Dominance Attributes for Alternatives’ Perception in Choice Set Formation: an Application to Spatial Choices

Ennio Cascetta  
Department of Transportation Engineering  
University of Naples Federico II  
Via Claudio, 21 – 80125 Naples – ITALY  
Phone: +39-081-7683351  
Fax: +39-081-7683946  
E-mail address: cascetta@unina.it

Francesca Pagliara*  
Department of Transportation Engineering  
University of Naples Federico II  
Via Claudio, 21 – 80125 Naples – ITALY  
Phone: +39-081-7683949  
Fax: +39-081-2390366  
E-mail address: fpagliar@unina.it

Kay W. Axhausen  
IVT – ETH  
CH – 8093 Zürich  
Phone: +41-44-6333943  
Fax: +41-44-6331057  
E-mail address: axhausen@ivt.baug.ethz.ch

SUBMISSION DATE: 15/07/2006

Words: 6012  
Plus 4 Tables: 1000  
Plus 1 Figure: 250  
TOTAL: 7262

• Corresponding author

Paper for the 86th Annual Meeting of the  
Transportation Research Board  
Washington D.C.  
January 21st-27th 2007

Committee on Transportation Demand Forecasting (ADB40)
ABSTRACT

Few practitioners interested in spatial choice would contest the importance of properly specifying the choice set in destination choice modelling applications to avoid model parameter bias. The number of 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 of alternatives is considered instead. The traditional approaches consist of delineation based on a restricted list of deterministic criteria selected by the analyst. This opens the door to likely mis-specification of choice sets and, in turn, it can lead to an incorrect setting of the parameters of the utility function and incorrectly predicted choice probabilities. The objective of this paper is to provide a methodology which aims to bypass the question of excluding a priori any alternative from the choice set. The approach considered is that of assigning to each alternative a perception degree, which varies among alternatives, but which make them perceived simultaneously by the decision-maker. This methodology will be applied to model residential location choice in the canton of Zurich.
INTRODUCTION

The importance of properly specifying the choice set to avoid biasing the model parameters is well recognized in the theoretical literature (12) and it is particularly relevant for spatial choice related models where alternatives are generally very 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 (5, 26). This set merely provides a starting point from which sets of greater interest may be constructed by the decision-maker, either accidentally or purposefully (29).

A number of approaches to define choice sets or develop choice set generation procedures have been pursued, most in the marketing and travel mode choice literature. Consequently, the majority of empirical tests of these approaches are in the non-spatial domain of brand choice or travel mode, even though a few examples have recently been proposed (10). It has been argued that defining a choice set is a more arduous task in the context of spatial choice modelling versus the relatively straightforward setting of the choice set in non-spatial choice contexts (13).

As it name implies, the awareness or knowledge set consists of the subset of items in the universal set of which, for whatever reason, the decision-maker is “aware” (whether they come to mind on a given occasion or not) and which are believed appropriate for the decision-maker’s goals or objectives. Knowledge of the items in this set is presumed to reside in individual long-term memory; any item could potentially be selected for processing (24). It is from the awareness set that the consideration set evolves. A consideration set is purposefully constructed and can be viewed as consisting of those goal-satisfying alternatives salient or accessible on a particular occasion. While an individual may have knowledge of a large number of alternatives, it is likely that only a few of these will come to mind for a relevant use or purpose.

The decision-maker needs not, and typically does not, to possess the same level of knowledge about each alternative in any set. More information may be acquired once it is realized that a decision has to be made, but often a decision will reflect only the available information about the alternatives.

Because of its dynamic nature, it is sometimes useful to define another, closely related set in more static terms. In this interpretation, the choice set is defined as the final consideration set, i.e. The set of alternatives considered immediately prior to choice.

Most spatial choices are made from large universal choice sets. For instance, the number of stores within a reasonably sized urban perimeter that an individual can possibly go to for grocery shopping is conceivable beyond 100. Possible stores can number in the thousands when shopping for soft drinks, books, clothes or shoes. The number of places in a medium-sized city where a household might choose to live can also number in the thousands. The number of 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 of alternatives is considered instead (34).

