characteristics.

Various factors have been found to influence people’s residential location choices. It has long been a challenge to determine these factors and the degree of their influence. The spatial analysis, at the disaggregate level, considers the decision-maker who decides to locate his/her residence within the urban area under study. According to the random utility approach, the utility of an alternative is expressed as a function of the attributes of the alternative and characteristics of all possible factors that may influence residential location choice.

There is a large number of studies of residential location choice behaviour in urban areas. The pioneer was McFadden (1978) who considered the problem of translating the theory of economic behaviour into models suitable for the empirical analysis of housing location. Studies like the ones of Sermons and Koppelman (1998); Wardman et al. (1998); Cooper et al. (2001); Simmonds and Skinner (2001); Bhat and Guo (2004); Kim et al. (2005) and others can be referred to for a review of the main factors influencing residential location choice. All of them have been useful in defining the explanatory variables employed in this paper.

Model estimation

In 2005 an RP survey was conducted in the canton of Zurich in Switzerland covering the mobility and moving biography of the respondents (see Beige and Axhausen, 2005; Beige, 2006; for details on the instrument and fieldwork). A sample of 1100 residents was obtained. Among them 658 respondents were considered useful for our purpose on the basis of those living and working within the canton of Zurich. For each resident included it is known the respondent’s residential place and workplace, the age, income, number of household members. Residents considered are both those living in a zone and working in another and those living and working in the same zone of the canton. The sample included also residents working outside the canton of Zurich. The study area has been divided in 182 traffic zones (of which 12 make up the municipality of Zurich) that represent the universal choice set of the model. The zonal data was obtained from an IVT database described in Tschopp et al. (2003).

The residential location model specified is a Multinomial Logit model according to equation (3) and the utility variables considered are
Price\(_d\) is the average land price of zone \(d\);  
\(\ln\text{Stock}_{d}\) is the natural logarithm of the housing stock in zone \(d\);  
\(\text{Logsum}_{od}^{LM}\) is the logsum of the mode choice model for work purpose for low-medium income residents; (attributes are of the mode choice and reference to these models) (Vrtic et al., 2005);  
\(\text{Logsum}_{od}^{H}\) is the logsum of the mode choice model for work purpose for high income residents;  
\(\ln\text{Workplaces}_{serv}\) is the natural logarithm of the workplaces in services (summation of retail, leisure and services to the households such as education, health) in zone \(d\) and it represents a measure of the quality of services to households in the zone itself.

The availability/perception variables have been obtained through a combination of the rules defined in the second section and they are:

**Dom1\(_d\)** is the number of zones \(d^*\) strongly dominating, i.e:
(a) \(d^*\) has average land price lower than \(d\);  
(b) the distance from the respondent’s workplace zone \(d^*\) (\(\text{dist}_{od^*}\)) is shorter than that to \(d\) (\(\text{dist}_{od}\));  
(c) \(d^*\) is along the path to reach the respondent’s workplace zone \(d\) from \(o\): \(\text{dist}_{od^*} + \text{dist}_{d^*d} < 1.2 \times \text{dist}_{od}\)

**STRONG GLOBAL DOMINANCE RULE**

**Dom2\(_d\)** is the number of zones \(d^*\) for which conditions (a) and (b) simultaneously occurs.  
**WEAK GLOBAL DOMINANCE RULE**

**Dom3\(_d\)** is the number of zones \(d^*\) strongly dominating, i.e. satisfying conditions (b) and (c) simultaneously.  
**STRONG SPATIAL DOMINANCE RULE**

**Dom4\(_d\)** is the number of zones \(d^*\) weakly dominating, i.e. satisfying only condition (b).  
**WEAK SPATIAL DOMINANCE RULE**

The calibration of the MNL model has been carried out with the software BIOGEME version 1.4 (Bierlaire, 2005). Calibration results are reported in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Std.dev.</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price(_d)</td>
<td></td>
<td>630.93</td>
<td>610</td>
<td>300</td>
<td>1560</td>
<td>276.14</td>
<td>CHF</td>
</tr>
<tr>
<td>lnStock(_d)</td>
<td></td>
<td>7.27</td>
<td>7.29</td>
<td>4.65</td>
<td>10.68</td>
<td>1.26</td>
<td>SQM</td>
</tr>
<tr>
<td>Logsum(_{od}^{LM})</td>
<td></td>
<td>-0.46</td>
<td>-0.21</td>
<td>-2.50</td>
<td>0.64</td>
<td>0.55</td>
<td>REAL</td>
</tr>
<tr>
<td>Logsum(_{od}^{H})</td>
<td></td>
<td>-0.35</td>
<td>0</td>
<td>-2.54</td>
<td>0.64</td>
<td>0.52</td>
<td>REAL</td>
</tr>
<tr>
<td>lnWorkplaces(_{serv})</td>
<td></td>
<td>5.69</td>
<td>5.56</td>
<td>2.19</td>
<td>9.91</td>
<td>1.72</td>
<td>INTEGER</td>
</tr>
<tr>
<td>Dom1(_d)</td>
<td></td>
<td>3.52</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>6.47</td>
<td>INTEGER</td>
</tr>
<tr>
<td>Dom2(_d)</td>
<td></td>
<td>20.54</td>
<td>11</td>
<td>0</td>
<td>147</td>
<td>24.86</td>
<td>INTEGER</td>
</tr>
<tr>
<td>Dom3(_d)</td>
<td></td>
<td>23.28</td>
<td>17</td>
<td>0</td>
<td>151</td>
<td>21.68</td>
<td>INTEGER</td>
</tr>
<tr>
<td>Dom4(_d)</td>
<td></td>
<td>87.43</td>
<td>87</td>
<td>0</td>
<td>181</td>
<td>52.86</td>
<td>INTEGER</td>
</tr>
</tbody>
</table>

