# The ORVal Recreation Demand Model

Authors: Prof Brett Day, University of Exeter
Dr Greg Smith, University of Exeter

Date: 24th January 2017

# Contents


1. Introduction ........................................................................................................ 1

2. Recreation Demand Modelling ................................................................................ 1

3. Data ..................................................................................................................... 2

3.1 MENE Data ......................................................................................................... 2

3.2 ORVal Greenspace Map ........................................................................................ 3

4. Data Processing ...................................................................................................... 6

4.1 Basic Observation Classification ............................................................................. 6

4.2 Destination Matching ............................................................................................. 6

4.3 Respondent Sampling ............................................................................................ 8

4.4 Choice Set Sampling ............................................................................................ 10

4.5 Travel and Time Cost Calculation ......................................................................... 12

5. The Econometric Model ........................................................................................ 13

5.1 Econometric Specification ..................................................................................... 13

4.3 Welfare Estimation .............................................................................................. 17

5.4 Econometric Corrections ....................................................................................... 18

4.7 Covariate Choice ................................................................................................. 19

6. Results .................................................................................................................. 22

6.1 Parameter Estimates ............................................................................................ 22

6.2 Visitation and Welfare Estimation ......................................................................... 30

References ................................................................................................................ 32

# List of Tables

Table 1: Annual sample sizes in the MENE survey ........ 3

Table 2: Recreation sites in the ORVal greenspace map ........ 4

Table 3: Landcovers present at recreation sites ........ 4

Table 4: Designations present at recreation sites ........ 5

Table 5: Points of Interest present at recreation sites ........ 5

Table 6: Choice-based sampling scheme and weights for (WESML) estimation ........ 9

Table 7: Choice set sampling scheme ........ 11

Table 8: Model coefficients for participation decision ........ 22

Table 9: Travel cost coefficient of site choice decision ........ 27

Table 10: Greenspace type coefficient of site choice decision ........ 27

Table 11: Size, diversity and commonality coefficients of site choice decision ........ 28

Table 12: Landcover coefficients of site choice decision ........ 28

Table 13: Similarity coefficients of site choice decision ........ 29

Table 14: Designation coefficients of the site choice decision ........ 29

Table 15: Points of interest coefficients of the site choice decision ........ 30

# List of Figures

Figure 1: Matching scores for each focus trip plotted in ascending order of score .............................. 7

Figure 2: Relative magnitude of participation likelihood across the year (base case December)........ 24

Figure 3: Relative magnitude of participation likelihood across the week (base case Tuesday) ......... 25

Figure 4: Relative magnitude of participation likelihood across Government Office Regions (base case East Midlands) ........................................................................................................ 25

Figure 5: Relative magnitude of participation likelihood across age-gender groups (base case females in their 20s)........................................................................................................... 26

Figure 6: Relative magnitude of participation likelihood across socioeconomic groups (base case segment D) .......................................................................................................... 26

## 1. Introduction

The 'grand challenge' of the ORVal project was to develop a tool that could estimate the recreational value derived from any existing or new greenspace in the whole of England. That challenge is not insignificant. Of course, environmental economists have a long tradition of using data recording people's use of greenspaces in order to develop models that can be used to estimate the economic value derived from those natural resources. In general, however, those efforts have usually focused on one form of greenspace (e.g. beaches, lakes or municipal parks) rather than all forms of greenspace, have considered just a single location or region rather than an entire nation and have not attempted to deliver a tool that allows users to interrogate the underpinning model to answer question of their own design. In this report, we outline the steps that were taken in estimating a model that meets that challenge.

## 2. Recreation Demand Modelling

The approach economists normally adopt to estimate the welfare derived from a good is to observe how demand for that good changes as its price changes. In essence, that relationship traces out how much money individuals are willing to give up in order to enjoy that good; a quantity that (roughly speaking) defines the measure of welfare that economists call economic value. Indeed, throughout this report when we talk about 'value' or 'valuation' we are referring to this particular monetary measure of welfare.

More often than not, however, access to greenspaces does not command a price, or if it does that price is often minimal and without sufficient variation to directly estimate the demand-price relationship. Hence, conventional techniques of welfare estimation are frequently not applicable to the valuation of greenspace.  A solution to this problem was first forwarded by Harold Hotelling in a letter to the National Park Service of the United States in 1947 (Smith and Kaoru 1990). He noted that though the greenspace is not itself a market good, in undertaking a recreational trip individuals incur time and travel costs that in effect can be considered the 'price' of access. In other words, when we observe an individual taking a trip to a greenspace, we can presume that the value they derive from that experience is worth at least the costs incurred in travelling to the site.

When considering just one site, this travel cost method progresses by examining how many trips individuals living at different distances, and hence with different travel costs, choose to make to the recreational greenspace. Information of that nature is sufficient to inform on the value for that particular site. The challenge for the ORVal project was considerably different from the single site case. In particular, we were concerned with recreational activities over all greenspaces in England where those greenspaces were differentiated not only in their location but in the recreational experience they offered.

A related framework that better suits our needs is one that focuses on an individual's choice of which of the array of different greenspaces to visit rather than how many trips to take to a particular greenspace. This discrete choice approach is also a form of travel cost modelling. The intuition of how information on discrete choices provides evidence for welfare valuation progresses as follows. Imagine, an individual has a choice between just two greenspaces. Both greenspaces provide visitors with 2ha of open grassland but the more distant greenspace also possesses 2ha of woodland. If we observe the individual choosing to visit the more distant greenspace we can conclude that the extra welfare derived from being able to visit a greenspace with woodland must be worth at least as much as the extra costs in travelling to that more distant location rather than the closer greenspace. Given sufficient observations on individuals choosing between quality-differentiated greenspaces at different distances from their homes, the discrete choice approach can inform on the economic value that individuals realise from greenspaces with different qualities. Moreover it can be used to predict how likely it is that an individual will choose to visit a particular greenspace from the set of greenspaces available to them.

The econometric method used to estimate discrete choice models are known as Random Utility Models (RUMs). We review the particular RUM approach used in estimating the ORVal model in Section 4. The approach is data intensive. It requires information on the choices individuals make on each recreational choice occasion (in our modelling we assume that each day represents such a choice occasion). In particular, we need to know whether an individual took a trip to greenspace or not and, if they did what the qualities of that site were and the time and travel costs incurred in getting there. Moreover, since this is a choice model, we need details of the qualities and travel costs associated with each other recreational greenspace that individual might have visited instead. In Section 2 we describe the data sources used to construct such a data set then in Section 3 how that data was processed to generate the estimation data set. Finally, Section 5 describes the modelling results and their use in predicting welfare values and visitation rates at existing and new greenspaces.

## 3. Data

### *3.1 MENE Data*

The primary data set supporting estimation of the ORVal model is provided by the Monitor of Engagement with the Natural Environment (MENE) survey. Administered on behalf of Natural England, DEFRA and the Forestry Commission, the MENE survey provides a large, random location sample of recreational day trips taken by adults (over 16 years of age) residents of England. As a consequence, the estimates of visits and values that are estimated from the ORVal model are limited to:

- Recreational day trips
- Residents of England
- Adults

The survey is administered face-to-face, recording the recreational trips to greenspace taken by the respondent over the seven days prior to the interview. Moreover, for one randomly selected trip, the survey elicits detailed information regarding the respondent’s activities on that trip as well as the location and characteristics of the recreational site visited. In this report we describe this trip as the *focus visit*.

The MENE survey runs throughout the year sampling at least 800 respondents each week making the data seasonally representative. As recorded in Table 1, the annual sample amounts to approximately 50,000 and the ORVal model used data from the six years of data collected since the survey began in 2009.

**Table 1: Annual sample sizes in the MENE survey**

| Year | Sample |
|---|---|
| 2009-10 | 48,514 |
| 2010-11 | 46,099 |
| 2011-12 | 47,418 |
| 2012-13 | 46,749 |
| 2013-14 | 46,785 |
| 2014-15 | 45,225 |
| Total: | 280,790 |

The MENE data is provided with a demographic weight for each observation. The weight is calculated so as to ensure that the sample of respondents collected in one month can be adjusted so as to be representative of the adult population of the UK in that year. The demographic characteristics used in calculating the weights are:

- age and sex (for example, males 16- 24, females 85+),
- region of residence,
- social grade,
- presence of children in the household,
- sex and working status (for example, male full time),
- presence of a dog in the household and
- urban/rural residence

Put simply the weight for each observation indicates the number of people in the population represented by that respondent. Accordingly, the weighted sum of observations of, for example, male respondents aged 16-24 will equal the number of males in that age group in England, with the same being true of all the other demographic categories in the list above.

A detailed description of the MENE survey, its administration and the calculation of demographic weights can be found in the MENE Technical Report (Natural England 2015).

### *3.2 ORVal Greenspace Map*

The second key dataset used in the estimation of the ORVal model is provided by the ORVal greenspace map. The ORVal greenspace map is a detailed spatial dataset compiled through the combination and manipulation of a large number of primary data sources that describes the location and characteristics of accessible greenspace across England. Construction of the ORVal greenspace map is provided in the companion report to this document (Day, 2016).

As described in Table 2, the ORVal greenspace map identifies some 129,575 greenspace sites in England that could form the focus of a recreational trip. Those features come in three basic forms;

- <u>parks</u> which consist of areas of accessible greenspace within well-defined boundaries over which visitors usually have freedom to wander at will,
- <u>paths</u> which consist of accessible, walkable routes that pass through the landscape, often traversing a variety of different greenspaces and tending to restrict visitors to defined routes of passage.
- <u>beaches</u>.

**Table 2: Recreation sites in the ORVal greenspace map**

| Type | Number of Sites |
|---|---|
| Parks: | |
| Municipal Park | 19,377 |
| Cemetery | 9,494 |
| Woods | 7,359 |
| Allotment | 6,865 |
| Nature | 2,846 |
| Country Park | 413 |
| Path Access Points | 82,591 |
| Beaches | 630 |
| Total | 129,575 |

Each recreation site is described by various aspects of its physical characteristics; particularly the site's dimensions, landcovers, designations and points of interest.

Table 3 provides an indication of landcovers used to describe sites and how frequently those landcovers were present at the various sites. Note that sites are characterised by a diversity of land covers so the columns of Table 3 do not sum to the number of sites of different types shown inTable 2. Moreover, for paths the presence of a landcover is determined by whether that landcover was found along the path network accessed by a path access point with in 10km of that access point. Further details can be found in the ORVal Greenspace Map report (Day, 2016).

**Table 3: Landcovers present at recreation sites**

| Lancover | Parks | | Paths | | Beaches | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| Woods | 16,478 | 35.5% | 69,838 | 84.6% | 0 | 0.0% |
| Wood Pasture | 1,047 | 2.3% | 11,382 | 13.8% | 0 | 0.0% |
| Agriculture | 5,503 | 11.9% | 76,138 | 92.2% | 0 | 0.0% |
| Natural Grass | 3,139 | 6.8% | 59,565 | 72.1% | 0 | 0.0% |
| Moors | 758 | 1.6% | 15,047 | 18.2% | 0 | 0.0% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Coastal | 334 | 0.7% | 2,941 | 3.6% | 630 | 100.0% |
| Saltmarsh | 198 | 0.4% | 2,023 | 2.4% | 0 | 0.0% |
| Marsh & Fen | 470 | 1.0% | 5,896 | 7.1% | 0 | 0.0% |
| Managed Grass | 16,680 | 36.0% | 77,133 | 93.4% | 0 | 0.0% |
| Sports Pitches | 3,960 | 8.5% | 4,284 | 5.2% | 0 | 0.0% |
| Gardens | 519 | 1.1% | 1,701 | 2.1% | 0 | 0.0% |
| Allotments | 6,820 | 14.7% | 923 | 1.1% | 0 | 0.0% |
| Cemeteries | 9,393 | 20.3% | 2,973 | 3.6% | 0 | 0.0% |
| Sea | 533 | 1.1% | 2,252 | 2.7% | 630 | 100.0% |
| Estuary | 329 | 0.7% | 1,948 | 2.4% | 0 | 0.0% |
| River | 8,223 | 17.7% | 49,493 | 59.9% | 0 | 0.0% |
| Lake | 1,298 | 2.8% | 12,802 | 15.5% | 0 | 0.0% |

Similar data on the presence of different forms of formal designation are provided in Table 4. Note that the category 'nature' includes numerous form of designation for nature protection including local and national nature reserves, Natura 2000 sites, Ramsar Sites, SSSIs and Ancient Woodlands.

**Table 4: Designations present at recreation sites**

| Designation | Parks | | Paths | | Beaches | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| National Park | 1,974 | 4.3% | 9,453 | 11.4% | - | 0.0% |
| AONB | 3,631 | 7.8% | 16,360 | 19.8% | 213 | 33.8% |
| Heritage Coast | 457 | 1.0% | 2,097 | 2.5% | 195 | 31.0% |
| National Trail | 641 | 1.4% | 5,427 | 6.6% | 278 | 44.1% |
| Historic Park | 1,456 | 3.1% | 6,431 | 7.8% | 43 | 6.8% |
| Millennium Green | 444 | 1.0% | 163 | 0.2% | 5 | 0.8% |
| Nature | 6,781 | 14.6% | 36,348 | 44.0% | 455 | 72.2% |
| No Designation | 34,526 | 74.5% | 35,505 | 43.0% | 82 | 13.0% |

Table 5 provides details of the presence of different points of interest at recreational sites.

**Table 5: Points of Interest present at recreation sites**

| Designation | Parks | | Paths | | Beaches | |
|---|---|---|---|---|---|---|
| | Number | Percent | Number | Percent | Number | Percent |
| Archaeological Feature | 589 | 1.3% | 7,930 | 9.6% | - | 0.0% |
| Historic Building | 622 | 1.3% | 3,822 | 4.6% | - | 0.0% |
| Scenic Feature | 263 | 0.6% | 3,904 | 4.7% | - | 0.0% |
| Playground | 4,995 | 10.8% | 1,472 | 1.8% | - | 0.0% |
| Viewpoint | 462 | 1.0% | 17,062 | 20.7% | - | 0.0% |
| No Points of Interest | 39,875 | 86.0% | 65,188 | 78.9% | 630 | 100.0% |

## 4. Data Processing

### *4.1 Basic Observation Classification*

The basic unit of observation in our data is a respondent-day; that is to say, the choice of outdoor recreation activity made by a respondent on a particular day. Since each respondent in the MENE data set provides information on their recreation activity over 7 days, each respondent contributes 7 different observations to the data.

