Can dominance affect spatial choices?

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

Publication Date:
2010

Permanent Link:
https://doi.org/10.3929/ethz-a-006070894

Rights / License:
In Copyright - Non-Commercial Use Permitted
Can dominance affect spatial choices?

Ennio Cascetta and Francesca Pagliara*

Department of Transportation Engineering – University of Naples Federico II

* corresponding author: e-mail: fpagliar@unina.it

Kay W. Axhausen

IVT, ETH – Zurich

e-mail: axhausen@ivt.baug.ethz.ch
Dear editors

I hope you will consider the paper *Can dominance affect spatial choices?*, suitable for being published in Journal of Regional Science.

Regards

Francesca Pagliara

Naples 21st May 2010
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.

spatial choices, dominance variables, random utility models, sampling techniques
1. 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) stated 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 in order to show how weighted sampling of alternatives provide parameters estimates “closer” to the full choice set.

Previously used in the context of Multicriteria comparison methods for alternative projects (Haimes and Chankong, 1983), the concept of dominance was conceived to
generate a set of non-dominated solutions (projects) and assist the decision-maker in selecting a reasonable compromise between contrasting objectives.

A project $j$ is *dominated* if there exists at least one project $h$ satisfying all the objectives better than, or at least as well as, project $j$:

$$x_{mj} \leq x_{mh} \quad \forall \ m = 1, \ldots, M \quad (1)$$

with at least one of the inequalities (2) holding strictly.

This paper is organized as follows. Section 2 defines the use of dominance variables to model spatial choices within random utility theory. Section 3 provides the application of dominance variables to the residential location choice as an example of spatial choices. In section 4 the use of dominance criteria in sampling techniques is discussed. Conclusions and further perspectives are reported in section 5.

2. Dominance within Random Utility (RU) models

In Random Utility (RU) models it is assumed that the choice set of each individuals’ category is known (Cascetta, 2009). Actually, the analyst doesn’t have this information and therefore the choice set should be generated.

Many approaches have been proposed to generate choice sets. A comprehensive review of these approaches is provided by Thill (1992) in the context of travel destination
choice, while Ben-Akiva and Lerman (1985) and Cascetta (2009) can be referred to for the more general problem. In general, it is possible to define deterministic and stochastic approaches, which can be further classified as exhaustive and selective, implicit and explicit.

The traditional approaches work with a restricted list of deterministic criteria selected by the analyst. This opens the door for a likely misspecification of choice sets. 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 are 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 that are 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 (Williams et al. 1982).

The stochastic, exhaustive and implicit, was proposed by Mansky (1977). He proposed a way of identifying all the possible choice sets \( C \) starting from a given choice set and a way of simulating the choice probability of each of them \( p(C) \). The choice probability of an alternative \( d \), \( p(d) \) is given by:
\[ p(d) = \sum_C p(C) \cdot p(d/C) \quad (2) \]

where \( p(d/C) \) is the probability of choosing alternative \( d \) within choice set \( C \).

The main problem that is associated with Mansky’s approach is that the number of elements to which the sum is extended increases exponentially when the number of alternatives increase. It is not possible to apply this method when several alternatives are available as is destination choice. To avoid the computational problems connected with the enumeration of all the possible choice sets, an implicit approach can be used. In this case the probability that an alternative \( d \) belongs to the decision-maker’s choice set \( C \) is jointly simulated with its probability of being chosen within this set, \( p(d/C) \).

Fotheringham (1983) and Cascetta and Papola (2001), starting from a different theoretical perspective, proposed a similar approach which can be synthesised in the following equation:

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

Various approaches have been proposed for the simulation of \( p(d \in C) \). For a review see Cascetta et al. (2007).
The contribution by Cascetta and Papola (2001; 2005) was that of providing a methodology which aimed 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. Following Cascetta (2009), dominance attributes were developed as perception attributes and they were introduced in the utility function together with other structural attributes.