Because choice sets are not directly observable, this approach is usually simplified in the modelling of revealed-preference spatial data by ad hoc, deterministic, and sometimes arbitrary, assumptions on the size, composition and interpersonal variability of the choice set of alternatives (35). The traditional approaches work with a restricted list of deterministic criteria selected by the analyst. This opens the door to the likely misspecification of choice sets. Actually, two misspecification scenarios are likely to arise. First, the choice set assigned by the analyst can be a subset of the individual’s true choice set. In this case, the model parameters can still be estimated consistently and choice probabilities correctly predicted by a random utility based spatial choice model under well-defined conditions. In contrast, erroneous model parameter estimates are expected from the case where the choice set defined by the analyst includes alternatives actually never evaluated by the decision-maker. Here, the choice model assigns non-negative probabilities to all alternatives in the choice set, including those that are not in the true choice set. This results in inconsistent estimates of the choice function, and faulty interpretations of individual behaviour (40).

The objective of this paper is to provide a methodology which aims to bypass the question of excluding *a priori* any alternative from the choice set. The approach considered is that of assigning to each alternative a perception degree, which varies among alternatives, but which make them perceived simultaneously by the decision-maker. Following Cascetta (8,9,10) dominance attributes will be developed as perception attributes and they will be introduced in the utility function together with other structural attributes.

This paper is organized in five sections. In the second section a review of choice set generation approaches, with a focus on their use in random utility modelling, is presented. The third section deals with the research methodology, highlighting the role of dominance among alternatives. In the fourth this methodology will be applied to the residential location choices of the residents of the canton of Zurich in Switzerland. Results are reported and discussed as well. Conclusions are reported in the last section.
REVIEW OF PREVIOUS APPROACHES TO CHOICE SET FORMATION IN SPATIAL CHOICES

As a starting point, it is commonly assumed that an individual’s true choice set on a given choice occasion contains those alternatives considered by the individual (32,33). However, given imperfect individual level information – including the subset of alternatives that individuals are aware of, their familiarity with these alternatives, and constraints limiting the viability of some of the alternative as choices – the true choice set is not perfectly observable.

Therefore, three second best approaches for choice set formation have emerged in the literature. The first, deterministic, approach entails excluding alternatives from the choice sets used for estimation if they fail to satisfy a specific rule defined by the analyst. Examples in the destination choice literature include the early work by Black (7), whereby sites located beyond a maximum travel time or distance threshold are discarded; the contribution by Hicks and Strand (17); the maximum travel time and distance criteria examined by Parsons and Hauber (25) and Whitehead and Haab (39). The second, behavioural, approach follows from Manski’s (21) two-stage model of choice set formation, whereby the probability an individual chooses a given alternative is conditional upon the alternative belonging to the choice set. The probability that the alternative belongs to the choice set is conditioned upon individual characteristics and then modelled jointly with the probability of the alternative choice. The third one is the random selection approach (4) where all the alternatives have equal probability of being chosen.

Whether to adopt simple deterministic criteria, random selection procedures or a more rigorously behavioural approach, the development of a formal model of choice set formation will depend primarily upon the size of the universal choice set, the availability of individual information specific on choice set formation, and the analysts’ technical expertise. The deterministic approach is ad hoc as the analyst must decide upon both the exclusion criterion and the rigour with which it will be applied. In contrast, the behavioural approach is consistent with utility maximization and by design provides insight into the decision making process that is unobtainable with the deterministic approach. However, behavioural models of choice set formation become computationally intractable if the universal choice set contains more than a few alternatives. As destination choice analysts frequently consider universal choice sets containing large numbers of alternative locations (e.g. Dozens or hundreds of alternatives), deterministic criteria have prevailed for generating choice sets in empirical applications. For instance, in Miller and O’Kelly (23), the choice set of a certain individual is composed of all the destinations actually chosen by individuals from the same geographic area. In other cases, all individuals share an identical choice set made of all destinations in the urban study area. Gautschi (14) restricts the set of feasible locations for major non-grocery shopping in suburban San Francisco to four major retail centres. The criterion is that these centres generate the bulk of the non-grocery retail business in the study area. Smaller centres are discarded from potential destinations for computational reasons. As a result, estimation results can be expected to be significantly skewed, in particular for those individuals who chose the discarded alternatives.