For each of those observations the MENE data reveals whether or not the respondent took an outdoor recreation trip on that day. For the observation constituting the focus trip (the randomly selected trip for which detailed information is selected), MENE also provides information from which we might identify the recreation site visited. Accordingly, at a basic level we can classify observations into one of three groups;

- No trip taken
- Trip taken to unidentified site
- Trip taken to identified site

For reasons not reported in the MENE documentation, the home location of some respondents (reported as a Lower Super Output Area) is not recorded. Since the ORVal model requires information on how far different recreation sites are from a respondent's home, we were forced to drop these 7,706 observations from the dataset. Likewise, for some observations the focus trip was reported as starting out from a location that was not the respondent's home. One possibility for explaining such responses is that the respondent was not at their home for the period covered by the survey perhaps staying with friends or on holiday. Since, the ORVal model focuses exclusively on day trips (as opposed to overnight trips) for the purposes of outdoor recreation, the 4,646 observations for those respondents were also dropped from the data.

Following the removal of observations from the dataset, demographic weights for the remaining sample were recalculated using the 'Anesrake' package for the R statistical software.

### *4.2 Destination Matching*

A first step in bringing together the MENE dataset and the ORVal Greenspace map requires matching the geocoded destinations for focus visits with the recreation sites identified in the greenspace map. Using destination details provided by the respondent (but not recorded in the released data) the survey administrators managed to attribute a six digit BNG reference to some 80% of the focus visits recorded in the survey (Natural England 2015).

The procedure for matching the MENE destination locations with the ORVal Greenspace map focused on the 104,978 respondents that had taken an outdoor recreation trip during the week in which they were interviewed.

As a first step, the 17,862 observations for which the survey administrators had failed to geolocate the focus visit destination were re-classified as having taken a trip to an unidentified location. More complex to deal with, were focus trips associated with recreational activities that were not site-based interactions with the natural environment. Those included the 4,061 observations where the focus trip was described as either "off-road driving or motorcycling", "road cycling" or "appreciating scenery from a car". An additional, 531 observations claimed to have been to a 'village' rather than a greenspace. Since these destinations could not be matched with the ORVal greenspace map, we considered whether they should be dropped from the dataset. Unfortunately, that would have biased the sample insomuch as we could not be sure that non-focus visits by respondents had also been of this nature. Ultimately we decided that the least-worst option was to recategorise these focus visit observations as being trips to an unidentified location.

The remaining 82,524 observations contained information on focus trips where a respondent had travelled from their home to a recreation site and for which the MENE data recorded both home and site locations. Note that our analysis does not address the complicating issue of multi-site trips; the MENE data fails to record the information that would allow a proper characterisation of such trips. Accordingly, each trip is assumed to be solely for the purpose of visiting the site identified by the MENE destination location.

A scoring procedure was developed to facilitate the process of matching MENE destination locations with the ORVal Greenspace Map. In short, for each focus visit all recreation sites within 2.5km of the destination location recorded in MENE were identified. Details of each of those sites were then compared to information provided by the MENE survey and scored according to how well they tallied with details of the actual site visited in terms of their location, environs, site type and landcovers. The weights used to determine scores in the matching procedure were calibrated through examining how well the matching algorithm performed with a training data set where the actual destination could be readily determined from the data provided in MENE. Details of the matching algorithm and the weights used in the procedure can be found in Appendix I.

The matching algorithm took approximately 10 hours to run and identified a best guess as to the site on the ORVal Greenspace Map that was considered the mostly likely destination of each focus visit. As shown in Figure 1, where the score for each observation has been plotted in ascending order of score, matching scores varied across the range of 0 to 128.

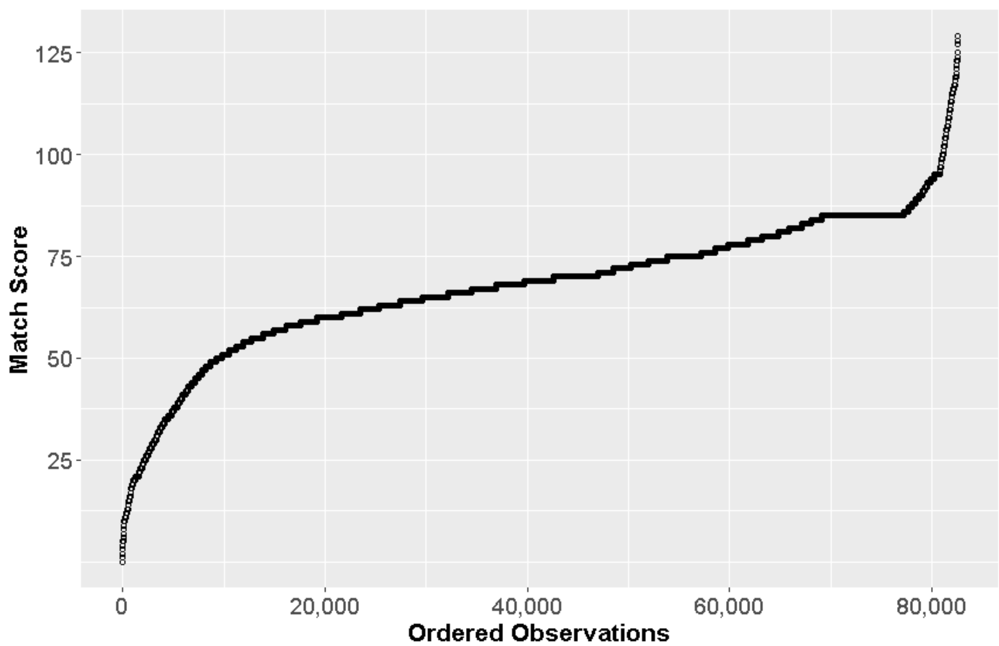


**Figure 1: Matching scores for each focus trip plotted in ascending order of score**

Through inspection of Figure 1, the change in slope of the data around the 50 pts mark was identified as a natural point to split the data. For the roughly 10% of focus trip observations below that threshold, the level of match was deemed too low to believe we had identified the actual site visited. Accordingly, those observations were reclassified into the “trip to unidentified site” category.

Of the 72,141 matched destinations, 1,193 were to sites identified as golf courses. Again these visits complicated the data since most golf courses charge to play, though we cannot be sure that the visits were for the purposes of playing. As with cycling, driving and village trips we decided to take the least-worst option and classify these trips as being to an unidentified site.

### *4.3 Respondent Sampling*

Even with unusable observations removed the remaining dataset contained 268,438 observations. Since the estimation procedure to be used in the analysis (to be described subsequently) was relatively complex, it was decided to further reduce the dataset by drawing a smaller sample from those observations.

To ensure the richness of the data was maintained in that sampling procedure, a process of stratified random sampling was adopted where strata were defined by a respondent’s choice of recreation activity on the focus trip. Accordingly, strata were defined as; (i) observations where no trip was taken; (ii) observations where a trip was taken to an unidentified site and (iii) a further 42 strata defined for observations where trips were taken to sites of different types. Those types were based on a classification of sites defined along four dimensions;

- Type: Beach, Park, Path, Woods, Nature, Allotment, Cemetery or Country Park
- Dominant Land Cover/Use: Woods, Sea Water, Fresh Water, Managed Grass, Agriculture, Natural Grass, Wetlands, Moors & Heath, Allotment or Cemetery
- Dominant Designation: National Park, AONB, Heritage Coast, Nature (including local and national nature reserves, Natura 2000 and Ramsar sites, ancient woodland), National Trail, Forestry Commission, Millennium & Doorstep Green, Historic Park or Country Park
- Points of Interest: Whether or not the site had archaeological remains, a historic building or a scenic feature or a viewpoint.

Applying this this four-dimensional classification scheme to the 129,575 recreation sites in the ORVal Greenspace Map resulted in 493 unique classes of sites. Naturally some of those classes contained very few sites such that classes were further aggregated so as to ensure that the MENE data set contained at least 100 focus visits to sites in each group. Definitions of the 42 groups identified through this procedure are shown in Table 6 and represent the choice-based strata used for sampling. The third column of Table 6 shows the number of observations with a focus trip to sites in each strata.

In order to establish our reduced sample, we used a stratified random sampling method in which we randomly sampled a fixed proportion of observations from each strata. To ensure representation of less commonly taken trips in the sample, the proportion taken for each strata was increasing in the rarity of visits. So a 20% sample was taken from strata with greater than 10,000 observed visits, a 30% sample for strata with between 7,500 and 10,000 visits, a 40% sample from strata with between 4,000 and 7,500 visits, a 50% sample from strata with between 3,000 and 4,000 visits, a 60% sample from strata with between 2,000 and 3,000 visits and a 75% sample for strata with less than 2,000 visits. The sampling probabilities and number of sampled observations are shown in columns 4 and 5 of Table 6.

**Table 6: Choice-based sampling scheme and weights for (WESML) estimation**

| | Description | Num Obs | Sample Probability | Num Sample | WESML Weight |
|---|---|---|---|---|---|
| 1 | All Beaches | 4,389 | 0.4 | 1,756 | 0.578 |
| 2 | All Cemeteries | 3,168 | 0.5 | 1,584 | 0.455 |
| 3 | All Allotments | 516 | 0.75 | 387 | 0.296 |
| 5 | All Country Park | 4,357 | 0.4 | 1,743 | 0.579 |
| 6 | Path, Agriculture | 3,230 | 0.5 | 1,615 | 0.489 |
| 7 | Park, Managed Grass | 17,555 | 0.2 | 3,511 | 1.063 |
| 8 | Path, Managed Grass | 1,260 | 0.75 | 945 | 0.323 |
| 9 | Path, Agriculture, Nature | 1,283 | 0.75 | 962 | 0.325 |
| 10 | Path, Agriculture, AONB | 361 | 0.75 | 271 | 0.340 |
| 11 | Path, Managed Grass, Nature | 649 | 0.75 | 487 | 0.326 |
| 12 | Path, Managed Grass, AONB | 295 | 0.75 | 221 | 0.335 |
| 13 | Woods, Woods | 1,354 | 0.75 | 1,016 | 0.295 |
| 14 | Path, Woods, Nature | 626 | 0.75 | 470 | 0.326 |
| 15 | Path, Managed Grass, Natl Park | 227 | 0.75 | 170 | 0.326 |
| 16 | Woods, Woods, Nature | 1,279 | 0.75 | 959 | 0.321 |
| 17 | Path, Agriculture, AONB, POI | 301 | 0.75 | 226 | 0.326 |
| 18 | Path, Woods, AONB | 289 | 0.75 | 217 | 0.338 |
| 19 | Path, Agriculture, Nature, POI | 370 | 0.75 | 278 | 0.334 |
| 20 | Path, Agriculture, POI | 460 | 0.75 | 345 | 0.323 |
| 21 | Path, Woods | 284 | 0.75 | 213 | 0.314 |
| 22 | Park, Woods | 3,337 | 0.5 | 1,669 | 0.422 |
| 23 | Park, Agriculture | 1,160 | 0.75 | 870 | 0.296 |
| 24 | Path, Fresh Water | 834 | 0.75 | 626 | 0.308 |
| 25 | Path, Moors & Heath, Nature, POI | 245 | 0.75 | 184 | 0.339 |
| 26 | Path, Woods, Nature, POI | 328 | 0.75 | 246 | 0.327 |
| 27 | Nature, Woods, Nature | 1,530 | 0.75 | 1,148 | 0.298 |
| 28 | Path, Managed Grass, AONB, POI | 249 | 0.75 | 187 | 0.318 |
| 29 | Other Fresh Water | 3,443 | 0.5 | 1,722 | 0.453 |
| 30 | Other Sea Water | 1,429 | 0.75 | 1,072 | 0.298 |
| 31 | Other Moors & Heath | 477 | 0.75 | 358 | 0.323 |
| 32 | All Wetlands | 203 | 0.75 | 152 | 0.308 |
| 34 | All National Trail | 557 | 0.75 | 418 | 0.306 |
| 35 | All National Park | 1,204 | 0.75 | 903 | 0.332 |
| 36 | Others No Designation | 2,001 | 0.6 | 1,201 | 0.380 |
| 37 | Other Nature Designation | 2,994 | 0.5 | 1,497 | 0.453 |
| 38 | All Historic Designation | 7,771 | 0.3 | 2,331 | 0.739 |
| 39 | All Heritage Coast | 179 | 0.75 | 134 | 0.306 |
| 40 | All Millennium & Doorstep Greens | 440 | 0.75 | 330 | 0.286 |
| 41 | All Forestry Commission | 327 | 0.75 | 245 | 0.335 |
| 42 | Other AONB Designation | 1,169 | 0.75 | 877 | 0.336 |

Note that to correct for sampling bias in the MENE survey and to ensure representativeness of the sample, observations were drawn from strata in proportion to their demographic weights; that procedure increased the likelihood of drawing respondents with under-represented demographic profiles and decreased the likelihood of drawing respondents with over-represented demographic profiles.

A sampling weight was determined for each strata (described as the WESML weight in Table 6) that would late be used in estimation to correct for the choice-based sampling in the selection of observations (see Section 5.4). That weight indicates the ratio of the likelihood of a respondent drawn at random from the population having a focus trip to a site in a certain strata, to the likelihood of such an observation being in the sample. The population likelihood was estimated using the demographic weights for the full sample (see section 3.1) and the sample likelihood calculated from the numbers drawn from each strata;

$$w_{s(j_i)} = \frac{Q_{s(j_i^*)}}{H_{s(j_i^*)}} = \frac{Prop\ of\ Pop\ in\ s}{Prop\ of\ Sample\ in\ s} \quad (i = 1, 2, \dots N) \tag{1}$$

where $j_i^*$ indicates the site chosen by respondent $i$, $s(j_i^*)$ identifies the sampling strata for that site and $Q_{s(j_i^*)}$ and $H_{s(j_i^*)}$ are defined as shown in Equation (1).

The final estimation dataset comprised a sample of 51,807 observations, where each observation identified recreation behaviour over 7 consecutive days.

### 4.4 Choice Set Sampling

With the sample of observations to be used in estimation established, the next step in developing the dataset was to define the choice set for each respondent in the sample. The choice set represents the set of sites a respondent might choose from in making the decision as to where to visit for a recreational day trip … as well as the option of not taking a trip at all.

