Report

The use of dominance variables in choice set generation

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

Publication Date:
2007

Permanent Link:
https://doi.org/10.3929/ethz-a-005378355

Rights / License:
In Copyright - Non-Commercial Use Permitted
The use of dominance variables in choice set generation

Ennio Cascetta
University of Naples Federico II
Department of Transportation Engineering “Luigi Tocchetti”
Via Claudio 21 – 80125 – Naples – Italy
Ph. +39 081 7683351
Fax + 39 081 7683946
E-Mail: cascetta@unina.it

Francesca Pagliara(*)
University of Naples Federico II
Department of Transportation Engineering “Luigi Tocchetti”
Via Claudio 21 – 80125 – Naples – Italy
Ph. +39 081 7683932
Fax + 39 081 7683946
E-Mail: 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

(*) Corresponding author
**Abstract**

This paper is the second experiment for the authors on choice set generation in spatial contexts. Two approaches are proposed: in the first each alternative is assigned a perception degree, using the concept of dominance, so that all of them are perceived simultaneously, if differentially, by the decision-maker. Alternatively, choice sets are formed by random sampling exploring the dominance criteria as weights for the sampling probability. This methodology has been applied to model residential location choice in the canton of Zurich resulting in improved model fits.

**Abbreviated title:** dominance in choice set generation

1. **Introduction**

In the usual approach to random-utility modelling, the choice set is assumed to be known and fixed. Actually, the analyst doesn’t have this information and therefore the choice set should be generated (Horowitz, 1991; Cascetta et al., 2006a).

The paradigm that has come to dominate behavioural research in geography depicts choice behaviour as the process of selecting a certain alternative from a limited set of discrete opportunities in accordance with some decision rules (Timmermans and Golledge, 1990).

The choice set refers to the group of discrete alternatives faced by an individual in the decision-making process which is a subset of the universal set that describes the overall environment of alternatives available to the decision-maker.

Most spatial choices are made from large sets of possible alternatives. For example, the
number of stores within a reasonably sized urban perimeter that an individual can possibly go to for grocery shopping is around a hundred. 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 an individual 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 (Thill, 1992; Thill and Horowitz 1997). A realistic way in which humans deal with a large amount of information is that of considering a portion of the universe of alternatives. In fact, the traditional approaches consist of the elimination of the alternatives based on a restricted list of deterministic criteria selected by the analyst. Theoretically, choice set misspecification may lead to incorrect parameter estimation of the utility function for the analysts, and subsequently, to erroneous interpretations of individual behaviour. Given the widespread popularity of spatial choice modelling in a variety of planning and policy contexts, such as transportation planning, consumer shopping behaviour, inter-regional and intra-urban migration, housing choice, industrial location, and recreation choice, the importance of understanding the sensitivity of model parameters to choice set definition is obvious (Pellegrini et al., 1997).

This paper represents some progress made on a previous paper (Cascetta et al., 2006b) where two different approaches for choice set formation were proposed. In the first each alternative is assigned a perception degree, using the concept of dominance, so that all of them are perceived simultaneously by the decision-maker. Therefore none of the alternatives is excluded from the choice set. Alternatively, choice sets are formed by random sampling using the dominance criteria as weights for the sampling probabilities. In the first case, following Cascetta (2001), the dominance attributes will be developed as perception attributes and they will be introduced in the utility function together with
other structural attributes.

The remainder of the paper is organized as follows. In section 2 possible approaches to choice set formation are reported. Section 3 deals with the methodology proposed. The choice context of concern in this paper is specified in section 4. Concluding comments are provided in section 5.

2. Choice set generation: possible approaches

Most contemporary accounts of human decision-making give a prominent role to simplification (Shocker et al., 1991). Simple models are to be preferred because they are tractable, a fact that is particularly important when the analyst’s task is to make predictions for large numbers of users. On the other hand, many behavioural scientists have questioned the adequacy of such models as explanation since they often find a process that is too complex to be modelled simply.