The recent work of Scrogin et al. (27) developed a multi-variate efficient frontiers approach to choice set formation. Choice set are formed with respect to the deviation of sites from the efficient frontier. Alternatively, the behavioural approach has its roots in the work of Manski (21). Manski’s two stage modelling procedure has been dominating in the existing literature, where the first stage treats generation of a choice set and the second stage consists of choosing an alternative from the choice set. However, it has been pointed out that from behavioural perspective, choice needs not be modelled as the above-mentioned two-stage process because separate specification of choice set provides no information useful for predicting choices beyond that contained in decision makers’ utility functions. Furthermore, for the two-stage modelling procedure, when the number of possible alternatives becomes large, the number of possible choice sets increases drastically. To curtail the problem, Swait and Ben-Akiva (33) advocated the use of a priori restrictions to be placed on the membership of the set of possible choice sets. Constraints work as exclusion criteria.

A more recent contribution comes with the paper by Swait (32), who introduced a new member of the generalized extreme value (GEV) family of discrete choice models that directly incorporates choice set generation modelling into the specification via the GEV generating function. An interesting feature of the model is that the choice set probabilities need to make no use of exogenous information, but are instead taste-driven. This is more in line with the proposal of Horowitz and Louviere (18), who questioned the two-stage model of Manski and developed models of choice set generation on the assumption that choice sets are not separate constructs per se, but are simply another expression of preferences, just as are choices.

Another approach considered is the time-geographic approach, where an individual’s activity profile in space and through time (the path) is confined to a feasible “prism” shaped by three forces – or constraints – defined as “capacity”, “authority” and “coupling” Hagerstrand’s (15). Given its focus on constraining environments, the time-geographic approach has naturally grown into a suitable framework for the modelling of choice set.
The rank can be considered as a dominance attribute for the utility function specification, creating in such a way
occupied by the alternatives with few dominations and, therefore, by those better perceived by the individual.

The rank itself. The lowest rank (1) will be occupied by the alternative with the greatest number of dominations and, in
which the rank occupied by the alternatives is defined by the number of alternatives dominating the alternative

times an alternative is dominated by the others. A dominance ranking of the alternatives can be provided as well,
can be a Boolean variable 0/1, it can be a variable assuming values between 0 and 1; it can be the number of
utility model:

In this paper, the choice of an alternative by an individual will be simulated in two steps: by introducing into

The innovative aspect consists of specifying the perception variables through the concept of dominance,

The concept of dominance among alternatives within random utility

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 (4, 8).

In this paper, the choice of an alternative by an individual will be simulated in two steps: by introducing into the
utility function specification some variables reproducing the perception of the alternative and by estimating for this,
within the model itself, a parameter (9).

The innovative aspect consists of specifying the perception variables through the concept of dominance,

The innovative aspect consists of specifying the perception variables through the concept of dominance,

In many choice contexts, like the residential location one, one can observe that some alternatives are not
taken into account since they are dominated by other alternatives. In particular a general methodology has been
developed to define:

a) when two alternatives are comparable;
b) when an alternative dominates (or is dominated by) another;
c) a method for using such information.

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 \( c_{od'} \). Moreover, a stronger domination (spatial domination) can be constructed which recalls the concept of
intervening opportunities (31). 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 \( od' \)). In this case, \( d \) represents an intervening (and better)

opportunity along the path, or bundle of paths, towards \( d' \).

In particular, the perception of a zone can be approximated through such dominance attributes to be inserted
in the model utility function as availability/perception attributes. Considering a multinomial logit (MNL) random utility model:

\[
p(d) = \frac{\exp(V_d)}{\sum_d \exp(V_d)} \quad (1)
\]

with

\[
V_d = \sum_n \beta_n X_{dn} + \sum_k \gamma_k Y_{dk} \quad (2)
\]

where \( \beta_n \) and \( \gamma_k \) are coefficients of the utility and availability/perception attributes respectively.

A dominance value will be assigned to each alternative. The new variable can be defined in several ways: it
can be a Boolean variable 0/1, it can be a variable assuming values between 0 and 1; it can be the number of
times an alternative is dominated by the others. A dominance ranking of the alternatives can be provided as well,
in which the rank occupied by the alternatives is defined by the number of alternatives dominating the alternative
itself. The lowest rank (1) will be occupied by the alternative with the greatest number of dominations and, in
turn, the model will give it a lower probability of belonging to the set of alternatives. The highest ranks will be
occupied by the alternatives with few dominations and, therefore, by those better perceived by the individual.
The rank can be considered as a dominance attribute for the utility function specification, creating in such a way
a perception degree of a given alternative. The proposed methodology aims to bypass the question of excluding a priori any alternative from the choice set and so avoiding model misspecification.