The issue of how to establish choice sets remains an open question in the literature; for a recent review see Thiene, Swait et al. (2017). In this research we assumed that each respondent's choice set consists of all recreation sites in England though, of course, many would be too distant from a respondent's house to ever compete with more proximate recreation sites offering similar experiences. Since the ORVal Greenspace Map identifies 129,575 sites, including each of these explicitly in the choice set of each observation would result in an intractably large estimation dataset. Accordingly, we adopt a form of importance sampling in order to select a sample of sites for each observation with which to model the full choice set.

To select the choice set sample for each respondent, we wanted to ensure that the selected sites included;

(i) a diverse range of different outdoor greenspaces and
(ii) sites that were likely to be important possible recreation locations for that respondent.

To achieve (i), we again used a stratified sampling approach. Sites were categorised into 19 different strata according to their type and dominant landcover. The descriptions of those strata definitions are provided in the second column of Table 7.

**Table 7: Choice set sampling scheme**

| Category | Description | Num Sites | Num Sampled |
|---|---|---|---|
| 0 | No Trip | 0 | 1 |
| 1 | Beaches | 630 | 2 |
| 2 | Cemeteries | 9,494 | 5 |
| 3 | Allotments | 6,865 | 5 |
| 4 | Parks mostly woods | 11,151 | 10 |
| 5 | Parks mostly wetland | 117 | 2 |
| 6 | Parks by sea | 207 | 2 |
| 7 | Parks mostly natural grass | 854 | 2 |
| 8 | Parks mostly moorland | 194 | 2 |
| 9 | Parks mostly managed grass | 14,549 | 10 |
| 10 | Parks by fresh water | 1,141 | 5 |
| 11 | Parks mostly agricultural | 1,782 | 5 |
| 12 | Paths mostly woods | 10,230 | 10 |
| 13 | Paths mostly wetland | 93 | 2 |
| 14 | Paths by sea | 800 | 2 |
| 15 | Paths mostly natural grass | 2,355 | 5 |
| 16 | Paths mostly moorland | 2,833 | 5 |
| 17 | Paths mostly managed grass | 22,504 | 10 |
| 18 | Paths by fresh water | 3,358 | 5 |
| 19 | Paths mostly agricultural | 40,418 | 10 |
| | Total: | 129,575 | 100 |

As shown in the final column of Table 7, a sampling scheme was devised in which the number of sites sampled from a category to be included in an individual's choice set was selected according to the number of sites in each category type. So a category containing more than 10,000 sites (e.g. paths through agricultural land) was sampled 10 times, a category with greater than 1,000 but less than 10,000 sites was sampled 5 times and a category with less than 1,000 sites was sampled twice. Where the respondent had taken a trip to particular greenspace, that greenspace was included in their choice set and one less alternative sampled from the category corresponding to the chosen site. A final category of not taking a trip at all was added to the choice set giving a (sampled) choice set size of 100 options.

To ensure (ii) (that is, that 'important' sites were selected from each strata) we first calculated the straight line distance between the centroid of each site and that respondent's home location. Home locations were defined as the centroids of the Lower Super Output Area (LSOA) identified in the MENE data as the residence location of that respondent. Under the assumption that respondents were more likely to consider visiting larger sites, closer to a respondent's we calculated a weight for each site according to:

$$\omega_{i,j} = \frac{Size_j}{{Distance_{i,j}}^2} \quad (i = 1,2, \dots, N, j = 1, 2, \dots, J) \tag{2}$$

We then randomly selected sites from each strata with replacement and a likelihood of selection proportional to $\omega_{i,j}$.

Finally, we needed to define a weight for each site in the choice set that would be used in estimation to correct for the fact that we are using a sample from the choice set rather than the full choice set itself (see Section 5.4). That weight was calculated as follows:

$$w_{j_i}^1 = \frac{Actual\ \#j\ sampled}{Expected\ \#j\ sampled} \quad (i = 1, 2, \dots N) \tag{3}$$

Where $Expected\ \#j\ sampled$ is the inverse of the probability of sampling site $j$ in strata $s(j)$ as determined by the importance weight, multiplied by the number of observations drawn from $s(j)$ and $Actual\ \#j\ sampled$ is the number of times that site is randomly chosen to appear in the choice set[1]

### 4.5 Travel and Time Cost Calculation

The final step in creating the estimation dataset was to calculate the time and travel costs that would have been incurred by a respondent in travelling to and from each greenspace included in their choice set. For this purpose we extracted the roads network from the OS Meridian 2 product and generated a Network Dataset in the ArcGIS software, defining travel times along roads that differentiated between standard driving speeds on motorways, primary roads, a-roads, b-roads and minor roads. We then employed the 'Origins-Destinations Matrix' function of the ArcGIS tool to calculate the distance and travel time along the roads network between a respondent's house (particularly, the nearest point on the road network and the centroid of the LSOA of residence of the respondent) and each site in their choice set (particularly, the nearest point on the road network to the centroid of that greenspace).

Given the size of the sample, a short Python procedure was written to loop through the sample calling the ArcGIS engine to perform the OD Matrix calculations for each respondent in turn.

[1] To provide a brief intuition as to the functioning of the weight in (3), in estimation we are going to need to calculate a sum across all the sites in a respondent's real choice set; roughly speaking adding up the utility the respondent might have got if they had chosen to visit each site. So imagine that there were four sites in the real choice set and let us label the utility from visiting each of those sites as $u_1$, $u_2$, $u_3$ and $u_4$. Our best guess is that $u_1 > u_2 > u_3 = u_4$ in the ratio 4:2:1:1. Now imagine we wanted to estimate the sum $u_1 + u_2 + u_3 + u_4$ but could only base our guess on that sum through drawing a sample of one observation. Given our best guess of the relative sizes of the four utilities, we could use importance sampling which means we would sample $u_1$ with probability $4/8 = 1/2$, $u_2$ with probability $2/8 = 1/4$ and $u_3$ and $u_4$ both with probability $1/8$. Now if we were to draw site 1 as the single observation in our sample, the weight in (3) would be 1 divided by $1/2$ which is 2. So our best bet at the sum $u_1 + u_2 + u_3 + u_4$ given this single observation would be $2u_1$. Likewise if we were to draw site 2, our best estimate using our importance weights would be $8u_2$. If instead we were to draw a sample of two observations, there would be 2 chances of selecting any site into the choice set such that in calculating weights the denominator of (3) would be doubled. Say we drew sites 2 and 3. The weight for 2 would be $1/(2 \times 2/8) = 2$ and the weight for 3 would be $1/(2 \times 1/8) = 4$. Accordingly, given that sample of two observations selected through importance sampling our best estimate of the sum $u_1 + u_2 + u_3 + u_4$ would be $2u_2 + 4u_3$.

Using data supplied by the AA Motoring Costs publications[2] we estimated the cost of fuel for an average family car over the period of the data to be 9p per km and using that figure calculated the fuel cost of travelling to and from each site for each respondent. In addition we drew on recent research for DfT on the value of travel time to establish a monetary value for time spent travelling for non-work activities (Department for Transport 2015). Those values were £2.30 per hour for trips under 8km, £3.47 per hour for trips between 8km and 32km, £6.14 per hour for trips between 32km and 160km and £9.25 per hour for trips greater than 160km (see Table 7.18 of DfT report). A total monetary cost for travel was taken by adding the time costs to the fuel costs for the return journey.

## 5. The Econometric Model

### *5.1 Econometric Specification*

Our approach to estimating a recreational demand model adopts the long-established random utility framework first proposed by McFadden (1973). That framework characterises recreational decisions as discrete choices in which, on any particular choice occasion, an individual has the opportunity to visit one of an array of sites each offering different opportunities for outdoor recreational activities. In essence, the modelling approach seeks to establish the value of the recreational opportunities offered by sites by observing data recording which particular sites individuals chose to visit given the set of sites that they could have possibly visited.

More formally, our dataset records the outdoor recreational choices of a sample of individuals, indexed $\mathrm{i} = 1, 2, \ldots, \mathrm{N}$, on each of series of days indexed $t = 1, 2, \ldots, T$. That recreational choice concerns which greenspace to visit where greenspaces are indexed $j = 1, 2, \ldots, J$ or whether to undertake some other activity, an option indexed $j = 0$.

The choice as to which greenspace to visit depends on a number of factors, but two important considerations are the quality of the recreational experience offered by a site and the cost in time and money of visiting that site. In our model, the quality of recreational experience offered by site $j$ is determined by the vector of site characteristics $\boldsymbol{x_j}$ and the costs of making a trip to that site $tc_{ij}$.

To construct our econometric model, we first need to posit a function which describes the utility an individual will enjoy if they decided to visit site $j$. In line with the vast majority of the literature we choose the simple linear approximation;

$$v_{ijt} = \alpha_j + \boldsymbol{x}_j \boldsymbol{\beta_1} + \gamma\left(I_{i,t} - tc_{ij}\right) \qquad (j = 1, 2, \ldots, J_i \text{ and } \forall i, t) \tag{4}$$

where, $I_{i,t}$ is individual $i$'s per period income, $\alpha_j$ is a site-specific utility element, $\boldsymbol{\beta_1}$ is the vector of coefficients describing the marginal utilities of site qualities and $\gamma$ is the marginal utility of income.

Alternatively, an individual may choose not to make an outdoor recreational trip. We give that "no trip" option the index $j = 0$, and specify the utility from that option as;

$$v_{i0t} = \alpha_0 + \boldsymbol{z}_{i,t} \boldsymbol{\beta_0} \qquad (\forall i, t) \tag{5}$$

where $\boldsymbol{z}_{i,t}$ is a vector capturing characteristics of the time period (e.g. month of the year, day of the week) and of the individual (e.g. gender, age, socioeconomic segment) whose importance in

[2] Accessed from www.theaa.com/motoring_advice/running_costs/advice_rcosts_guide.html

determining participation in outdoor recreation is captured by the vector of coefficients $\boldsymbol{\beta_0}$, while $\alpha_0$ is some constant utility associated with choosing not to take a trip to greenspace.

Adopting the familiar random utility framework, we develop our econometric specification from (4) and (5) by constructing the conditional indirect utility function;

$$u_{ijt} = v_{ijt} + \varepsilon_{ijt} \qquad (j = 0, 1, \ldots, J \; and \; \forall i, t) \tag{6}$$

where $\varepsilon_{ijt}$ is an econometric error term introduced to capture the divergence between our model of utility $\left(v_{ijt}\right)$ and the individual's experienced utility $\left(u_{ijt}\right)$. Since the scale on which utility is measured is not known, we can make any arbitrary decision as to what quantity represent zero. For the purposes of this analysis we set $\alpha_j = \alpha = 0 \; \forall j$. Given the very large number of sites in the analysis, this assumption amounts to relegating utility derived from idiosyncratic features of each parks to the error term.

In making recreational trip decisions it is assumed that individuals choose from the set of options $j = 0, 1, \ldots, J_i$, selecting that option which gives them the highest utility. Accordingly, the probability of observing individual $i$ choosing to visit site $k$ can be written as;

$$\begin{aligned} P_{ikt} &= Prob\left[u_{ikt} > u_{ijt} \;\; \forall \, j \neq k \,\right] \\ &= Prob\left[v_{ikt} + \varepsilon_{ikt} > v_{ijt} + \varepsilon_{ij} \;\; \forall \, j \neq k \,\right] \\ &= Prob\left[v_{ikt} - v_{ijt} \; > \varepsilon_{ijt} - \varepsilon_{ikt} \;\; \forall \, j \neq k \,\right] \end{aligned} \tag{7}$$

Given $v_{ikt}$ and $v_{ijt}$ are, given parameters $\alpha_0$, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are deterministic, the probability in (7) is determined by the assumptions made regarding the joint distribution of the error terms, $\varepsilon_{it} = \left[\varepsilon_{i0t}, \varepsilon_{i1t}, \ldots, \varepsilon_{iJt}\right]$.

Perhaps the simplest assumption, and one used extensively in the choice modelling literature, is to assume that the error terms are drawn from the family of distributions described as Generalised Extreme Value (GEV) (McFadden 1978). In that case, the probability in (7) is given by;

$$P_{ijt} = \frac{e^{v_{ijt} + lnG_j}}{\sum_{k=1}^{J} e^{v_{ikt} + lnG_k}} \tag{8}$$

Where the function $G(\cdot)$ follows from the particular assumptions made regarding the join distribution of the error terms and must conform to certain properties outlined by McFadden (1978). Also $G_j = \partial G / \partial e^{v_{ijt}}$. The simplest form for GEV results from the assuming that;

$$G = \sum_j e^{v_{ijt}} \tag{9}$$

Which, from (8), results in choice probabilities that define the familiar multinomial logit (MNL) model;

$$P_{ijt} = \frac{e^{v_{ijt}}}{\sum_{k=0}^{J} e^{v_{ikt}}} \tag{10}$$

The great advantage of the MNL model is the simplicity of calculation of choice probabilities which greatly increases computational efficiency in estimating the model parameters. On the other hand the MNL model fails to allow for any form of correlation in the error terms of the different options or their observed attributes, an assumption that leads to predictions of somewhat implausible substitution patterns often referred to as *independence from irrelevant alternatives* (IIA) (McFadden, Tye et al. 1977). In effect, the IIA assumption does not allow for the expectation that the addition of a new option to the choice set will tend to reduce the probability of choosing options than have attributes more like that new option by a greater extent than it will options that are more dissimilar.

WIth the ORVal data, we hypothesised that one feature of greenspaces that will determine the degree to which individuals regard them as substitutes was the similarity of sites in terms of the type of landcovers and land uses present at those sites. Accordingly, we took the range of landcovers used to describes sites (see Table 3) and organised those into 9 broad groups;

- Salt Water
- Fresh Water
- Natural Grass
- Managed Grass
- Agriculture
- Wetlands
- Moors & Heath
- Allotments
- Cemeteries & Graveyards

Such that the landcover of any site could be described in terms of the proportion of its area in each of these different groups.

One way to proceed, would be to identify each site with its dominant landcover thereby classifying all sites as belonging to one of the nine groups. Under the assumption that sites in the same group are considered closer substitutes, an alternative specification of the GEV model would result from assuming;

$$G = \sum_{m=0}^{M} \left( \sum_{j \in B_m} e^{\mu_m v_{ijt}} \right)^{1/\mu_m} \tag{11}$$

Where the sites in each distinct landcover group form the set $B_m$ and $m = 1, 2, \dots, M$ indexes the different landcover groups. Notice we have added an additional group, $m = 0$, which has the single member consisting of the option not to take a recreational trip. Notice also, the group-specific parameters, $\mu_m$, which allow for the fact that sites in a group may be similar to each other in some unobserved way. As shown by McFadden (1978), this similarity parameter should vary on the range from 1 to ∞. When $\mu_m$ is large then individuals regard the sites in group $m$ as very similar and hence treat them as close substitutes. In contrast when the $\mu_m = 1$ the sites in the group are considered no more similar to each other then they are to any other site; indeed if $\mu_m = 1$ for all $m$ (11) reduces to (9) and we are back at the MNL model. Replacing (11) in (8) results in the specification of a GEV known as the nested multinomial logit model (NMNL) which is thoroughly reviewed and described in Morey (1999).