The first contributions on choice set generation came from the brand literature. The notion of evoked set was the foundation concept in this literature and was defined by Howard and Sheth (1969) as “those brands the buyer considers when he (or she) contemplates purchasing a unit of the product class”. Drawing on this work Narayana and Markin (1975) extended it by developing a model of the evoked set formation process, which posited that the awareness set is composed of evoked, inert and inept sets. The awareness set refers to those brands that are known to the consumer in a given product class. Moreover, the evoked set contains brands that are acceptable to the consumer, the inept set consists of brands that are unacceptable, while the inert set consists of brands for which the consumer holds neutral views.

Using the term consideration set as opposed to evoked set, Shocker et al. (1991) developed a process which views individual decision making as being based upon four
hierarchical or nested sets of alternatives: the universal set, awareness set, consideration set and choice set. Furthermore, they proposed that, except for the first, these sets are processed by the decision-maker prior to choice. The first set, the universal, refers to the totality of all alternatives.

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 (Hicks and Strand, 2000; Parsons and Hauber, 1998; Whitehead and Haab, 1999). The second, behavioural, approach follows from Manski’s (1977) 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 (Ortuzar and Willumsen, 2001) where all the alternatives have equal probability of being chosen.

However each of the three approaches present some limits. In the first case the choice set generation portion of the model deals with a number of choice sets that grows geometrically with the size of the individual’s universal choice set. The scale of the task involved represents a major challenge for the implementation of the model to a large variety of destination-choice situations where numerous alternatives are available.

An obvious danger of the a priori choice set formation, through the elimination of some alternatives with deterministic criteria, is that the specified choice set may be incorrect.
That is, the specified choice set may include an alternative whose choice probability, conditional on the explanatory variables, is zero or it may exclude an alternative whose conditional choice probability is non zero.

The main limit of the random selection approach is that the analyst may run the risk of eliminating many observations as the chosen alternative may not belong to the randomly selected set. The other thing is that the analyst does not know how “big” should be the set, i.e. how many alternatives should belong to the set. This point is strictly connected with the zoning of the system.

For a detailed description of the different approaches to choice set generation the paper by Cascetta et al. (2006b) can be referred to.

3. Research methodology

Random-utility models are widely used to analyse 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).

In this paper, a methodology is provided which aims to bypass the question of excluding a priori alternatives 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. Dominance attributes are developed as perception attributes and they are introduced in the utility function together with other structural attributes. Considering a multinomial logit (MNL) random utility model:
\[
p[d] = \frac{\exp(V_d)}{\sum_{d'} \exp(V_{d'})}
\]  
(1)

with

\[
V_d = \sum_{n} \beta_n X_{dn} + \sum_{k} \gamma_k Y_{d\alpha}
\]  
(2)

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

In particular, the choice of an alternative by an individual is 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.

The novelty of the approach consists of specifying the perception variables through the concept of dominance, previously used in the context of comparison methods for alternative transport projects (Haimes and Chankong, 1985).

In many choice contexts, especially the ones with many alternatives, one can observe that some alternatives are not taken into account since they are dominated by other alternatives. In particular a general methodology is developed to define (Cascetta and Papola, 2005):

a) when two alternatives are comparable;

b) when an alternative dominates (or is dominated by) another;

c) a method for using such information.

Concerning point (a) two alternatives, \(d\) and \(d'\), are comparable if they are characterized by the same attributes (i.e. if their utility functions are specified in the same way).

Concerning point (b) a simple rule is to define \(d\) dominating \(d'\), if all the utility attributes are larger (not smaller) in \(d\) than in \(d'\), and all disutility attributes are smaller
(not larger) in d than in \( d' \), with at least one inequality strictly satisfied. Stronger dominance rules can be generated, for example by introducing some thresholds in these comparisons between attribute values (e.g. \( d \) dominates \( d' \), if the utility (disutility) attributes assume in \( d \) a value greater (less) than that assumed in \( d' \), by some threshold) or some specific factors generated by the comparison between \( d \) and \( d' \) (e.g. the spatial effects).

In random utility destination choice models, the same attributes are generally used for each alternative zone, trip purpose and demand segment (individual category).

As availability/perception attributes \( Y_d \), different dominance attributes have been proposed in previous works (Cascetta and Papola, 2005). In particular, two general dominance degrees can be generated using distance as an impedance attribute. With the first, it is assumed that alternative \( d \) dominates 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 cost \( c_{od} \) is smaller than \( c_{od'} \). With the second, a stronger domination (spatial domination) is constructed referring to the concept of intervening opportunities (Stouffer, 1960). 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 odd' 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 \( d' \). Figure 1 reports an example of spatial domination.