**METHODOLOGY APPLICATION TO RESIDENTIAL LOCATION CHOICES IN ZURICH**

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 have an influence on 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 (22) 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 Bhat and Guo (1); Cooper et al. (11); Kim et al. (19); Sermons and Koppelman (28); Simmonds and Skinner (30); Wardman et al. (38) and others can be referred to for a review of the main factors influencing residential location choice behaviour. All of them have been useful in defining the explanatory variables employed in this paper.

**Model estimation**

Model estimation requires that the model is specified (i.e. The functional form and the variables are defined), calibrated (i.e. The unknown coefficients are estimated) and validated (i.e. The ability to reproduce the available data is tested) (8).

In 2005 an RP survey was conducted in the canton of Zurich in Switzerland covering the mobility and moving biography of the respondents (see 2 and 3 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 we know 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 (see Figure 1).

The zonal data was obtained from an IVT database described in Tschopp et al. (36). The residential location model specified is a Multinomial Logit model and the proposed specification is:

\[
\sum \beta \cdot X_{ad} = \beta_{Price, d} \cdot \text{Price}_{d} + \beta_{\ln\text{Stock}, d} \cdot \ln\text{Stock}_{d} + \beta_{\text{Logsum}_{LM}, d} \cdot \text{Logsum}_{LM, d} + \beta_{\text{Logsum}_{H}, d} \cdot \text{Logsum}_{H, d} + \beta_{\ln\text{Workplaces_serv}, d} \cdot \ln\text{Workplaces_serv}_{d} + \gamma_{\text{Dom1}, d} \cdot \text{Dom1}_{d} + \gamma_{\text{Dom2}, d} \cdot \text{Dom2}_{d}
\]

where:

- \(\text{Price}_{d}\) is the average land price of zone \(d\);
- \(\ln\text{Stock}_{d}\) is the natural logarithm of the housing stock in zone \(d\);
- \(\text{Logsum}_{LM, d}\) is the logsum of the mode choice model for work purpose for low-medium income residents;
- \(\text{Logsum}_{H, d}\) is the logsum of the mode choice model for work purpose for high income residents (37);
- \(\ln\text{Workplaces_serv}_{d}\) is the natural logarithm of the workplaces in services (summation of retail, leisure and services to the households such as education, health) in zone \(d\);
- \(\text{Dom1}_{d}\) is the number of times the zone \(d\) is dominated, i.e. the number of zones \(d’\) for which the following occurs simultaneously:
(a) $d’$ has average land price lower than $d$;
(b) the distance from the respondent’s workplace zone $d’$ ($dist_{d’}$) is shorter than that to $d$ ($dist_d$);
(c) $d’$ is along the path to reach the respondent’s workplace zone $d$ from $o$: $dist_{od} + dist_{d’d} < 1.2 \cdot dist_{od}$

$Dom_{2d}$ is the number of times the zone $d$ is dominated using a less strict definition, i.e. it indicates the number of zones for which conditions (b) and (c) simultaneously occurs regardless condition (a).

**FIGURE 1 ZOOM ON THE ZONING SYSTEM OF THE CANTON OF ZURICH**

Table 1 reports the descriptive statistics of the variables.

**TABLE 1 Descriptive statistics of the variables**

<table>
<thead>
<tr>
<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>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>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>-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>-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>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>$Dom_{1d}$</td>
<td>3.52</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>6.47</td>
<td>INTEGER</td>
</tr>
<tr>
<td>$Dom_{2d}$</td>
<td>23.28</td>
<td>17</td>
<td>0</td>
<td>151</td>
<td>21.68</td>
<td>INTEGER</td>
</tr>
</tbody>
</table>
In the following the logic of the rules introduced with the concept of dominance is explained in detail. Consider an example with only four alternatives and one respondent. Suppose that the respondent lives in zone 1 and works in zone 4. Average land prices per square meters are 660 (CHF), 300 (CHF), 670 (CHF) and 680 (CHF) for zones 1, 2, 3 and 4 respectively. The distances (km) among the zones are: \( d_{12} = 1.5 \) km (and vice versa); \( d_{13} = 2.5 \) km (and vice versa) and \( d_{14} = 3 \) km (and vice versa).