Constraining similarity between sites to be dictated by the dominant landcover ignores the fact that each site is actually a mosaic of landcovers such that each site may share similarities with a variety of groups. A specification of $G(\cdot)$ that captures that possibility is given by;

$$G = \sum_{m=0}^{M} \left( \sum_{j=0}^{J} \alpha_{jm} e^{\mu_m v_{ijt}} \right)^{1/\mu_m} \tag{12}$$

(12) differs from (11) with regards to the parameters $\alpha_{jm}$ which dictate the 'share' of site $j$ that should be apportioned to similarity group $m$.

In the ORVal model we set $\alpha_{jm}$ to be the proportion of site $j$'s land area that is of landcover $m$. With this specification, therefore, a site is seen as similar to other sites with which it shares landcovers but more similar to sites with which it has more landcover in common. Replacing (12) in (8) results in a the specification of a GEV model known as the cross nested logit model (CNMNL) first proposed by Ben-Akiva and Bierlaire (1999) and reviewed in detail by Bierlaire (2006).

Compared to other possible GEV specifications, the CNMNL admits rich patterns of substitution between greenspaces that reflect the similarities in environmental experience offered by the different sites. From (8) we see that the mathematical form of the CNMNL choice probability, while more complex than the MNL model, remains reasonably tractable.

In passing we note that the partial derivative in (8) for the CNMNL takes the form;

$$G_j = \frac{\partial G}{\partial e^{v_{ijt}}} = \sum_{m=0}^{M} \alpha_{jm} e^{v_{ijt}(\mu_m - 1)} \left( \sum_{j=0}^{J} \alpha_{jm} e^{v_{ijt}\mu_m} \right)^{1/\mu_m - 1} \tag{13}$$

such that the choice probabilities can be written as;

$$P_{ijt} = \frac{e^{v_j + \ln G_j\left(\sum_{h=0}^{J} \alpha_{hm} e^{v_h \mu_m}\right)}}{\sum_{k=0}^{J} e^{v_k + \ln G_k\left(\sum_{h=0}^{J} \alpha_{hm} e^{v_h \mu_m}\right)}} \tag{14}$$

Where the notation $G_j\left(\sum_{h=0}^{J} \alpha_{hm} e^{v_h \mu_m}\right)$ is included to make explicit the fact that the partial derivative (13) is a function of a sum across all greenspaces in the choice set.

One thing to note about the form of CNMNL model shown in (14) and adopted in the ORVal Greenspace Model is that it makes no accommodation for the fact that our data contains observations of the same individual making choices across multiple choice occasions. This is an assumption we hope to relax in future advances of the model.[3]

Given data on the recreational choices of the $N$ individuals in $T$ time periods, it follows from (14) that the log of the likelihood of observing those choices is;

[3] Note that subsequently we employ clustered robust standard errors to account for the lower information content provided by repeated responses from the same individual.

$$ln\,L(\alpha_0, \boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \gamma, \boldsymbol{\mu}) = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{J} Y_{ijt}\,ln\,P_{ijt} \tag{15}$$

Where $Y_{ij}$ is a dummy variable which takes the value 1 if individual $i$ chose recreational option $j$, or zero otherwise and $\boldsymbol{\mu}$ is the vector of similarity parameters. The parameters of the model can be estimated using maximum likelihood methods by optimising (15) with respect to the parameters $\alpha_0, \boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \gamma$ and $\boldsymbol{\mu}$.[4]

Since (15) can be highly non-linear we use a global search algorithm, the Nelder-Mead simplex algorithm (Nelder and Mead 1965) using the parameterisation suggested by Gao and Han (2012) for high dimensional problems.

### 4.3 Welfare Estimation

As shown in equation (7), an important feature of GEV models like the CNMNL is that they are firmly based on a theory of random utility maximisation. Indeed, provided empirical estimation of the model results in $\mu_m \geq 1$ $(m = 0,1, \dots, M)$, then the model is globally consistent with that theory (Kling and Herriges 1995).

One useful property of GEV models that follows from that fact, is that there exists a simple closed-form expression for the expectation of the maximum utility a respondent might expect to derive from being able to choose an option from their choice set. In the case of the CNMNL model that expression amounts to;

$$V_{it}(J') = \ln \sum_{m=1}^{M} \left( \sum_{j \in J'} \alpha_{jm} e^{v_{ijt}\mu_m} \right)^{1/\mu_m} + \lambda \tag{16}$$

where $V_{it}(J')$ is the expectation of maximum utility realised by individual $i$ in time period $t$ given the opportunity to choose from the choice set $J'$, and $\lambda$ is the Euler-Mascheroni constant (that takes a value of 0.5772 to 4 decimal places).[5]

It follows that the expected level of welfare change that an individual would experience if the nature of their choice set were to change, perhaps through the loss or gain of sites from the choice set and/or changing the qualities of sites (Small and Rosen 1981);

$$\Delta W = \frac{1}{\gamma}\big(V_{it}(J'') - V_{it}(J')\big) \tag{17}$$

[4] Under two circumstances the MENE data records that the respondent has taken a trip on that choice occasion but does not record which greenspace was the visited. First, for days in the respondent's week-long dairy record where a trip was taken but that choice occasion was not randomly selected as the focus trip. Second, where we were unable to identify the location of the focus trips (see Section 4.2). On those occasions all we know is that the respondent chose to take a trip to some greenspace or, put another way, decided not to choose the outside option indexed 0. Accordingly, under both those circumstances, we record the probability of the choice as $1 - P_{i0t}$

[5] The derivation of this formula arises from interpreting the conditional indirect utilities of each option (see equation (6)) as random variables and calculating the expected maximum of that set.

where $J'$ is the original choice set and $J''$ the changed choice set. In simple terms, equation (17) describes the analyst's best estimate of how an individuals' utility will change as a result of changes in the choice set with that quantity translated into money terms by dividing that utility change by the marginal utility of income, $\gamma$.

### 5.4 Econometric Corrections

The econometric model as defined by (15) fails to correct for a number of features of the data used in estimation of the model. IN the first instance, the specification in (15) assumes random sampling, where the data used in estimating the ORVal model were drawn using the choice-based sampling strategy described in Section 4.3. To correct the likelihood we use the weighted exogenous sampling maximum likelihood (WESML) estimator as follows;

$$ln\, L = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=0}^{J} Y_{ijt} w_{s(j_i^*)}\, ln\, P_{ijt} \tag{18}$$

where $w_{s(j_i^*)}$ is the choice-based sampling weight defined in (1). In effect the weight acts to correct the log likelihood function, decreasing the importance of observations that have been over-sampled in drawing a choice-based sample and increasing the importance of observations that have been under-sample. Manski and Lerman (1977) show that the WESML estimator provides consistent estimates of the model parameters.

A second issue with the ORVal model data set is the sampling of sites for inclusion in the choice set for each sample respondent. Recall from (8) that the GEV probability for choice option $j$ takes the form of a relatively simple proportion relating the utility from a visit to site $j$ to an aggregations fo the utilities of visits to all sites in the choice set. More specifically, in the CNMNL model the numerator of the probability is the exponentiated utility from a visit to site $j$ plus a term that captures the degree of similarity of that site to others in the choice set, while the denominator is the sum of exponentiated utilities plus similarity terms for the entire choice set. Clearly, when we use a sample of options in the choice set two errors arise in the choice probability. First, the ratio of numerator to denominator is biased since we fail to sum over the full choice set in the denominator. Second the similarity term is biased since it fails to aggregate over all sites similar to site $j$.

To address these biases, Guevara and Ben-Akiva (2013) propose a correction to the choice probability of the form;

$$P_{ijt} = \frac{w_{ij}^1 e^{v_{ijt} + \ln \hat{G}_j\left(\sum_{h\in \tilde{J}_i^2} w_{ih}^2 \alpha_{ihm} e^{v_h \mu_m}\right)}}{\sum_{k\in \tilde{J}_i^1} w_{ik}^1 e^{v_{ikt} + \ln \hat{G}_k\left(\sum_{h\in \tilde{J}_i^2} w_{ih}^2 \alpha_{ihm} e^{v_h \mu_m}\right)}} \tag{19}$$

where the weights $w_{ij}^2$ $\left(j \in \tilde{J}_i^1\right)$ are calculated as per (3) to reflect the relatively probability of an alternative appearing as one of the options in the sampled choice set, $\tilde{J}_i^1$. Likewise $w_{ih}^2$ corrects the aggregation over sites that appears in the similarity terms. Accordingly, we denote these similarity terms by the functions $\hat{G}_j(\cdot)$ to make clear that the aggregation over choice sets used in their calculation is an estimate based on the choice set sampling weights, $w_{ih}^2$ $\left(h \in \tilde{J}_i^2\right)$. Notice that as per the recommendation of Guevara and Ben-Akiva (2013) we sample a second set of options to form the choice set used to calculate the similarity terms, $\tilde{J}_i^2$ and the weights $w_{ih}^2$ are calculated from this sample as per (3).

### 4.7 Covariate Choice

The final step in developing the ORVal recreation demand model is to determine the set of covariates that will be used to describe the participation choice, $\boldsymbol{z}_{i,t}$ $(i = 1, \dots, N, t = 1, \dots, T)$, and those to be used to describe the utility benefits of a trip to a site $\boldsymbol{x}_j (j = 1, \dots, J)$.

In the model, the choice of whether to take a trip or not was made a function of three groups of variables; those that described the time when the trip was taken, those that described the location of residence of a respondent, and those that described a respondent's socio-demographic characteristics of a respondent;

1. Time: We captured the time dimension through a set of dummy variables for the year (using 2009) as the base case, a set of dummy variables for month of the year (using December as the base case) and a set of dummy variables for day of the week (using Monday as the base case).
2. Location: Location of residence was represented by a set of dummy variables coding for a respondent's Government Office Region (GOR) using the East Midlands as a base case.
3. Sociodemographics: A key consideration in defining variables to describe a respondent's sociodemographic characteristics was the subsequent need to transfer the ORVal recreation demand model to predict the behaviour of all (adult) individuals in England (see Section 6.2). While the MENE survey collected numerous details of the sociodemographics of the survey respondents, our information for the wider population is limited to data provided by the 2011 census and collated at the level of Lower Super Output Area (LSOA). Accordingly, our selection of variables by which to describe the sociodemographics of respondents was restricted to those provided in both the MENE survey and the 2011 census. In particular, we defined dummy variables identifying age-gender groups, individuals with children, working status, and socioeconomic segment (using the six category – A, B, C1, C2, D, E - socioeconomic classification produced by the ONS). While ownership of a dog was not recorded in the census we included this as a covariate and used data from elsewhere to approximate that value in the transfer exercise (see Section 6.2).

As shown in (4), we assume that the utility derived from visiting a greenspace comprises the trade-off between a cost and a benefit. The cost comes in the form of the time and travel expenses incurred in getting to and from that greenspace; the travel cost, $tc_{ij}$, calculated as explained in Section 4.5. The benefits, it is assumed, are derived from the various qualities of the greenspace. We capture those qualities through a series of sets of covariates;

4. Greenspace Type: To establish differences in utility offered by different broad categories of greenspace, we created a dummy variable set distinguishing paths from beaches, from country parks, from allotments from graves/cemeteries leaving other parks (see definition in Section 3.2) as the base case.
5. Size and Landcover Composition: The nature of the greenspace with which an individual interacts when visiting a site is captured in the ORVal greenspace model through a series of variables that record the natural log of the total area of the greenspace (in hectares), and

the natural log of the areas of each landcover from which that greenspace is composed. With regards to the latter we identify the area (in hectares) of each park dominated by the following 17 landcover types;

- Woods
- Wood Pasture
- Agriculture
- Natural Grass
- Moors
- Coastal
- Saltmarsh
- Marsh & Fen
- Managed Grass
- Sports Pitches
- Gardens
- Allotments
- Cemeteries
- Sea
- Estuary
- River
- Lake

In addition to the quantities of different landcovers, the specification includes a variable which describes the diversity of landcovers accessible from a greenspace calculated using Simpson's Index of Diversity. In particular, we calculate the proportion of a greenspace under each land cover type according to;

$$prop_{l,j} = \frac{area\ under\ landcover\ l\ at\ site\ j}{total\ area\ of\ site\ j} \quad (l = 1, \dots, L; j = 1, \dots, J) \tag{20}$$

We then calculate the diversity index as;

$$diversity_j = \frac{1}{\sum_l^L prop_{l,j}{}^2} \quad (j = 1, \dots, J) \tag{21}$$

Observe that the lowest possible value of the index is 1 where the greenspace has only one landcover but increases in the number of landcovers accessible at that site. For example, with two land covers the index can take a value in the range 1 to 2, where an index near 1 would indicate only a small part of the greenspace having the second landcover and an index of 2 would arise when the greenspace has equal areas of the two landcovers. Likewise, with three landcovers the index can take a value in the range 1 to 3 with the upper bound again identifying an equal split of area between the three landcovers.

6. Commonalities: One complication with the definition of greenspaces is in defining what constitutes an independent recreation site, for example in circumstances where the ORVal greenspace map identifies greenspaces that share common boundaries (though see section 4.15 of the ORVal Greenspace Map Report). Ignoring the fact that a greenspace borders another greenspace may understate its qualities since individuals visiting that site may also take advantage of the greenspace provided by the adjoining site. In that case, we might

think that commonalities between the borders of greenspaces might indicate complementarities not otherwise captured in our model.
For path sites, defined as access points to a path network, the issue of commonality is likely to act in the opposite direction. In this case we define the commonality to be the area of overlap in the path network accessible from a particular access point. Where multiple path sites access the same path network then those different sites are likely to represent close substitutes. Accordingly, we might expect that a path site with more commonalities (and hence more close substitutes) will receive fewer visits than for an identical path site with no commonalities.
The issue of commonalities has received attention in the transport literature concerned with route choice. In the ORVal model we adopt the proposal of Cascetta, Nuzzolo et al. (1996) where they define a variable that captures the degree of commonality for option according to;

$$CF_j = \ln \sum_k \left( \frac{L_{jk}}{L_j^{1/2} L_k^{1/2}} \right)^{\rho} \quad (22)$$

Where $CF_j$ is the commonality factor for option $j$, $L_{jk}$ is the extent of commonality between site $j$ and site $k$, $L_j$ is the total extent of site $j$, $L_k$ is the total extent of site $k$ and $\rho$ is a parameter that we set to the value 1 in calculation of the commonality factors. In the case of parks, the extent of a site is taken to be its perimeter and the extent of commonality with another park is the extent of that perimeter that lies within 25m of that other park. In the case of paths, the extent of a site is taken to be the linearly decayed area of path accessible from a path access point, and the commonality with another access point is the extent of that area accessible from that other access point.