**FIGURE 1**

A dominance degree of each alternative \( d \), that is the number of alternatives dominating it, can be identified. Dominance attributes can be generated for each alternative (e.g. the
dominance degree itself) and used as perception attributes $Y$ in a choice set formation approach.

The proposed methodology aims to bypass the question of excluding a priori any alternative from the choice set and so avoiding model misspecification.

4. **Methodology application**

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.

The choice of residence generally involves trade-offs among several factors which give the household the highest possible utility. They can be classified into housing attributes like land price housing stock; transportation attributes like generalized cost to the workplace and neighbourhood attributes such as quality of the immediate environment.

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 Bhat and Guo (2004); Cooper et al. (2001); Kim et al. (2005); Sermons and Koppelman (1998); Wardman et al. (1998) 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.

4.1 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. 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 (see Figure 2).

![FIGURE 2](image)

The residential location model specified is a Multinomial Logit model and the variables considered are:

- $Price_d$ is the average land price of zone $d$;
- $lnStock_d$ is the natural logarithm of the housing stock in zone $d$;
- $Logsum_{ad}^{LM}$ 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)
*Logsum\(_{ad}^H\)* is the logsum of the mode choice model for work purpose for high income residents;

*lnWorkplaces\(_{servd}\)* 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;

*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\(_{od'}\)**) is shorter than that to \(d\) (**dist\(_{od}\)**);

(c) \(d'\) is along the path to reach the respondent’s workplace zone \(d\) from \(o\): **dist\(_{od'}\)** + **dist\(_{d'd}\)** < 1.2 **dist\(_{od}\)**

*Dom2\(_d\)* 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).

*DOM3\(_d\)* a simple dominance degree of each alternative \(d\), i.e. the number of zones \(d'\) that are closer, with respect to \(d\), to the respondent’s workplace zone.

*DOM4\(_d\)* a simple dominance degree; it indicates the number of zones for which conditions (a) and (b) simultaneously occurs regardless condition (c).

Table 1 reports the descriptive statistics of the variables.
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 2

In particular, eight different specifications are reported. As it can be seen, all coefficients’ signs are consistent with expectations: utility attributes ($\beta_{\text{PR}}$, $\beta_{\text{ST}}$, $\beta_{\text{logsum\_LM}}$, $\beta_{\text{logsum\_H}}$, $\beta_{\text{WP\_SERV}}$) have the expected sign ($\beta_{\text{PR}}$ is negative as it is a disutility, all the others are positive) while negative perception attributes ($\beta_{\text{dom}1}$, $\beta_{\text{dom}2}$, $\beta_{\text{dom}3}$, $\beta_{\text{dom}4}$) 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 and when passing from the basic model (specification 1) to model specification 5. The latter shows an improvement in the goodness of fit equal to 20% (from model 1 to 5) and equal to 14.37% (from model 2 to 5). A further improvement can be obtained by adding two dominance variables (see specification 8), 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 20.61% (from model 1 to model 8) and 0.51% (from model 5 to model 8).

In particular, the dominance variable that works better is DOM3 (specification 5), i.e.
for the residents of the canton of Zurich it is very important to consider in their location choice zones which are closer to their workplaces. The best model specification (in terms of $\rho^2$) is the one which combines $DOM1$ and $DOM3$ (specification 8). The significance of $DOM3$ (-4.788) becomes half the value presented in specification 5 (-8.377), because $DOM1$, which comprises $DOM3$ as well, captures some effects that are not considered in specification 5.

It is interesting to see how the values of the different parameters change from one specification to the next. For example the parameters of the variables $Logsum_{od}^{LM}$ and $Logsum_{od}^{H}$ tend to increase from specification 1 to 5 (the parameter value of $Logsum_{od}^{LM}$ being always greater than the corresponding value of $Logsum_{od}^{H}$), with a slight decrease when two dominance variables are combined together and assuming again almost the same values of specification 5 in the last specification (specification 8).