For the respondent who lives in zone 1 and works in zone 4, the first dominance variable is equal to 1, i.e. there is just one alternative (zone) dominating the origin zone 1. The zone dominating zone 1 is zone 2 because it has lower average land price (300 CHF) with respect to zone 1 (660 CHF); it is closer to the respondent’s workplace \( d_{24} < d_{14} \) (1 < 3) and it is along the path to reach the respondent’s workplace, i.e. \( D_{12} + d_{24} < 1.2 \cdot d_{14} \) (2.5 < 4.2).

Applying the same procedure to alternatives 2, 3, and 4, it has come out that there is no alternative dominating them. Therefore, the ranking of the alternatives is: 1 (1); 2, 3, 4 (0) at the same level. In brackets the number of dominations is reported.

For the same respondent, the second dominance variable is equal to 3, because there are three zones dominating zone 1 regardless the first condition on price. The three zones are 2, 3 and 4. In this case also zone 3 verifies the condition of being closer to the respondent’s workplace \( d_{34} < d_{14} \) (2.5 < 3) and it is along the path to reach the respondent’s workplace, i.e. \( D_{13} + d_{34} < 1.2 \cdot d_{14} \) (3.5 < 4.2), while zone 4 coincides with the respondent’s workplace. Applying the same procedure for the other alternatives, it follows that alternative 2 has two zones dominating it, i.e. 3 and 4; alternative 3 has one zone dominating it, i.e. Zone 4 and alternative 4 has no alternative dominating it. Therefore the ranking of the alternatives is: 1 (3); 2 (2); 3 (1) and 4 (0).

The calibration of the MNL model has been carried out with the software BIOGEME version 1.4 (6). Calibration results are reported in Table 2.

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Basic model (1)</th>
<th>Basic model (2)</th>
<th>Basic model (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{PR} ) (t-statistic)</td>
<td>-0.00191 (-14.357)</td>
<td>-0.00064 (-4.234)</td>
<td>-0.00216 (-13.515)</td>
</tr>
<tr>
<td>( \beta_{ST} ) (t-statistic)</td>
<td>0.45028 (5.083)</td>
<td>0.17205 (1.992)</td>
<td>0.46287 (4.929)</td>
</tr>
<tr>
<td>( \beta_{logsum_{LM}} ) (t-statistic)</td>
<td>3.09962 (25.768)</td>
<td>3.11859 (25.013)</td>
<td>3.56172 (25.503)</td>
</tr>
<tr>
<td>( \beta_{logsum_{HR}} ) (t-statistic)</td>
<td>2.67418 (20.294)</td>
<td>2.72389 (20.110)</td>
<td>3.09427 (20.842)</td>
</tr>
<tr>
<td>( \gamma_{dom1} ) (t-statistic)</td>
<td>0.72390 (9.617)</td>
<td>0.38477 (5.737)</td>
<td>0.84832 (10.383)</td>
</tr>
<tr>
<td>( \gamma_{dom2} ) (t-statistic)</td>
<td>-0.25340 (-8.3922)</td>
<td>-0.26805 (-7.259)</td>
<td>-0.01716 (-21.102)</td>
</tr>
<tr>
<td>No. observations</td>
<td>658</td>
<td>658</td>
<td>658</td>
</tr>
<tr>
<td>Mean no. of alternatives per respondents</td>
<td>182</td>
<td>182</td>
<td>182</td>
</tr>
<tr>
<td>( \ln (\theta) )</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
</tr>
<tr>
<td>( \ln (\beta) )</td>
<td>-1195.27</td>
<td>-1086.98</td>
<td>-1007.82</td>
</tr>
<tr>
<td>( \rho^2 )</td>
<td>0.650</td>
<td>0.682</td>
<td>0.705</td>
</tr>
</tbody>
</table>