7. Designations: The ORVal greenspace map records a variety of special designations given to the different recreational sites. For the purposes of the ORVal model we assume that those designations may capture aspects of the environmental experience of visiting a greenspace that are not captured by descriptions of type or landcover. Accordingly we define a series of binary variables identifying sites with the following designations.
    - National Park
    - AONB
    - CROW
    - Heritage Coast
    - Historic Park
    - Millennium or Doorstep Green
    - Nature

   Note that the ‘Nature’ category includes designation as a local nature reserve, national nature reserve, a Natura2000 site, a RAMSAR site, SSSI and ancient woodlands.
8. Points of Interest: The final set of variables used to describe the quality of greenspaces are a set of binary variables identifying the presence of a series of possible points of interest;
    - Archaeology
    - Historic Building
    - Scenic Feature
    - Playground
    - Viewpoint

Since we suspected that the recreational experience associated with a path-type site may differ from that of a park-type site (see definitions in Section 3.2) we define separate sets of site quality variables for paths and parks; that is to say, we have one set of size & landcover, commonality, designation and points of interest variables defined for parks and another set for paths.

## 6. Results

### *6.1 Parameter Estimates*

The model described in Section 5 was estimated using a custom routine written in the Gauss programming language. Given the size of the data set upon which estimation was based (50,720 respondents with 100 choice options for each respondent on each of 7 choice occasions), and despite use of parallelisation to speed up estimation, the estimation procedure took several days to converge on maximum likelihood parameter estimates. As a result, of the short time frame for delivery of the project, this prevented detailed explorations of the model specification.

In addition, since the key function of the estimated model would be in predicting how recreational activity and the welfare it generates might change as the quality and availability of greenspaces is varied, the model was estimated imposing constraints on the signs that could be taken by coefficient estimates. For example, we suppose that adding more expanse to a site of any particular natural land cover cannot decrease the utility offered by that site. Accordingly, we constrain the parameter estimated on areas of land cover to be non-negative. Likewise we suppose that endowing a greenspace with some designation should not reduce the benefits it affords visitors and hence we again constrain designation parameters to the positive line.

Table 8 lists parameter estimates for the participation choice. The coefficient estimates listed in the second column should be interpreted as indicating how a unit increase in the variable impacts on the utility of choosing not to take a trip to an outdoor recreation site and hence the likelihood of choosing not to visit a greenspace on any particular choice occasion. To illustrate, the constant in Table 8 is positive and takes a large value of 7.9906, indicating that on any particular choice occasion respondents have a high probability of not making an outdoor recreation trip.

Alternatively, one can read positively-signed coefficients in Table 8 as indicating variables that decrease participation in outdoor recreation and negatively-signed ones as variables that increase the likelihood of participation. For ease of expression (i.e. avoiding having to phrase as "increase not participating"), in the following discussion we will use this latter interpretation. In other words, the large positive constant indicates that, in general, on any particular choice occasion there is a relatively low probability an individual will choose to take an outdoor recreation trip as compared to doing something else with their day. The final two columns of Table 8 describe the statistical significance of the coefficient estimates, where significances are based on clustered robust standard errors that account for repeated observations of choices by the same individuals.

**Table 8: Model coefficients for participation decision**

| Variable | Coefficient | *p*-Value | Signif. |
|---|---|---|---|
| Constant | 7.9906 | <0.0001 | *** |
| 2010 | 0.1430 | 0.0069 | *** |
| 2011 | -0.0055 | 0.9149 | |

| | | | |
|---|---|---|---|
| 2012 | -0.0038 | 0.9405 | |
| 2013 | -0.1125 | 0.0255 | ** |
| 2014 | -0.0219 | 0.6700 | |
| 2015 | -0.0117 | 0.9119 | |
| Jan | -0.1517 | 0.0290 | ** |
| Feb | -0.1270 | 0.0864 | * |
| Mar | -0.2773 | <0.0001 | *** |
| Apr | -0.2473 | 0.0002 | *** |
| May | -0.2427 | 0.0003 | *** |
| Jun | -0.2571 | 0.0001 | *** |
| Jul | -0.2668 | 0.0001 | *** |
| Aug | -0.2375 | 0.0005 | *** |
| Sep | -0.1642 | 0.0176 | ** |
| Oct | -0.1861 | 0.0079 | *** |
| Nov | -0.1361 | 0.0533 | * |
| Mon | 0.5320 | <0.0001 | *** |
| Tues | 0.6687 | <0.0001 | *** |
| Wed | 0.6094 | <0.0001 | *** |
| Thur | 0.5736 | <0.0001 | *** |
| Fri | 0.5646 | <0.0001 | *** |
| Sat | 0.2503 | <0.0001 | *** |
| East | -0.1060 | 0.1041 | |
| London | 0.8873 | <0.0001 | *** |
| North East | 0.0047 | 0.9543 | |
| North West | 0.3223 | <0.0001 | *** |
| South East | 0.0669 | 0.2787 | |
| South West | -0.3248 | <0.0001 | *** |
| West Midlands | 0.1761 | 0.0091 | *** |
| Yorks & Humber | 0.0258 | 0.7018 | |
| Male 16-25 | -0.1173 | 0.1387 | |
| Female 26-35 | -0.5292 | <0.0001 | *** |
| Male 26-35 | -0.4019 | <0.0001 | *** |
| Female 36-45 | -0.5628 | <0.0001 | *** |
| Male 36-45 | -0.4953 | <0.0001 | *** |
| Female 46-55 | -0.5379 | <0.0001 | *** |
| Male 46-55 | -0.4847 | <0.0001 | *** |
| Female 56-65 | -0.6317 | <0.0001 | *** |
| Male 56-65 | -0.6087 | <0.0001 | *** |
| Female 65+ | -0.3806 | <0.0001 | *** |
| Male 65+ | -0.6374 | <0.0001 | *** |
| Children | -0.0530 | 0.1396 | |

| | | | |
|---|---|---|---|
| Dog | -1.7680 | <0.0001 | *** |
| Seg A | -0.7555 | <0.0001 | *** |
| Seg B | -0.7233 | <0.0001 | *** |
| Seg D | -0.1142 | 0.0796 | * |
| Seg C1 | -0.5072 | <0.0001 | *** |
| Seg C2 | -0.2275 | 0.0002 | *** |
| Work Part-Time | -0.0703 | 0.1568 | |
| Work Full-Time | 0.2113 | <0.0001 | *** |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

The first observation to make with regards to the parameter estimates in Table 8 is the large number of significant coefficients; an observation which suggests that the model has identified considerable levels of regularity in the participation decisions of respondents relating to time, location and sociodemographics.

The monthly and daily indicator variables, for example, reveal evidence of significant variation in greenspace participation over the course of the year. Those parameters are plotted in Figure 2 and Figure 3 where we have reversed the sign of the coefficients so that the magnitudes can be interpreted as indicating the relative likelihood of taking a trip. In Figure 2, for example, the base case is taken as December and this is seen to be the month in which outdoor recreation participation is at its lowest. Perhaps not surprisingly, day trips to greenspaces are seen to be most frequent at Easter (March) and in the summer months particularly June and July.

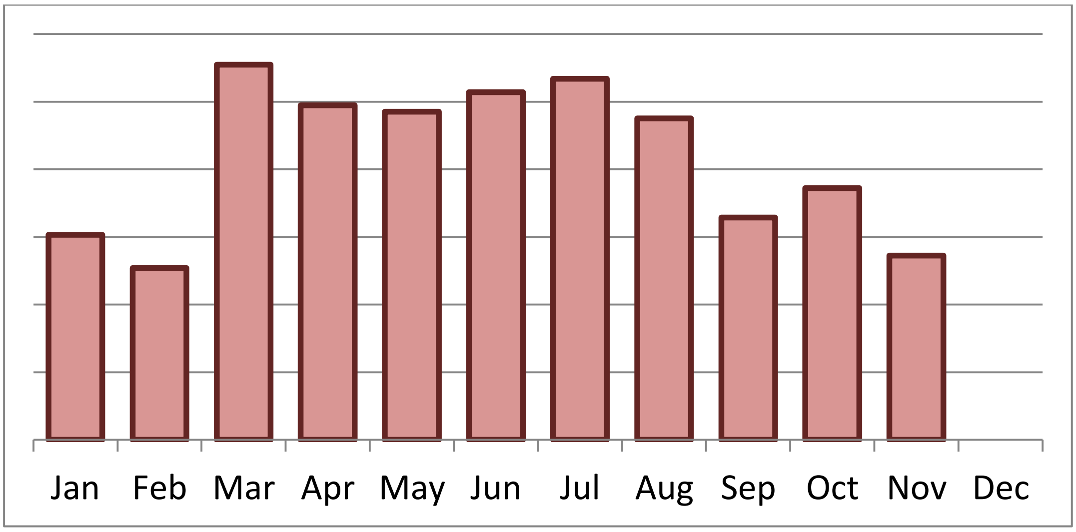


**Figure 2: Relative magnitude of participation likelihood across the year (base case December)**

In a similar vein, Figure 3 reveals a pattern of daily participation likelihoods that conform to prior expectations; individuals are most likely to take trips on weekends and least likely on Tuesdays.

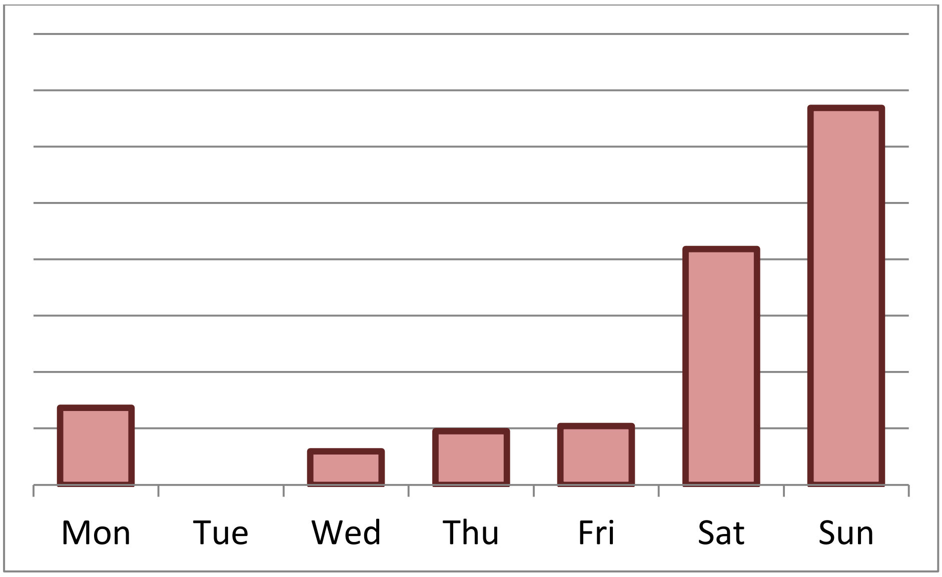


**Figure 3: Relative magnitude of participation likelihood across the week (base case Tuesday)**

Participation also demonstrates regional variation. As shown in Figure 4, participation rates tend to be significantly higher in the South West and significantly lower in the major urban areas of England in London, the North West and the West Midlands. Note that the model controls for the availability of greenspaces in this different regions, such that (barring misspecification bias) the coefficients depicted in Figure 4 show different propensities to participate in outdoor recreation amongst individuals residing in different regions. A possible explanation for this observation is that urban areas offer residents a access to more alternative activities that may substitute for trips to the outdoors. Alternatively, the data may reflect some form of sorting in which individuals that are more inclined to outdoor recreation locate themselves in more rural regions of the country.

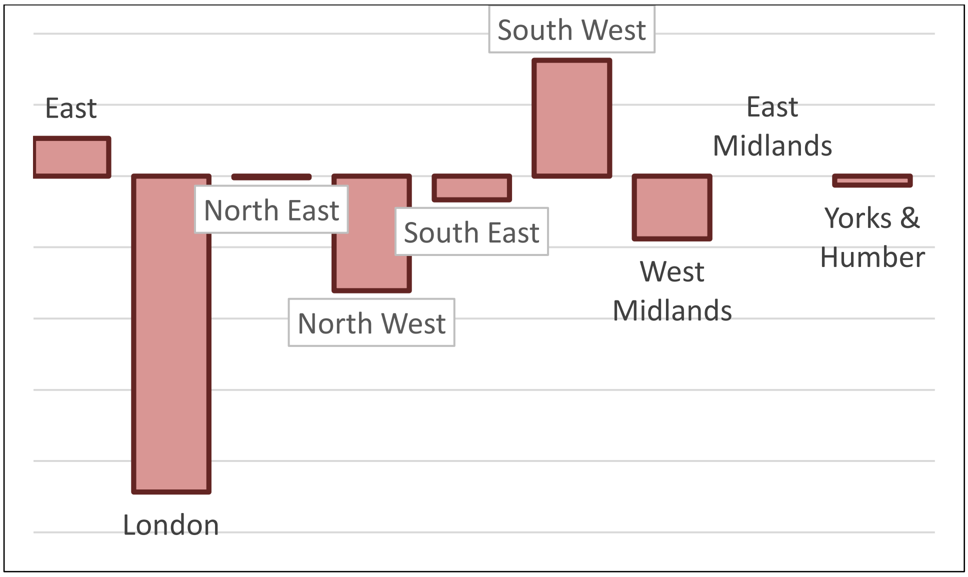


**Figure 4: Relative magnitude of participation likelihood across Government Office Regions (base case East Midlands)**

Figure 5 shows that there are significant differences in participation across age-gender groups. Here females in their 20s are the base case and that group are shown to have the lowest likelihood of participation in outdoor recreation. In general, participation grows in both genders with increasing age with exception of a notable down turn amongst females in the oldest age bracket.