Consider the following multinomial logit (MNL) random utility model:

\[ p[d] = \frac{\exp(V_d)}{\sum_{d'} \exp(V_{d'})} \]  

(4)

with

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

(5)

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 may 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.
The innovation with the above consists of specifying the perception variables through the concept of dominance and introducing them within RU models.

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 was developed to define:

a) when two alternatives are comparable,

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

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). In the case of 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 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 assumes 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). All the alternatives of this choice context are comparable by definition.
Obviously, simulating the perception of an alternative zone makes sense for those trip purposes (e.g. shopping, sport and leisure) in which the user actually chooses the destination zone, i.e. for those trips that are generally called non-systematic. Concerning point (c), the dominance information can be used both in deterministic and in stochastic choice set formation. First of all, the list of all the dominant alternatives has to be identified. Then a dominance degree of each alternative $d$, that is the number of alternatives dominating it, can be identified. Successively, all dominated alternatives (perhaps with a dominance degree greater than a certain threshold) can be deterministically excluded by the choice set or, alternatively, dominance attributes can be generated for each alternative (e.g. the dominance degree itself) and be used as perception attributes $Y$ in any of the stochastic choice set formation approaches.

As availability/perception attributes $Y_d$, different dominance attributes are generated and used together with other spatial attributes for each zone $d$. In particular, two different dominance degrees were generated using distance as an impedance attribute. In the first case, 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 $c_{od'}$. In the second case, a stronger domination (spatial domination) may be constructed which is consistent with the concept of intervening opportunities. 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 here

The dominance variables 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, or, it can be the number of times an alternative is dominated by the others. A dominance ranking of the alternatives is provided as well, in which the rank that is occupied by the alternatives is defined by the number of alternatives dominating the alternative itself. The lowest rank (1) is 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 are 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 a perception degree of a given alternative.

The proposed methodology aims to bypass the question of excluding \textit{a priori} any alternative from the choice set and so avoiding model misspecification. The model proposed may be calibrated for non-systematic destination trip purpose destination choices and the results obtained in Cascetta and Papola (2005) confirmed both the
importance of choice set formation in the choice context and the improvements that can be achieved with the introduction of the simulation of dominance among alternatives.

3. A residential location choice model involving dominance variables

In this paper the methodology described in the previous section 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); Habib and Kockelman (2008) 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.

3.1 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 (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 (5) and the utility variables considered are

\[ Price_d \] is the average land price of zone \( d \);

\[ \ln Stock_d \] is the natural logarithm of the housing stock in zone \( d \);

\[ \logsum_{od}^{LM} \] 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);

\[ \logsum_{od}^H \] is the logsum of the mode choice model for work purpose for high income residents;

\[ \ln Workplaces_{serv_d} \] is the natural logarithm of the workplaces in services
(summation of retail, leisure and services to the households such as education, health) in zone \( d \) 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 section 2 and they are:

\[ \text{Dom1}_d \]

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 \( d^* (dist_{od^*}) \) is shorter than that to \( d (dist_{od}) \);

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

**STRONG GLOBAL DOMINANCE RULE**

\[ \text{Dom2}_d \]

is the number of zones \( d^* \) for which conditions (a) and (b) simultaneously occurs.

**WEAK GLOBAL DOMINANCE RULE**

\[ \text{Dom3}_d \]

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

\[ Dom_{4d} \]

is the number of zones \( d^* \) weakly dominating, i.e. satisfying only condition (b).

**WEAK SPATIAL DOMINANCE RULE**

The descriptive statistics of the variables are reported in Table 1. The calibration of the MNL model has been carried out with the software BIOGEME version 1.4 (Bierlaire et al, 2006). Calibration results are reported in Table 2.