The results obtained by Cascetta and Papola (2005) in the context of destination choice modelling for non-systematic trips can be compared with the ones reported here. The attributes (utility attributes) of the basic model that they considered were workplaces in commerce and services; logsum of the mode choice model for work purpose and the two dominance attributes defined as follows:

$DOM1 = \text{spatial dominance degree of zone } d^*: \text{it indicates the number of zones } d \text{ for which the following occurs simultaneously:}$

(a) $d$ contains a number of workplaces larger than that in $d^*$;

(b) the distance from the user origin to $d$ ($dist_{od}$) is shorter than that to $d^*$ ($dist_{od^*}$);

(c) $d$ is along the path $o-d^*$: $dist_{od} + dist_{od^*} < 1.2 \cdot dist_{od^*}$.

$DOM2 = \text{simple dominance degree; it indicates the number of zones for which conditions (a) and (b) simultaneously occurs regardless condition (c).}$

When passing from the basic model to the one with one dominance attribute the $\rho^2$
increases of 11.16%, while passing from the basic model to the one with two dominance variables the increase of the $\rho^2$ is 11.65%. The improvement of the goodness of fit from the model with one dominance variable and the one with two dominance variables is 0.43%. These values compared with the residential location model specified in this paper are encouraging as they show that this methodology works in other choice contexts.

4.2 Alternative approach: randomly sampled choice set incorporating dominance criteria as weights

Many contemporary specifications employ random samples of alternatives. To provide a direct comparison, the second approach uses the dominance variable 1 (assuming that specification 8 represents the best improvement in the goodness-of-fit statistics considering dominance 1 (dominance 3 is partly included in dominance 1) as a sampling inverse weight. The sampling is performed with replacement, but an alternative was only included once in the choice set. Four different choice sets were considered with different number of alternatives (100, 75, 50 and 25 respectively)(see Table 3).

Comparing the models of Table 2 and 3 it can be seen, how the parameters approach the values reported above as the sample size increases. It is interesting to see, that model D1 provides a balance between the basic model and the model with dominance variable 1 and 3. In particular the change of $\rho^2$ is equal to 4.63% passing from Model D1 to Model D4 considering 100 and 25 alternatives respectively. Comparing Model D1, Model D2, Model D3 and Model D4 with the Basic Model (1) the improvement in the goodness of fit is 9.53%, 12.77%, 13.38% and 14.61% respectively.

Calibrations results obtained with this proposed approach have been further compared with a simple random selection approach. Results are reported in Table 4.

All the parameters of Models R1, R2, R3 and R4 have a lower value compared with
Models D1, D2, D3 and D4 respectively and are less significant. Moreover the $\rho^2$ in each model assumes a lower value confirming the validity of the proposed approach.

The weighted random sampling is noticeable. This approach enables capturing some perception variation among the alternatives, and reduces the bias due to omitted alternatives as it happens in the deterministic approach.

**TABLE 3**

**TABLE 4**

5. **Conclusions and further research**

Destination choice modelling and residential location choice modelling are both related to an entire class of behavioural modelling situations which are termed spatial choices, i.e. behavioural choice problems characterized by large sets of geographically distributed alternatives. In this paper different dominance variables have been defined and used for generating the choice set in the context of residential location choice. Estimation results provide evidence in support of the introduction of these attributes in the utility function as well as selection weights in a random sampling choice set generation approach, thanks to their significance and improvement in the goodness of fit statistics of the model specified.

The dominance approach offers a finer understanding of residential location choice processes with the introduction of the concepts of availability and perception of the different choice alternatives. The correct modelling 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 proved to be flexible tools as utility functions can be specified considering all attributes found significant like the intervening opportunities factors here. This work, together with Cascetta and Papola (2005), represents an example of the gains possible with the dominance approach in a completely different environment.
References


Hicks, R. L. and Strand, I. E. (2000) 'The extent of information: its relevance for


Wardman, M., Bristow, A. and Hodgson F. (1998) 'Noise and air quality valuations: evidence from stated preference residential and business choice models'. Presented at 8th World Conference on Transport Research, Antwerp, Belgium.