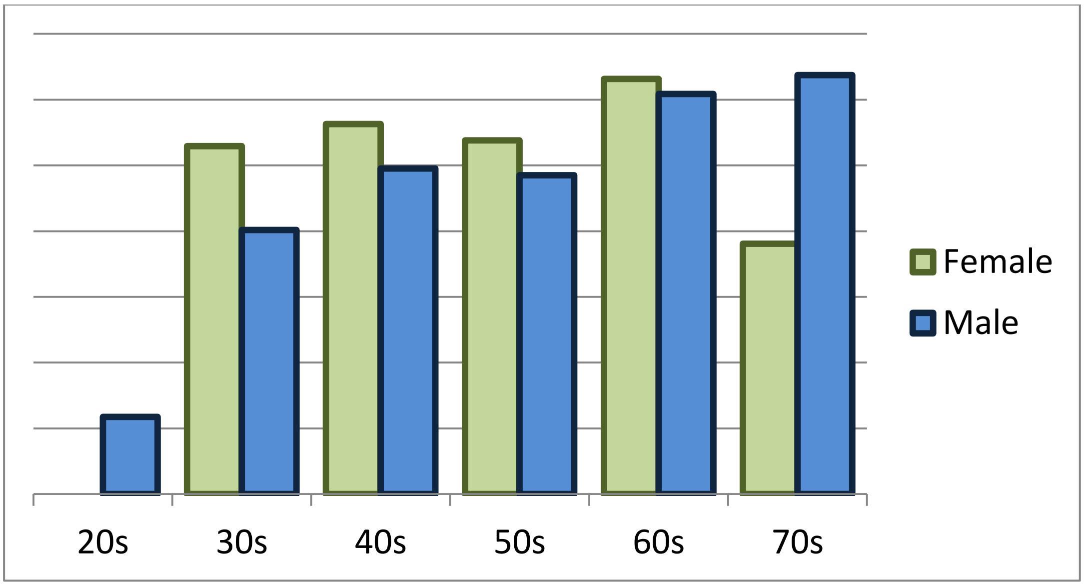


**Figure 5: Relative magnitude of participation likelihood across age-gender groups (base case females in their 20s)**

The data also reveal very significant differences in participation according to socioeconomic segment. As illustrated in Figure 6, participation is highest amongst the A and B categories (defined as higher & intermediate managerial, administrative, professional occupations) falling to the D and E categories (defined as semi-skilled & unskilled manual occupations, unemployed and lowest grade occupations). Again, the strong effect of social grade on participation should be considered as being an effect that is independent of the availability of greenspace; availability is controlled for in the model.

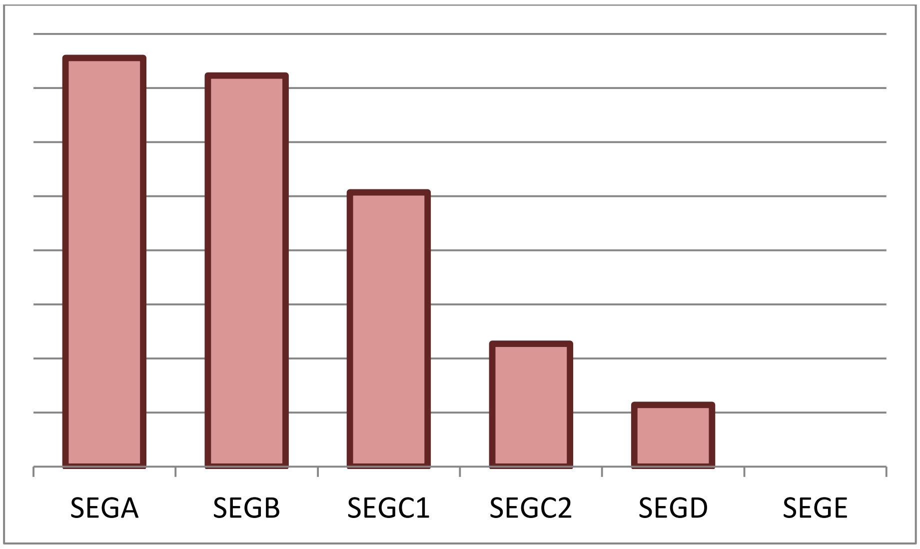


**Figure 6: Relative magnitude of participation likelihood across socioeconomic groups (base case segment D)**

Taken at face value, the coefficient estimates provide evidence to support the idea that outdoor recreation has the properties of what economists call a 'luxury good'. Alternatively, the different participation rates amongst socioeconomic grades could reflect some form of mis-specification bias in the model. One possibility is that the model assumes that travel to sites is possible using a privately owned car. Clearly for some socioeconomic grades car ownership may be limited such that the model overestimates the availability of sites to such groups and mistakes this for a lack of propensity to engage in outdoor recreation. Possibilities for addressing this issue are being investigated.

Of the other variables included in the participation decision equation, we observe that having a dog significantly increases the frequency of trips to greenspaces while working full time significantly decreases participation.

Table 9 to Table 15 document coefficient estimates for the parameters of the site choice decision. As might be expected travel cost turns out to be a hugely important factor in determining choices. As shown in Table 9 the travel cost parameter is negative indicating that site utility is declining in the costs of accessing that site and hence reducing the probability that more distant sites will be visited.

**Table 9: Travel cost coefficient of site choice decision**

| Variable | Coefficient | P-Value | Signif. |
|---|---|---|---|
| Travel Cost | -0.1885 | <0.0001 | *** |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

The data also reveal significant differences in the utility provided by trips to different broad types of greenspace. As recorded in Table 10, compared to a standard park sites, paths, allotments and cemeteries tend to offer less on-site utility for recreation and hence attract less visits while beaches tend to attract significantly more. Those empirical findings chime well with prior expectations.

**Table 10: Greenspace type coefficient of site choice decision**

| Variable | Coefficient | P-Value | Signif. |
|---|---|---|---|
| Beach | 1.0965 | <0.0001 | *** |
| Path | -0.4097 | 0.0036 | *** |
| Country Park | 0.0683 | 0.280 | |
| Allotment | -2.8323 | <0.0001 | *** |
| Cemetery/Graveyard | -1.8604 | 0.0067 | *** |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

In the main, the coefficients on the size, diversity and commonalities of greenspaces also conform to prior expectations. For both parks and paths the utility of a visit is increasing in the log of area. The same is true for the diversity of landcovers at park sites, though for paths the coefficient on paths hits the a priori positivity constraint and is recorded in the model as having no impact on visit utility. The commonality coefficients show significant but oppositely signed impacts on parks and paths. For parks the sign is positive reflecting our expectation that greenspaces that border each other offer complementarities that tend to increase visitation. In contrast, path sites that provide access to stretches of path network serviced by other path sites receive less than anticipated visits, an observation we ascribe to the close substitutability of visits to path sites which access the same network.

**Table 11: Size, diversity and commonality coefficients of site choice decision**

| Variable | Park | | | Path | | |
|---|---|---|---|---|---|---|
| | Coef | *p*-value | Sig | Coef | *p*-value | Sig |
| Area (ln) | 0.2424 | <0.0001 | *** | 0.1732 | 0.034 | ** |
| Diversity | 0.0920 | <0.0001 | *** | 0 | . | |
| Commonality | 0.0309 | 0.0087 | *** | -0.4584 | <0.0001 | *** |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

Table 12 reports the coefficients estimated on areas of land cover. The first thing to observe is that the model does not apparently provide much evidence of a strong impact of landcovers on site utility. We find that for parks, wood pasture, managed grass and sports pitches all have significant positive effects while for paths wood pasture, agriculture and natural grass are all significant and positive. The remaining coefficients tend to be small or set to zero by the non-negativity constraints.

**Table 12: Landcover coefficients of site choice decision**

| Variable | Park | | | Path | | |
|---|---|---|---|---|---|---|
| | Coef | *p*-value | Sig | Coef | *p*-value | Sig |
| Woods (ln) | 0 | . | | 0.0461 | 0.1775 | |
| Wood Pasture (ln) | 0.0108 | 0.5629 | *** | 0.1226 | 0.0061 | *** |
| Agriculture (ln) | | | | 0.1186 | 0.0036 | *** |
| Natural Grass (ln) | 0 | . | | 0.0646 | 0.0978 | ** |
| Moors (ln) | 0 | . | | 0 | . | |
| Coastal (ln) | 0 | . | | 0 | . | |
| Saltmarsh (ln) | 0 | . | | 0 | . | |
| Marsh & Fen (ln) | 0 | . | | 0.0055 | 0.9521 | |
| Managed Grass (ln) | 0.071 | 0.0002 | *** | 0.0379 | 0.2865 | |
| Sports Pitches (ln) | 0.1152 | 0.0001 | *** | | | |
| Gardens (ln) | 0.066 | 0.2411 | | | | |
| Allotments (ln) | 0.2447 | 0.2754 | | | | |
| Cemeteries (ln) | 0 | . | | | | |
| Sea (ln) | 0 | . | | 0 | . | |
| Estuary (ln) | 0 | . | | 0 | . | |
| River (ln) | 0 | . | | 0.0621 | 0.1453 | |
| Lake (ln) | 0 | . | | 0.0096 | 0.8255 | |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

Some of this apparent insensitivity of the model to differences in landcover can be explained through the additional estimation of the similarity parameters also defined on the basis of landcover. Table 13 presents the similarity parameter estimates. Recall from Section 5.1 that increasing values for these parameters signals that respondents regard sites in a landcover group as increasingly similar and hence as closer mutual substitutes. Notice that barring the salt water group all similarity parameters take a value that is significantly different from 1 (where a value of 1 indicates no greater level of similarity between sites within that group than without).

**Table 13: Similarity coefficients of site choice decision**

| Variable | Coefficient | P-Value | Signif. |
|---|---|---|---|
| Salt Water | 1.000 | . | - |
| Fresh Water | 1.3147 | <0.0001 | *** |
| Managed Grass | 1.3881 | <0.0001 | *** |
| Agriculture | 1.4966 | <0.0001 | *** |
| Natural Grass | 1.5634 | <0.0001 | *** |
| Wetlands | 1.3137 | <0.0001 | *** |
| Moors & Heath | 1.2119 | 0.0009 | *** |
| Woods | 1.3795 | 0.0001 | *** |
| Allotments | 1.3458 | 0.0001 | *** |
| Cemeteries | 1.1077 | 0.4251 | |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

Importantly, given the apparent lack of significance of the landcover parameters, it is clear that the similarity parameters lead to a model which will make different predictions regarding the likelihood of choosing a site based on the composition of landcovers present at that site. Likewise, as a result of the similarity parameters, the models predictions of the welfare change resulting from changing a site's landcovers will differ according to the type of landcover change. To illustrate if we differentiate equation (16) and calculate the marginal value of expanding site $k$ by adding 1 more hectare of landcover $r$, denoted $lc_{rk}$ we derive the equation;

$$\frac{\partial V_{it}}{\partial lc_{rk}} = \sum_{m=1}^{M} P_i(m)P_i(k|m)\left(\frac{\beta_{area}}{A_k} + \frac{\beta_{lc_r}}{lc_{rk}} + \beta_{div}\frac{\partial div_k}{\partial lc_{rk}} - \frac{1}{\mu_m A_k}\right) + P_i(r)P_i(k|r)\left(\frac{1}{\mu_r lc_{rk}}\right)$$

where $V_{it}$ is the welfare individual $i$ realises from their current choice set, $P_i(m)P_i(k|m)$ is the probability that individual $i$ chooses option $k$ in landcover category $m$, $A_k$is the size of site $k$, $\beta_{area}$ is the parameter on the site size variable, $\beta_{lc_r}$ is the parameter on the landcover $r$ variable, and $\beta_{div}$ is the parameter on the site diversity variable, $div_k$. Note that the utility change includes two elements that contain similarity parameters. One, $(\mu_m A_k)^{-1}$, can be thought of as a similarity effect on utility working through the change in the overall size of the site. The second, $(\mu_r lc_{rk})^{-1}$, can be thought of as a similarity effect on utility working through the expansion of landcover $r$. In other words, measures of the welfare change of changing landcovers will be differentiated through the effect of the similarity parameters even of the direct effect of land cover captured by $\beta_{lc_r}$ is zero.

Table 14 describes coefficient estimates for the designation indicator variables. We observe positive effects on utility from National Parks though this is only significant for path sites, and positive and significant effects for paths with Heritage Coast designation and parks classified as being of historical significance.

**Table 14: Designation coefficients of the site choice decision**

| Variable | Park | | | Path | | |
|---|---|---|---|---|---|---|
| | Coef | *p*-value | Sig | Coef | *p*-value | Sig |
| National Park | 0.035 | 0.7625 | | 0.1988 | 0.0486 | ** |
| AONB | 0 | . | | 0 | . | |
| CROW | 0 | . | | 0 | . | |
| Heritage Coast | 0 | . | | 0.2793 | 0.0702 | ** |
| Historic Park | 0.5023 | <0.0001 | *** | | | |
| Millennium Green | 0.1479 | 0.2963 | | | | |
| Nature | 0 | . | | 0 | . | |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

Table 15 records coefficient estimates for the points of interest indicator variables. We observe positive and significant effects for each of the identified points of interest barring viewpoints.

**Table 15: Points of interest coefficients of the site choice decision**

| Variable | Coefficient | *p*-value | sig. |
|---|---|---|---|
| Archaeology | 0.1340 | 0.0435 | ** |
| Historic Building | 0.1826 | 0.0033 | *** |
| Scenic Feature | 0.3380 | <0.0001 | *** |
| Playground | 0.5065 | <0.0001 | *** |
| Viewpoint | -0.0323 | 0.55 | |

* >90% level of confidence, ** >95% level of confidence, *** >99% level of confidence

### *6.2 Visitation and Welfare Estimation*

The fundamental purpose of the ORVal recreation demand model is to provide predictions of the visits and welfare value generated by access to greenspace. More specifically, to estimate how many visits and how much welfare might be generated by the creation of a new greenspace and how those same measures might change for an existing site as a consequence of changing its landcover composition.