In particular, three specifications of increasing complexity are reported. As it can be seen, all coefficients’ signs are consistent with expectations: utility attributes (\( \beta_{PR}, \beta_{ST}, \beta_{logsum_{LM}}, \beta_{logsum_{HR}}, \beta_{WP\_SERV} \)) have the expected sign (\( \beta_{PR} \) is negative as it is a disutility, all the others are positive) while negative perception attributes (\( \beta_{dom1}, \beta_{dom2} \)) have a negative coefficient. It is interesting to highlight the different behaviour of low-medium income
residents and high income residents and how it is much more important for the former to have different alternatives modes available. Moreover, all coefficients in all specifications are very significant and there is a considerable improvement in the goodness of fit statistic when passing from one specification to the next. In particular, a substantial improvement in the goodness of fit of the model is achieved by adding a dominance attribute, that is very significant (see basic model 2); while a bigger improvement can be obtained by adding both dominance variables (see basic model 3), thereby confirming the importance of this kind of approach in simulating residential location choice. The improvement in the goodness of fit of the model is, comparing the $\rho^2$: 4.9% (from model 1 to model 2) and 8.4% (from model 1 to model 3). Moreover, in specification (3) it can be observed that the coefficient relative to the simple dominance degree ($\beta_{\text{dom1}}$) is always much more negative than that relative to the stronger dominance degree ($\beta_{\text{dom2}}$). This confirms the importance of considering zone spatial positions (distance to the workplace and residential zone located along the path to reach the workplace), when defining and identifying the dominated alternatives and in general in the choice set formation of a residential choice context. This approach captures the alternatives’ perception by assigning to each alternative a dominance degree. In this way, none of the alternatives is excluded from the choice set.

Following the deterministic approach

The exclusion of some alternatives and, at the same time observations, through the consideration of the dominance criteria (see Table 3) brings an increase of $\rho^2$ compared to the basic model for the remaining observations. Dominance variables are not included in the new models specification, because when they are, in this approach, they become not significant, as expected. This testifies how the a priori exclusion of some alternatives modifies the model estimates.

TABLE 3 Calibration results with the deterministic approach

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Model D1 Alternatives exclusion on the basis of first criterion</th>
<th>Model D2 Alternatives exclusion on the basis of the second criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{PR}$ (t-statistic)</td>
<td>-0.00226 (-9.957)</td>
<td>-0.00225 (-6.204)</td>
</tr>
<tr>
<td>$\beta_{ST}$ (t-statistic)</td>
<td>0.50985 (3.960)</td>
<td>0.40716 (2.358)</td>
</tr>
<tr>
<td>$\beta_{\text{logsum,LM}}$ (t-statistic)</td>
<td>4.31289 (21.333)</td>
<td>3.66446 (11.810)</td>
</tr>
<tr>
<td>$\beta_{\text{logsum,H}}$ (t-statistic)</td>
<td>3.62897 (16.971)</td>
<td>2.93383 (9.336)</td>
</tr>
<tr>
<td>$\beta_{WP_SERV}$ (t-statistic)</td>
<td>0.89634 (7.805)</td>
<td>0.68957 (4.468)</td>
</tr>
<tr>
<td>$\gamma_{\text{dom1}}$ (t-statistic)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma_{\text{dom2}}$ (t-statistic)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No. observations</td>
<td>454</td>
<td>255</td>
</tr>
<tr>
<td>Mean no. of alternatives per respondents</td>
<td>69</td>
<td>12</td>
</tr>
<tr>
<td>$\ln (0)$</td>
<td>-1898.14</td>
<td>-584.88</td>
</tr>
<tr>
<td>$\ln (\hat{\beta})$</td>
<td>-440.86</td>
<td>-178.32</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.767</td>
<td>0.695</td>
</tr>
</tbody>
</table>

In particular, for each respondent the alternatives excluded are the ones for which there is at least one alternative dominating them. With the exclusion on the basis of the first dominance criterion, the number of observations decrease from 658 to 454. The decrease in the number of respondents from the initial sample is due
to the fact that the alternatives chosen by them have been excluded on the basis of the dominance criteria. Results are interesting. All the parameters’ values increase (with respect to the basic model) and the $\rho^2$ increases as well. On the basis of the second criterion the number of observations decrease from 658 to 255. Again, the $\rho^2$ increases compared to the basic model. The deterministic application of the dominance criteria leaves no room for different subjective perceptions, which are clearly present for those many excluded observations. In future work, one will need to look at less rigorous definitions of dominance for the choice set determination, or use the dominance variables only indicators of relative rank, as was done in the first group of models here.