Table 1 here

Table 2 here

In particular, eight different specifications are reported. As it can be seen, all coefficients’ signs are consistent with expectations: utility attributes (\( \beta_{PR}, \beta_{ST}, \beta_{logsum_{LM}}, \beta_{logsum_{H}}, \beta_{WP\_SERV} \)) have the expected sign (\( \beta_{PR} \) is negative as it is a disutility, all the others are positive) while negative perception attributes (\( \beta_{Dom1}, \beta_{Dom2}, \beta_{Dom3}, \beta_{Dom4} \)) have a negative coefficient.
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\textsubscript{od}\textsuperscript{LM} and Logsum\textsubscript{od}\textsuperscript{H} tend to increase from specification 1 to 5 (the parameter value of Logsum\textsubscript{od}\textsuperscript{LM} being always greater than the corresponding value of Logsum\textsubscript{od}\textsuperscript{H} showing how it is much more important for low-medium income households to have a smaller transport cost),
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).
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 Dom4 (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 with less emphasis on land prices. The best model specification (in terms of $\rho^2$) is the one which combines $Dom1$ and $Dom4$ (specification 8).

In general it is possible to state that models with dominance variables perform better with respect to model without with a decrease of $ln(\beta)$ up to 38%, which provides a better predictive model.

Weak dominance variables seem to perform better than strong dominance variables (e.g. $Dom4$); 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. $Dom1+Dom4$).

4. 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:

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

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:

\[
\left( \bar{\beta}_{\text{BASIC}} - \bar{\beta}_{\text{WEIGHT}} \right) \tilde{V}^{-1} \left( \bar{\beta}_{\text{BASIC}} - \bar{\beta}_{\text{WEIGHT}} \right)
\]  

(7)

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

Table 3 here

5. Conclusions and further perspectives

This article aims to make an original contribution to residential location choice modeling in particular and to any choice context modeling in general by proposing and testing new general methodologies, and new operative attributes that can be used within the RU models.

This methodology has been applied to identify dominance attributes which may be defined in different ways, possibly in accordance with the specific choice context, and which can be introduced as perception attributes in RU models.
This general methodology can be applied in any choice context and, in this article, was applied in particular to the residential location choice context which, by defining specific dominance attributes and testing the corresponding statistical significance. These attributes can be implemented in whatever RU modeling approach and in this article, to test their intrinsic significance, they were introduced in the simplest and least effective RU model, that is, an MNL.

The estimation results show a generally high significance of all these attributes and a considerable improvement in the model’s goodness-of-fit statistics. This confirms the importance of enhancing residential location choice modeling with perception attributes in general and with dominance and spatial variables in particular. Further empirical evidence supporting the validity of the proposed methodology will have to be sought in successive application phases, which will be the first step in future research.

The use of dominance variables has been also tested 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.

It could be interesting to introduce dominance attributes in the different RU modeling approaches and compare them to identify the most effective and convenient way of utilizing this kind of attribute. It is, however, important to underline that the extreme
simplicity and operatività of the way these attributes are used in this article represents in the authors’ opinion the strength of their proposal.

References


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


Figures

Fig. 1- Example of spatial domination

O = origin
D = destinations
d_{OD} = OD distance

d_{OD1} = d_{OD2} = d_{OD3} < d_{OD4}
D_2 spatially dominates D_4
area of possible zones spatially dominating D_4
## 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>$InStock_{d}$</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>$Logsum_{od}^{2m}$</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>$Logsum_{od}^{3l}$</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>$InWorkplaces_{ser}_{d}$</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>20.54</td>
<td>11</td>
<td>0</td>
<td>147</td>
<td>24.86</td>
<td>INTEGER</td>
</tr>
<tr>
<td>$Dom_{3d}$</td>
<td>23.28</td>
<td>17</td>
<td>0</td>
<td>151</td>
<td>21.68</td>
<td>INTEGER</td>
</tr>
<tr>
<td>$Dom_{4d}$</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 2 – Model calibration