The procedure we use for estimating the number of visits to a particular site (whether that be an existing greenspace or a proposed new greenspace) proceeds through a series of steps as follows;

1. LSOA Sociodemographics: For LSOA $r$, in a particular GOR, we use 2011 census data to calculate the average values for age, gender, working status, children and dog-ownership within that LSOA.
2. Travel Cost: Calculate the straight line distance between the LSOA centroid and the centroid of the site, indexed, $j$. For speed of calculation, rather than using routing software in a GIS to calculate travel time and distance in real time, we took a large sample of data derived from GIS routing calculations to estimate a polynomial equation relating straight line distances to driving distances and driving times as follows;

$$drive\ dist_{rj} = 2\ \times \left(1.3642 \times d_{rj} - 0.5890 \times \frac{d_{rj}^2}{10^3} + 0.8959 \times \frac{d_{rj}^3}{10^6} + 0.3772 \times \frac{d_{rj}^4}{10^9}\right)$$

$$drive\ time_{rj} = \frac{2}{60} \times \left(1.7120 \times d_{rj} - 4.2609 \times \frac{d_{rj}^2}{10^3} + 10.9042 \times \frac{d_{rj}^3}{10^6} - 9.4810 \times \frac{d_{rj}^4}{10^9}\right)$$

Using this driving distance and time data we calculated travel costs using the procedure described in Section 4.5.

3. Probability Calculation: Assuming a participation constant equal to that seen in the 2015 data and feeding in details of the park's characteristics and landcovers and using the travel cost information from 2 we use equation (14) to calculate the probability that an individual from socioeconomic segment AB will take a trip to the site on a Monday in January.
4. Individual Trips per Year: We multiply the probability from 3 up by the number of Mondays in an average January and then repeat that procedure for each day of the week and then again for each other month and day of the week combination. Adding up those probabilities provides an estimate of the expected number of trips to the site by an individual in the AB socioeconomic segment living in LSOA $r$.
5. LSOA Trips per Year: Using the 2011 census data, multiply up the trips from 4 by the number of adults in the AB segment in LSOA $r$. Then repeat steps 3, 4 and 5 for the C1, C2 and DE segments to calculate the total number of expected visits by adults from LSOA $r$ to the site.
6. National Trips per Year: Repeat steps 1 through 5 for each of the 34,753 LSOAs in England.

The procedure for estimating welfare changes follows identical logic but in step 4, rather than calculating a probability of visitation, we calculate welfare change using equation (17).

## References


Ben-Akiva, M. and M. Bierlaire (1999). Discrete choice methods and their applications to short term travel decisions. Handbook of transportation science, Springer: 5-33.

Bierlaire, M. (2006). "A theoretical analysis of the cross-nested logit model." Annals of Operations Research **144**(1): 287-300.

Cascetta, E., A. Nuzzolo, F. Russo and A. Vitetta (1996). A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Pergamon Lyon, France.

Department for Transport (2015). Provision of Market Research for Value of Travel Time Savings and Reliability, Phase 2 Report, Department for Transport.

Gao, F. and L. Han (2012). "Implementing the Nelder-Mead simplex algorithm with adaptive parameters." Computational Optimization and Applications **51**(1): 259-277.

Guevara, C. A. and M. E. Ben-Akiva (2013). "Sampling of alternatives in Multivariate Extreme Value (MEV) models." Transportation Research Part B: Methodological **48**: 31-52.

Kling, C. L. and J. A. Herriges (1995). "An Empirical Investigation of the Consistency of Nested Logit Models with Utility Maximization." American Journal of Agricultural Economics **77**(4): 875-884.

Manski, C. F. and S. R. Lerman (1977). "The Estimation of Choice Probabilities from Choice Based Samples." Econometrica **45**(8): 1977-1988.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. Frontiers of Econometrics. P. Zarembka. New York, Academic Press.

McFadden, D. (1978). Modelling the choice of residential location. Spatial Interaction Theory and Planning Models. A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull. Amsterdam, Amsterdam: 75-96.

McFadden, D., W. B. Tye and K. Train (1977). An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model, Institute of Transportation Studies, University of California.

Morey, E. R. (1999). Two RUMs unCLOAKED: Nested-Logit Models of Site Choice and NestedLogit Models of Participation and Site Choice. In Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice, Citeseer.

Natural England (2015). Monitor of Engagement with the Natural Environment: Technical Report from year 6 of the survey. **Natural England Joint Report JP015**.

Nelder, J. A. and R. Mead (1965). "A simplex method for function minimization." Computer Journal **7**(4): 308-313.

Small, K. A. and H. S. Rosen (1981). "Applied Welfare Economics with Discrete Choice Models." Econometrica **49**(1): 105-130.

Smith, V. K. and Y. Kaoru (1990). "What have we learned since hotelling's letter?" Economics Letters **32**(3): 267-272.

Thiene, M., J. Swait and R. Scarpa (2017). "Choice set formation for outdoor destinations: The role of motivations and preference discrimination in site selection for the management of public expenditures on protected areas." Journal of Environmental Economics and Management **81**: 152-173.

**APPENDIX 1: Destination Matching Algorithm**

For each observation, $i$, the scoring procedure progressed through the following steps.

- Potential Sites: To identify greenspaces that were potential matches to the visit destination, all parks, beaches and path networks with a within 2.5km of the location recorded as the MENE destination location were selected from the ORVal Greenspace Map. For beaches and parks the proximity of the potential site was recorded as the straight line distance from the centroid of that site to the MENE destination location. For path networks the proximity was taken as the straight line distance to the nearest location on a path network.
- Location Score: A proximity index was calculated for each site in the list of potential matches (indexed by $s$) using the following formula:

$$1 - \frac{Proximity_s \text{ (in m)}}{2{,}500} \tag{23}$$

which ascribes an index of 1 to sites exactly on the recorded MENE destination location and declines linearly with distance to 0 for the most distant potential match sites 2.5km from the recorded destination location.
For path networks, sites are defined by access points such that a second round of logic was required. First, we identified all access points to each site network in the list of potential matches. We then ranked those according to how far the access point was from the point we had previously identified as the nearest point on that network to the recorded MENE destination location with rank 0 being the closest, 1 the second closest, 2 the third closest, and so on. Under the assumption that it was more likely that we calculated the proximity index for access point $p$ on path network $s$ as follows;

$$\left(1 - \frac{Proximity_s \text{ (in m)}}{2{,}500}\right) \times 0.95^{rank_{s,p}} \tag{24}$$

Such that the highest ranked path access point on the network was given the highest proximity index and that index declined geometrically with increasing rank.
A *location score* was calculated first by multiplying the proximity index by a positive weighting factor. As with the other weighting factors to be described subsequently this weighting factor was adjusted in a process of calibration that ultimately set its value to 50.
Finally the *location score* for each site was adjusted to reflect information provided by respondents in the MENE questionnaire on the distance they had travelled to get to the site.
That information was provided as a range such that if the distance between the respondent's home and a possible match site was less than half the distance of the low end of that range then the proximity score was adjusted by a factor given by;

$$\frac{Distance\ to\ Home}{0.5\ \times Lower\ End\ of\ Distance\ Range} \tag{25}$$

Likewise, if the distance from a respondent's home was greater than 1.5 times the high end of the reported distance range then the proximity score was adjusted by a factor given by;

$$\frac{1.5 \times Lower\ End\ of\ Distance\ Range}{Distance\ to\ Home} \tag{26}$$

Clearly both adjustment factors lie between 0 and 1 ensuring that possible match sites located at a distance from a respondent's home considerably different from the distance they reported in the MENE questionnaire end up with a lower overall proximity score. The final *location score* varied on the range between 0 and 50.

- Environs Score: Questions 2 and 5 of the MENE survey provide information that helps identify the environs of the visited site particularly whether it was in a built-up or rural location, and whether that on was coastal or inland. To calculate an *environs score* for each possible match site, we began by defining a built-up indicator variable, *built-up%*, which established the proportion of a park's boundaries or a path's length that was within 100m of a built-up area. Where a respondent answered that they had visited a location in a town, city or seaside resort then we began the calculation of an *environs score* for each possible match site by multiplying a weighting factor (calibrated value: 10) by *built-up%*. Alternatively, if they indicated they had visited a location in the countryside a possible match sites location score was calculated as the weighting factor multiplied by 1 − *built-up%*. A similar calculation was carried out for coastal locations where coastal proximity was turned into a linearly declining index equal to 1 at the coast and falling to a value of 0 5km inland. Again if the visit destination was recorded as coastal then a weighting factor (calibrated value: 10) was multiplied by the coastal proximity index otherwise it was multiplied by one minus that amount. The *environs score* was incremented by that value reflecting the degree to which the coastal environs of the visited site matched that of possible match site.
  Answers to Question 5 of the MENE survey gave further clues as to the environs of the chosen site; for example, a respondent indicating that they had visited "Farmland" was assumed to have visited a site in a rural setting, while those indicating they had visited "A park in a town or city" had clearly chosen a site in a built-up setting. Such confirmatory information was given a weighting factor (calibrated value: 5) and added to the total *environs score*, which as a result could take a maximum value of 25.
- Type Score: Answers to Question 4 and 5 of the MENE survey allowed us to compare the type of recreation site visited by a respondent to the types of the possible match sites. Some explicit responses were given very high weighting factors; for example, if a respondent stated they had visited "an allotment", then all allotments in the list of possible sites were given a *type score* of 50 while all sites that were not allotments were given a *type score* of -10. Where the details of the Question 5 response were less explicit a lower type score was attributed; for example, if a respondent stated they had visited "a playing field or other recreation area" then a *type score* of 8 was given for all possible visit sites classified as 'parks' and a *type score* of 0 to all possible visit sites with a different classification.
- Landcover Score: Similar to the *type score* the landcover score used evidence from Questions 4 and 5 of the MENE survey to establish how closely the sorts of landcovers

present at the possible visit sites matched those present at the site actually visited. As an example, respondents indicating they had participated in fishing, swimming outdoors or watersports must have visited a site bordering water features including rivers, lakes, estuaries and sea. Accordingly, sites with such features amongst the list of possible match sites were given an increased *landcover score*. Similarly, where a respondent indicated they had visited a woodland or forest, then possible match sites with woodland cover were attributed landcover score.

- Total Match Score: To arrive at an overall match score for each possible site, the location score, environs score, type score and landcover score were summed. The site with the highest match score was chosen as the most likely location of that particular focus visit.