For all the three models the direct elasticities have been computed based on a 1% increase in Price, Stock, Workplaces and Logsum of the mode choice model. For all of them a distinction has been made between low-medium and high income residents. Mean, medians, minimum and maximum values are reported for the direct elasticity distributions in Table 4 for the basic model (3) with two dominance variables, model D1 and model D2, obtained by excluding the alternatives on the basis of the first and second dominance criterion respectively.

The analysis of the elasticities shows that most of the values for model D1 and D2 are higher than those of the model with two dominance variable (basic model 3). This, again, shows the lack of realism of the deterministic constraint approach and the subsequent parameter misspecification when using a priori deterministic rules for generating the choice set.

For most of the variables the maximum value of the direct elasticities is greater for the high income residents compared to those of the low-medium ones. Very interesting is the result of the elasticity with respect to the logsum of the mode choice model for work purpose. The unexpectedly high values are an indication of the high sensitivity to the transportation system in making location choices.

### TABLE 4 Direct elasticities

<table>
<thead>
<tr>
<th>Variables</th>
<th>Descriptive</th>
<th>Basic model (3)</th>
<th>Model D1</th>
<th>Model D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% increase Price</td>
<td>Mean</td>
<td>-1.33464</td>
<td>-1.33441</td>
<td>-1.39668</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>-1.30810</td>
<td>-1.30819</td>
<td>-1.36769</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.16437</td>
<td>0.17579</td>
<td>0.11277</td>
</tr>
<tr>
<td>1% increase Stock</td>
<td>Mean</td>
<td>0.45908</td>
<td>0.45656</td>
<td>0.50580</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.46103</td>
<td>0.46106</td>
<td>0.50783</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0.39600</td>
<td>0.00260</td>
<td>0.41434</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.46162</td>
<td>0.46163</td>
<td>0.50860</td>
</tr>
<tr>
<td>1% increase Workplaces</td>
<td>Mean</td>
<td>0.84298</td>
<td>0.84298</td>
<td>0.89092</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.84657</td>
<td>0.84665</td>
<td>0.89451</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>0.72690</td>
<td>0.73865</td>
<td>0.72951</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>0.84766</td>
<td>0.84767</td>
<td>0.89587</td>
</tr>
<tr>
<td>1% increase Logsum</td>
<td>Mean</td>
<td>6.82448</td>
<td>7.12366</td>
<td>8.13484</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>5.37173</td>
<td>5.55971</td>
<td>6.16707</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>-0.73609</td>
<td>-0.69358</td>
<td>-0.84981</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>27.44315</td>
<td>28.40461</td>
<td>34.79811</td>
</tr>
</tbody>
</table>

### CONCLUSIONS AND FURTHER PERSPECTIVES

From a formal point of view, modelling the choice of residential location can be compared to modelling the choice of a destination. In random utility models of destination choice, correct prediction of choices is conditional on correct information about the choice set. In this choice dimension, it is important to generate the individual choice set because of the large number of alternatives. In this paper different dominance variables have been defined and used for generating the choice set. Estimation results provide evidence in support of the introduction of these attributes in the utility function, thanks to their significance, and of the improvement in the goodness of fit statistics of the model specified.
Implication of the approach introduced on planning and policy developments is fundamental. The new approach may be a better predictive model than a conventional MNL logit. More research is warranted to corroborate the superiority of the approach for other location choices and other non-systematic destination choices and in other areas. Furthermore, the dominance approach offers a finer understanding of destination choice processes with the introduction of the concepts of availability and perception of destination choice alternatives. The correct modeling of residential location choice represents an important tool for analyzing urban housing policy, transportation planning policy, and urban social spatial structure and is represented in the transportation planning, urban economic, sociology, and urban geography arenas.

Random utility models once again represent more flexible tools as utility functions can be specified considering all the attributes found significant like the intervening opportunities factors. This work, together with the one by Cascetta and Papola (10), 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, with original and new specifications able to improve the reproducing/forecasting ability in this complex choice context.

REFERENCES


Ennio Cascetta, Francesca Pagliara and Kay W. Axhausen

13

report to the Amt für Verkehr, Kanton Zürich, IVT, Ernst Basler + Partner and PTV AG, Zürich and Karlsruhe, 2005.