### Tables

**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$_{ad}^{LM}$</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$_{ad}^{H}$</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$_{serv_d}$</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>
<tr>
<td>Dom$_3d$</td>
<td>87.43</td>
<td>87</td>
<td>0</td>
<td>181</td>
<td>52.86</td>
<td>INTEGER</td>
</tr>
<tr>
<td>Dom$_4d$</td>
<td>20.54</td>
<td>11</td>
<td>0</td>
<td>147</td>
<td>24.86</td>
<td>INTEGER</td>
</tr>
</tbody>
</table>
### Table 2

Model calibration

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Basic Model</th>
<th>(1)</th>
<th>(1) plus Dom1</th>
<th>(1) plus Dom4</th>
<th>(1) plus Dom2</th>
<th>(1) plus Dom3</th>
<th>(1) plus Dom1+ Dom2</th>
<th>(1) plus Dom2+ Dom4</th>
<th>(1) plus Dom1+ Dom3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{PR}$ (t-statistic)</td>
<td>-0.00191 (-14.357)</td>
<td>-0.00064 (-14.411)</td>
<td>-0.00212 (-15.596)</td>
<td>-0.00239 (-15.382)</td>
<td>-0.00263 (-15.315)</td>
<td>-0.00216 (-15.390)</td>
<td>-0.00247 (-9.520)</td>
<td>-0.00166 (-9.540)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{ST}$ (t-statistic)</td>
<td>0.45028 (5.083)</td>
<td>0.17205 (1.992)</td>
<td>0.56036 (5.952)</td>
<td>0.4469 (4.877)</td>
<td>0.53075 (5.482)</td>
<td>0.46287 (4.929)</td>
<td>0.55444 (5.811)</td>
<td>0.32360 (3.454)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{region, LM}$ (t-statistic)</td>
<td>3.09962 (25.768)</td>
<td>3.11859 (25.013)</td>
<td>3.60316 (25.808)</td>
<td>3.77224 (27.579)</td>
<td>4.34535 (29.304)</td>
<td>3.56172 (25.503)</td>
<td>3.90741 (27.281)</td>
<td>4.09870 (29.080)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{region, H}$ (t-statistic)</td>
<td>2.67418 (20.294)</td>
<td>2.72399 (20.110)</td>
<td>3.14261 (21.055)</td>
<td>3.27548 (22.447)</td>
<td>3.72915 (24.280)</td>
<td>3.99427 (22.478)</td>
<td>3.12534 (24.571)</td>
<td>3.63644 (24.571)</td>
<td></td>
</tr>
<tr>
<td>$\beta_{WP, Serv}$ (t-statistic)</td>
<td>0.72390 (5.737)</td>
<td>0.76135 (9.449)</td>
<td>0.72763 (9.3223)</td>
<td>0.73509 (8.810)</td>
<td>0.84832 (10.383)</td>
<td>0.78046 (9.473)</td>
<td>0.51554 (6.815)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{dom1}$ (t-statistic)</td>
<td>-0.35240 (-8.392)</td>
<td>-0.25805 (-7.259)</td>
<td>-0.16635 (-9.933)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{dom2}$ (t-statistic)</td>
<td>-0.03932 (-8.869)</td>
<td>-0.01716 (-21.102)</td>
<td>-0.04505 (-5.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{dom3}$ (t-statistic)</td>
<td>-0.03775 (-8.377)</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01020 (-4.788)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{dom4}$ (t-statistic)</td>
<td>-0.03775 (-8.377)</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01020 (-4.788)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ln(0)$</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td>-3424.24</td>
<td></td>
</tr>
<tr>
<td>$ln(\beta)$</td>
<td>-1195.27</td>
<td>-1086.98</td>
<td>-981.17</td>
<td>-909.94</td>
<td>-752.05</td>
<td>-1007.82</td>
<td>-1007.82</td>
<td>-1007.82</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>0.650</td>
<td>0.682</td>
<td>0.7134</td>
<td>0.734</td>
<td>0.780</td>
<td>0.705</td>
<td>0.742</td>
<td>0.784</td>
<td></td>
</tr>
</tbody>
</table>
Table 3
Calibration results of the proposed approach