Table 1 - Reports the descriptive statistics of the variables
In general, it is possible to state that models with dominance variables perform better with respect to model without a decrease of \( \ln(\theta) \) up to 38\%, which provides a better predictive model.

Weak dominance variables seem to perform better than strong dominance variables (e.g. \( \text{Dom}4 \); spatial...
dominance variables perform better than global dominance variables (e.g. Dom3 and Dom4) which tests how important is to consider the spatial component is such choice contexts and the best performances are obtained by combing weak spatial and strong global dominance variables together (i.e. Dom1 + Dom4).

APPLICATION OF DOMINANCE TO SAMPLING TECHNIQUES

Sampling techniques are applied techniques used to avoid the computational burden involved in estimating choice models with a large number of alternatives (Bierlaire et al., 2006).

Chapters 8 and 9 of the book by Ben-Akiva and Lerman (1985) are prototypical since they provide all the alternative methods of sampling alternatives and corresponding estimators for the choice model parameters.

In this paper a simple random selection of the alternatives and a weighted selection of the alternatives, using dominance criteria as weights for the sampling probabilities, are proposed.

In the first one, at each drawing each alternative has equal probability of being selected. In the second approach, dominance variables have been used as inverse sampling weights, i.e:

\[
p(d \in C) = \frac{w_d}{\sum w_d} = \frac{1/Dom_g}{\sum 1/Dom_g} \tag{4}
\]

The sampling is performed with replacement, but an alternative has been only included once in the choice set. Five different choice sets were considered with different number of alternatives (100, 75, 50, 25 and 20 respectively) and the corresponding models have been estimated. Estimations results obtained by using both Dom1 and Dom4 as inverse sampling weights have been compared with a simple random selection of the alternatives. The parameters of the weighted sampling models present always higher values compared to a simple random selection of the alternatives as well as the improvement in the goodness-of-fit of the different specifications. The three models have been further compared with the basic model which considers the universal set (182 alternatives). Comparisons between the vectors of coefficients has been carried out through the following statistics:

\[
\left(\hat{\beta}_{BASIC} - \hat{\beta}_{WEIGHT}\right)^T \Omega \left(\hat{\beta}_{BASIC} - \hat{\beta}_{WEIGHT}\right) \tag{5}
\]

and it results that the weighted sampling gives parameters’ estimates “closer” to those obtained with full choice set (see Table 3).

<table>
<thead>
<tr>
<th>No. of Alternatives</th>
<th>Basic Model - Weighted Sampling Dom1</th>
<th>Basic Model - Weighted Sampling Dom4</th>
<th>Basic Model - Simple Random Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.0217</td>
<td>0.0116</td>
<td>0.5918</td>
</tr>
<tr>
<td>75</td>
<td>0.0281</td>
<td>0.0170</td>
<td>0.5882</td>
</tr>
<tr>
<td>50</td>
<td>0.2248</td>
<td>0.0943</td>
<td>0.7981</td>
</tr>
<tr>
<td>25</td>
<td>0.1753</td>
<td>0.0816</td>
<td>0.2396</td>
</tr>
</tbody>
</table>

Table 3 – Comparison between vectors of parameters in the different specifications
CONCLUSIONS AND FURTHER PERSPECTIVES

In this paper different dominance variables have been defined and tested within random utility spatial choice models. Estimation results provide evidence in support of the introduction of these attributes in the utility function of a MNL model to provide a better forecasting model. Specifically, the best performances are obtained by specifying a model which combines weak spatial and strong global dominance variables together.

The use of dominance variables has been tested also as a sampling technique. In particular, dominance variables have been used as selection weights in a random sampling approach. Results obtained show that the weighted sampling gives parameters’ estimates “closer” to those obtained with full choice set.

Random utility models once again have proved to be flexible tools as utility functions can be specified considering all attributes found significant like the intervening opportunities factors. This work, together with Cascetta and Papola (2005), is another example of the gains possible with the dominance approach in a completely different environment. Thanks to this greater flexibility it is useful to consider it for further research in more sophisticated choice set approaches (e.g. Manski), as well as for other spatial choice contexts, such activity location choice or destination choice, with original and new specifications able to improve the reproducing ability in this complex choice context.

REFERENCES


AUTHORS INFORMATION

Ennio CASCETTA  
cascetta@unina.it  
Department of Transportation Engineering, University of Naples Federico II - Italy

Francesca PAGLIARA  
fpagliar@unina.it  
Department of Transportation Engineering, University of Naples Federico II - Italy

Kay W. AXHAUSEN  
axhausen@ivt.baug.ethz.ch  
IVT. ETH – Zurich