<table>
<thead>
<tr>
<th>Logit specifications</th>
<th>Basic Model (1)</th>
<th>(1) plus Dom1 (2)</th>
<th>(1) plus Dom2 (3)</th>
<th>(1) plus Dom4 (4)</th>
<th>(1) plus Dom3 (5)</th>
<th>(1) plus Dom1+Dom4 (6)</th>
<th>(1) plus Dom2+Dom4 (7)</th>
<th>(1) plus Dom1+Dom3 (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{PR} )</td>
<td>-0.00191 (14.357)</td>
<td>-0.00064 (-4.234)</td>
<td>-0.00212 (-14.411)</td>
<td>-0.00239 (-15.596)</td>
<td>-0.00263 (-15.382)</td>
<td>-0.00216 (-13.515)</td>
<td>-0.00247 (-15.390)</td>
<td>-0.00166 (-9.520)</td>
</tr>
<tr>
<td>( \beta_{ST} )</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>
</tr>
<tr>
<td>( \beta_{logsum_{LM}} )</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 (29.086)</td>
</tr>
<tr>
<td>( \beta_{logsum_{H}} )</td>
<td>2.67418 (20.294)</td>
<td>2.72389 (20.110)</td>
<td>3.14261 (21.055)</td>
<td>3.27548 (22.447)</td>
<td>3.72915 (24.280)</td>
<td>3.09427 (22.042)</td>
<td>3.41254 (22.478)</td>
<td>3.65644 (24.571)</td>
</tr>
<tr>
<td>( \beta_{WP_SERV} )</td>
<td>0.72390 (9.617)</td>
<td>0.38477 (5.737)</td>
<td>0.76135 (9.449)</td>
<td>0.72763 (9.322)</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>
</tr>
<tr>
<td>( \gamma_{Dom1} )</td>
<td>-0.25340 (-8.392)</td>
<td>-0.26805 (-7.259)</td>
<td>-0.26805 (-7.259)</td>
<td>-0.26805 (-7.259)</td>
<td>-0.16635 (-4.993)</td>
<td>-0.16635 (-4.993)</td>
<td>-0.16635 (-4.993)</td>
<td>-0.16635 (-4.993)</td>
</tr>
<tr>
<td>( \gamma_{Dom2} )</td>
<td>-0.03775 (-8.874)</td>
<td>-0.03775 (-8.874)</td>
<td>-0.03775 (-8.874)</td>
<td>-0.03775 (-8.874)</td>
<td>0.01935 (-4.111)</td>
<td>0.01935 (-4.111)</td>
<td>0.01935 (-4.111)</td>
<td>0.01935 (-4.111)</td>
</tr>
<tr>
<td>( \gamma_{Dom3} )</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01741 (-8.377)</td>
<td>-0.01020 (-4.788)</td>
<td>-0.01020 (-4.788)</td>
<td>-0.01020 (-4.788)</td>
<td>-0.01020 (-4.788)</td>
</tr>
<tr>
<td>( \gamma_{Dom4} )</td>
<td>-0.06933 (-8.869)</td>
<td>-0.06933 (-8.869)</td>
<td>-0.06933 (-8.869)</td>
<td>-0.06933 (-8.869)</td>
<td>-0.04505 (-5.004)</td>
<td>-0.04505 (-5.004)</td>
<td>-0.04505 (-5.004)</td>
<td>-0.04505 (-5.004)</td>
</tr>
<tr>
<td>( ln(0) )</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
<td>-3424.24 -3424.24</td>
</tr>
<tr>
<td>( ln(\beta) )</td>
<td>-1195.27 -1086.98</td>
<td>-981.17 -909.94</td>
<td>-752.05 -1007.82</td>
<td>-883.44 -739.33</td>
<td>-883.44 -739.33</td>
<td>-883.44 -739.33</td>
<td>-883.44 -739.33</td>
<td>-883.44 -739.33</td>
</tr>
<tr>
<td>( \rho^2 )</td>
<td>0.650 0.682</td>
<td>0.7134 0.734</td>
<td>0.780 0.705</td>
<td>0.742 0.784</td>
<td>0.742 0.784</td>
<td>0.742 0.784</td>
<td>0.742 0.784</td>
<td>0.742 0.784</td>
</tr>
</tbody>
</table>
Table 3 – Comparison between vectors of parameters in the different specifications

<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>