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Model D1</th>
<th>Model D2</th>
<th>Model D3</th>
<th>Model D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{PR}$</td>
<td>-0.002038 (-13.452)</td>
<td>-0.001972 (-12.763)</td>
<td>-0.00056 (-3.511)</td>
<td>-0.00095 (-5.621)</td>
</tr>
<tr>
<td>$\beta_{ST}$</td>
<td>0.45957 (4.834)</td>
<td>0.44697 (5.591)</td>
<td>0.16745 (1.790)</td>
<td>0.29243 (2.847)</td>
</tr>
<tr>
<td>$\beta_{logsum_{H}}$</td>
<td>2.96989 (21.604)</td>
<td>3.05883 (22.196)</td>
<td>2.40146 (19.787)</td>
<td>2.07940 (17.720)</td>
</tr>
<tr>
<td>$\beta_{WP_{SERV}}$</td>
<td>0.81357 (9.978)</td>
<td>0.78526 (9.512)</td>
<td>0.38954 (5.392)</td>
<td>0.52445 (6.527)</td>
</tr>
<tr>
<td>No. observations</td>
<td>658</td>
<td>658</td>
<td>658</td>
<td>658</td>
</tr>
<tr>
<td>Mean no. of alternatives per respondents</td>
<td>100</td>
<td>75</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>$ln (0)$</td>
<td>-3055.21</td>
<td>-3000.54</td>
<td>-2625.25</td>
<td>-2136.76</td>
</tr>
<tr>
<td>$ln (\beta)$</td>
<td>-873.79</td>
<td>-800.852</td>
<td>-690.70</td>
<td>-543.99</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.712</td>
<td>0.733</td>
<td>0.737</td>
<td>0.745</td>
</tr>
</tbody>
</table>
### Table 4
Calibration results with a simple random selection

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Model R1</th>
<th>Model R2</th>
<th>Model R3</th>
<th>Model R4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{PR} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (t)-statistic)</td>
<td>-0.001029 (-6.317)</td>
<td>-0.00178 (-9.898)</td>
<td>-0.00016 (-3.410)</td>
<td>-0.00088 (-4.265)</td>
</tr>
<tr>
<td>( \beta_{ST} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (t)-statistic)</td>
<td>0.28600 (2.809)</td>
<td>0.47042 (3.955)</td>
<td>0.13645 (1.790)</td>
<td>0.27143 (2.257)</td>
</tr>
<tr>
<td>( \beta_{logsum,LM} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (t)-statistic)</td>
<td>2.31226 (22.509)</td>
<td>2.39063 (20.703)</td>
<td>2.59360 (19.833)</td>
<td>1.33261 (13.218)</td>
</tr>
<tr>
<td>( \beta_{logsum,H} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (t)-statistic)</td>
<td>1.96660 (17.960)</td>
<td>2.05761 (16.863)</td>
<td>2.26049 (17.141)</td>
<td>1.17058 (10.882)</td>
</tr>
<tr>
<td>( \beta_{WP_SERV} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (t)-statistic)</td>
<td>0.50820 (6.265)</td>
<td>0.77430 (7.859)</td>
<td>0.27654 (5.282)</td>
<td>0.51618 (5.1082)</td>
</tr>
<tr>
<td>No. observations</td>
<td>658</td>
<td>658</td>
<td>658</td>
<td>658</td>
</tr>
<tr>
<td>Mean no. of alternatives per respondents</td>
<td><strong>100</strong></td>
<td><strong>75</strong></td>
<td><strong>50</strong></td>
<td><strong>25</strong></td>
</tr>
<tr>
<td>( \ln (\theta) )</td>
<td>-3055.21</td>
<td>-2500.54</td>
<td>-1300.25</td>
<td>-1256.49</td>
</tr>
<tr>
<td>( \ln (\beta) )</td>
<td>-980.79</td>
<td>-800.85</td>
<td>-390.70</td>
<td>-343.58</td>
</tr>
<tr>
<td>( \rho^2 )</td>
<td>0.679</td>
<td>0.680</td>
<td>0.700</td>
<td>0.726</td>
</tr>
</tbody>
</table>
Captions to illustrations

Figure 1 – Example of spatial domination.
Figure 2 – Zoom on the zoning system of the canton of Zurich.
$d_{OD1} = d_{OD2} = d_{OD3} < d_{OD4}$

$D_2$ spatially dominates $D_4$

area of possible zones
spatially dominating $D_4$

$O = \text{origin}$

$D = \text{destinations}$

$d_{OD} = \text{OD distance}$