The full matching algorithm is transcribed below:

```
CREATE TABLE MENE.NearSites
 (spid   bigint,
  type   varchar,
  supertype varchar,
  prox   float,
  areagrid float,
  urbanpct float,
  coastprox float,
  lc_woods float,
  lc_agrculture float,
  lc_moors_heath float,
  lc_mountain float,
  lc_coastal float,
  lc_wood_pasture float,
  lc_sports_pitches float,
  lc_golf float,
  lc_allotments float,
  lc_seaside float,
  lc_estuary float,
  lc_rivers_canals float,
  lc_lakes_reservoirs float,
  dg_ancient_woodland float,
  dg_sssi float,
  dg_CPark float,
  dg_natura2000 float,
  dg_nnr float,
  dg_lnr float,
  dg_ramsar float,
  poi_playground float,
  disthome float,
  near_score integer,
  loc_score integer,
  type_score integer,
  lc_score integer,
  score  integer);

DO
$$
<<DestinationMatching>>
DECLARE
 destination record;
 visitdata record;
 matches        record;
 proximity integer := 2500;
 counter        integer := 0;
 nprint  integer := 50;
 pcounter      integer := 0;
 numrows        integer;
 numsites      integer;
 matchcnt      integer;
 coastdist     integer := 5000;
 coastthrshld integer := 1000;
 ntoprocess float;
 nprocessed float;
 TEST   boolean := FALSE;
```

```
BEGIN

 TEST = FALSE;

 ntoprocess := (SELECT count(*) FROM mene.visit_match);

 FOR destination IN TABLE mene.visit_match LOOP

   IF destination.geom_dest IS NULL OR destination.geom_home IS NULL THEN
    -- No destination or home location data - MATCHCODE -2
    IF NOT TEST THEN
       EXECUTE 'UPDATE mene.visit_match SET matchcode = -2 WHERE visitid =
'||destination.visitid;
    END IF;
   ELSIF lower(destination.q9) != 'your home' THEN
    -- Not travelled from home - MATCHCODE -3
    IF NOT TEST THEN
       EXECUTE 'UPDATE mene.visit_match SET matchcode = -3 WHERE visitid =
'||destination.visitid;
    END IF;
   ELSIF (lower(destination.q4_06) = 'yes') OR (lower(destination.q4_09) = 'yes') OR
(lower(destination.q4_11) = 'yes') THEN -- Not a location based interaction with greenspace
     -- q4_6: Off-road driving or motorcycling
     -- q4_9: Road cycling
     -- q4_11: Appreciating scenery from your car (e.g. at a viewpoint)
     -- MATCHCODE -4
    IF NOT TEST THEN
       EXECUTE 'UPDATE mene.visit_match SET matchcode = -4 WHERE visitid =
'||destination.visitid;
    END IF;
   ELSIF (lower(destination.q5_05) = 'yes') AND
     (lower(destination.q5_01) = 'no') AND (lower(destination.q5_02) = 'no') AND
(lower(destination.q5_03) = 'no') AND (lower(destination.q5_04) = 'no') AND
     (lower(destination.q5_06) = 'no') AND (lower(destination.q5_07) = 'no') AND
(lower(destination.q5_08) = 'no') AND (lower(destination.q5_09) = 'no') AND
     (lower(destination.q5_10) = 'no') AND (lower(destination.q5_11) = 'no') AND
(lower(destination.q5_12) = 'no') AND (lower(destination.q5_13) = 'no') AND
     (lower(destination.q5_14) = 'no') AND (lower(destination.q5_15) = 'no') AND
     (lower(destination.q4_02) = 'no') AND (lower(destination.q4_03) = 'no') AND
(lower(destination.q4_04) = 'no') AND (lower(destination.q4_05) = 'no') AND
     (lower(destination.q4_06) = 'no') AND (lower(destination.q4_07) = 'no') AND
(lower(destination.q4_08) = 'no') AND (lower(destination.q4_10) = 'no') AND
     (lower(destination.q4_12) = 'no') AND (lower(destination.q4_13) = 'no') AND
(lower(destination.q4_15) = 'no') AND (lower(destination.q4_16) = 'no') AND
     (lower(destination.q4_17) = 'no') AND (lower(destination.q4_18) = 'no') AND
(lower(destination.q4_19) = 'no') THEN
    -- Then just visited a village - MATCHCODE -5
    IF NOT TEST THEN
       EXECUTE 'UPDATE mene.visit_match SET matchcode = -5 WHERE visitid =
'||destination.visitid;
    END IF;
   ELSE

    -- (1) FIND NEAR SITES
    -- -------------------
    -- Select all sites ST_Dwithin 1km of destination coordinates
    -- No need to worry about entrances
    TRUNCATE TABLE MENE.NearSites;
    numsites := 0;

    -- Parks:
    INSERT INTO MENE.NearSites (spid, type, supertype, prox, disthome, urbanpct, coastprox,
areagrid,
                lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal,
lc_wood_pasture, lc_sports_pitches, lc_golf,
                lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
                dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr,
dg_ramsar, poi_playground)
    SELECT spid, type, supertype,
         (1-(ST_Distance(geom, destination.geom_dest)/proximity)) AS prox,
         ST_Distance(geom, destination.geom_home) AS disthome,
         urbanpct, coastprox, areagrid,
      lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal, lc_wood_pasture,
lc_sports_pitches, lc_golf,
      lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
```

```
     dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr, dg_ramsar,
poi_playground
    FROM parks.parks_england
    WHERE ST_Dwithin(geom, destination.geom_dest, proximity) AND supertype IS NOT NULL;
    GET DIAGNOSTICS numrows = ROW_COUNT;
    numsites = numsites + COALESCE(numrows,0);

    -- Paths:
    INSERT INTO MENE.NearSites (spid, type, supertype, prox, disthome, urbanpct, coastprox,
areagrid,
                lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal,
lc_wood_pasture, lc_sports_pitches, lc_golf,
                lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
                dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr,
dg_ramsar, poi_playground)
    SELECT spid, type, supertype, (prox*0.95^(rank-1)) AS prox, disthome, urbanpct, coastprox,
areagrid,
     lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal, lc_wood_pasture,
lc_sports_pitches, lc_golf,
     lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
     dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr, dg_ramsar,
poi_playground
    -- ranks access points
    FROM (SELECT spid, type, supertype, prox, disthome, urbanpct, coastprox, rank() OVER
(PARTITION BY pid ORDER BY dist_acc ASC) AS rank, areagrid,
       lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal, lc_wood_pasture,
lc_sports_pitches, lc_golf,
       lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
       dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr, dg_ramsar,
poi_playground
    -- Find access points to close paths
     FROM (SELECT tbl1.pid, tbl1.spid, tbl1.type, tbl1.supertype,
                       (1-ST_Distance(tbl2.geom, destination.geom_dest)/proximity) AS prox,
                                    ST_Distance(tbl2.geom, destination.geom_home) AS
disthome,
                       tbl1.urbanpct, tbl1.coastprox, tbl1.areagrid,
        ST_Distance(tbl1.geom, destination.geom_dest) AS dist_acc,
        lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal, lc_wood_pasture,
lc_sports_pitches, lc_golf,
        lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
        dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr, dg_ramsar,
poi_playground
     FROM paths.paths_england AS tbl1 INNER JOIN
        -- Select paths that pass close to destination
        (SELECT DISTINCT pid, ST_ClosestPoint(geom_line, destination.geom_dest) AS geom
       FROM paths.paths
       WHERE ST_Dwithin(geom_line, destination.geom_dest, proximity)) AS tbl2
     ON tbl1.pid = tbl2.pid AND ST_DWithin(tbl1.geom, tbl2.geom, 10000)) AS tbl3) AS tbl4;
   -- Find nearest point on paths, then select all access points paths that are within 10km
   -- Rank access points on same pid and reweight prox score according to how close access
point is to destination
    GET DIAGNOSTICS numrows = ROW_COUNT;
    numsites = numsites + COALESCE(numrows,0);

    -- Beaches:
    INSERT INTO MENE.NearSites (spid, type, supertype, prox, disthome, urbanpct, coastprox,
areagrid,
                lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal,
lc_wood_pasture, lc_sports_pitches, lc_golf,
                lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
                dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr,
dg_ramsar, poi_playground)
    SELECT spid, type, supertype,
        (1-(ST_Distance(geom, destination.geom_dest)/proximity)) AS prox,
                ST_Distance(geom, destination.geom_home) AS disthome,
        urbanpct, coastprox, areagrid,
     lc_woods, lc_agrculture, lc_moors_heath, lc_mountain, lc_coastal, lc_wood_pasture,
lc_sports_pitches, lc_golf,
     lc_allotments, lc_seaside, lc_estuary, lc_rivers_canals, lc_lakes_reservoirs,
     dg_ancient_woodland, dg_sssi, dg_CPark, dg_natura2000, dg_nnr, dg_lnr, dg_ramsar,
poi_playground
    FROM beaches.beaches_england
    WHERE ST_Dwithin(geom, destination.geom_dest, proximity);
    GET DIAGNOSTICS numrows = ROW_COUNT;
    numsites = numsites + COALESCE(numrows,0);
    -- RAISE NOTICE ' Number sites near %: %', destination.visitid, numsites;
```

```
    IF numsites = 0 THEN
        -- MATCHCODE -1
        IF NOT TEST THEN
                       EXECUTE 'UPDATE mene.visit_match SET matchcode = -1 WHERE visitid =
'||destination.visitid;
               END IF;
    ELSE
         IF NOT TEST THEN
                EXECUTE 'UPDATE mene.visit_match SET matchcode = 0 WHERE visitid =
'||destination.visitid;
         END IF;

    -- (2) LOCATION SCORE
    -- -------------------
    -- 50 pts for proximity
    -- Linear Distance Decay: 50 * (1 - ST_Distance/proximity)
    UPDATE MENE.NearSites SET near_score = (50 * prox);

         -- Check for compatibility with stated travel distance
    UPDATE MENE.NearSites
    SET near_score = near_score * disthome/(destination.travdistlo*.5)
    WHERE disthome < destination.travdistlo*.5;

         UPDATE MENE.NearSites
    SET near_score = near_score * (destination.travdisthi*1.5)/disthome
    WHERE disthome > destination.travdisthi*1.5;

    -- (3) ENVIRONS SCORE
    -- ------------------
    -- urbanpct, rural, coastal, inland
    CASE destination.q2
      WHEN 'In a town or city' THEN
      UPDATE MENE.NearSites SET loc_score = (10*urbanpct     + 10*greatest(0,(coastprox-
coastthrshld)/(coastdist-coastthrshld)));
      WHEN 'In the countryside (including areas around towns and cities)' THEN
      UPDATE MENE.NearSites SET loc_score = (10*(1-urbanpct) + 10*greatest(0,(coastprox-
coastthrshld)/(coastdist-coastthrshld)));
      WHEN 'In a seaside resort or town' THEN
      UPDATE MENE.NearSites SET loc_score = (10*urbanpct     + 10*(1-greatest(0,(coastprox-
coastthrshld)/(coastdist-coastthrshld))));
      WHEN 'Other seaside coastline (including beaches and cliffs)' THEN
      UPDATE MENE.NearSites SET loc_score = (10*(1-urbanpct) + 10*(1-greatest(0,(coastprox-
coastthrshld)/(coastdist-coastthrshld))));
      ELSE
    END CASE;

    CASE destination.q5_02 -- Farmland
      WHEN 'Yes' THEN -- rural
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*(1-urbanpct));
      ELSE
    END CASE;

    CASE destination.q5_03 -- Mountain, Wood or Moorland
      WHEN 'Yes' THEN -- rural
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*(1-urbanpct));
      ELSE
    END CASE;

    CASE destination.q5_08 -- Another open space in the countryside
      WHEN 'Yes' THEN -- rural
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*(1-urbanpct));
      ELSE
    END CASE;

    CASE destination.q5_09 -- A park in a town or city
      WHEN 'Yes' THEN -- urban
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*urbanpct);
      ELSE
    END CASE;

    CASE destination.q5_13 -- Another open space in a town or city
      WHEN 'Yes' THEN -- urban
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*urbanpct);
      ELSE
    END CASE;

    CASE destination.q5_14 -- A beach
```

```
      WHEN 'Yes' THEN -- coastal
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*coastprox/coastdist);
      ELSE
    END CASE;

    CASE destination.q5_15 -- Other coastline
      WHEN 'Yes' THEN -- coastal
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*coastprox/coastdist);
      ELSE
    END CASE;

    CASE destination.q4_13 -- Visits to a beach (sunbathing or paddling in the sea)
      WHEN 'Yes' THEN -- coastal
      UPDATE MENE.NearSites SET loc_score = loc_score + (5*coastprox/coastdist);
      ELSE
    END CASE;


    -- (4) TYPE SCORE
    -- --------------
    -- path    park    beach   country park   allotment golf
    UPDATE MENE.NearSites SET type_score = 0;

    -- Allotment
    -- ---------
    CASE destination.q5_10 -- An allotment or community garden
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 50 WHERE supertype = 'allotment';
      WHEN 'No' THEN
      UPDATE MENE.NearSites SET type_score = type_score - 10 WHERE supertype = 'allotment';
      ELSE
    END CASE;

    -- Country Park
    -- ------------
    CASE destination.q5_07 -- Country parks
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 25 WHERE supertype = 'country_park';
      WHEN 'No' THEN
      UPDATE MENE.NearSites SET type_score = type_score - 5 WHERE supertype = 'country_park';
      ELSE
    END CASE;

    -- Cemetery
    -- --------
    CASE destination.q5_08 -- Another open place in the countryside
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 3 WHERE supertype = 'cemetery';
      ELSE
    END CASE;
    CASE destination.q5_13 -- Another open place in a town or city
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 3 WHERE supertype = 'cemetery';
      ELSE
    END CASE;

    -- Park
    -- ----
    CASE destination.q5_09 -- A park in a town or city
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 10 WHERE supertype = 'park';
--              WHEN 'No' THEN
--              UPDATE MENE.NearSites SET type_score = type_score - 2 WHERE supertype =
'park';
      ELSE
    END CASE;
    CASE destination.q5_12 -- A playing field or other recreation area
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 8 WHERE supertype = 'park';
--              WHEN 'No' THEN
--              UPDATE MENE.NearSites SET type_score = type_score - 2 WHERE supertype =
'park';
      ELSE
    END CASE;

    -- Beach
    -- -----
```

```
    CASE destination.q5_14 -- A beach
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 25 WHERE type = 'beach';
      WHEN 'No' THEN
      UPDATE MENE.NearSites SET type_score = type_score - 5 WHERE type = 'beach';
      ELSE
    END CASE;
    CASE destination.q4_13 -- Visits to a beach (sunbathing or paddling in the sea)
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 25 WHERE type = 'beach';
      WHEN 'No' THEN
      UPDATE MENE.NearSites SET type_score = type_score - 5 WHERE type = 'beach';
      ELSE
    END CASE;

    -- Golf
    -- -----
    CASE destination.q4_19 -- Informal games and sport
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 10 WHERE supertype = 'golf';
      WHEN 'No' THEN
      UPDATE MENE.NearSites SET type_score = type_score - 10 WHERE supertype = 'golf';
      ELSE
    END CASE;

    -- Path
    -- ----
    CASE destination.q5_06 -- A path
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 10 WHERE type = 'path';
      ELSE
    END CASE;

    -- Nature
    -- ------
    CASE destination.q4_18 -- Wildlife Watching
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET type_score = type_score + 10 WHERE type = 'nature';
      ELSE
    END CASE;


    -- (5) LANDCOVER SCORE
    -- -------------------
    -- woods, water, farmland, playground, sports, nature, leisure, country park
    UPDATE MENE.NearSites SET lc_score = 0;

    -- Woods
    -- -----
    CASE destination.q5_01 -- A woodland or forest
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + 5 WHERE type = 'woods';
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_woods/areagrid) +
(5*lc_wood_pasture/areagrid) + (5*dg_ancient_woodland/areagrid);
      ELSE
    END CASE;

    -- Farmland
    -- --------
    CASE destination.q5_02 -- Farmland
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (15*lc_agrculture/areagrid);
      ELSE
    END CASE;

    -- Mountain, Hill, Moorland
    -- ------------------------
    CASE destination.q5_03 -- Mountain, Hill, Moorland
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_moors_heath/areagrid) +
(10*lc_mountain/areagrid);
      ELSE
    END CASE;

    -- River, Canal, Lake or Reservoir
    -- -------------------------------
    CASE destination.q5_04 -- River, Lake or Canal
```

```
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_rivers_canals/areagrid) +
(10*lc_lakes_reservoirs/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_03 -- Fishing
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_rivers_canals/areagrid) +
(10*lc_lakes_reservoirs/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_12 -- Swimming Outdoors
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (5*lc_rivers_canals/areagrid) +
(5*lc_lakes_reservoirs/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_17 -- Watersports
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (5*lc_rivers_canals/areagrid) +
(10*lc_lakes_reservoirs/areagrid);
      ELSE
    END CASE;


    -- Seaside or Esturary
    -- ------------------
    CASE destination.q4_03 -- Fishing
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_seaside/areagrid) +
(10*lc_estuary/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_12 -- Swimming Outdoors
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_seaside/areagrid) +
(10*lc_estuary/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_17 -- Watersports
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_seaside/areagrid) +
(10*lc_estuary/areagrid);
      ELSE
    END CASE;

    -- Sports Pitches
    -- --------------
    CASE destination.q5_12 -- Playing Fields or Other Recreation Area
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_sports_pitches/areagrid);
      ELSE
    END CASE;
    CASE destination.q4_19 -- Informal games and sport
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*lc_sports_pitches/areagrid)+
(10*lc_golf/areagrid);
      ELSE
    END CASE;

    -- Nature
    -- ------
    CASE destination.q4_18 -- Wildlife watching
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (5*greatest(dg_nnr, dg_lnr,
dg_natura2000, dg_sssi, dg_ramsar)/areagrid);
      ELSE
    END CASE;

    -- Playgrounds
    -- -----------
    CASE destination.q5_11 -- A children's playground
      WHEN 'Yes' THEN
      UPDATE MENE.NearSites SET lc_score = lc_score + (10*poi_playground);
      ELSE
    END CASE;
```

```
    -- (6) RECORD BEST 3 MATCHES
    -- -------------------------
    UPDATE MENE.NearSites SET score = near_score + COALESCE(loc_score,0) +
COALESCE(type_score,0) + COALESCE(lc_score,0);
    IF NOT TEST THEN
        matchcnt := 1;
      FOR matches IN SELECT spid, near_score, loc_score, score FROM MENE.NearSites ORDER BY
score DESC LIMIT 3 LOOP
        EXIT WHEN NOT FOUND;
        EXECUTE 'UPDATE mene.visit_match SET match'||matchcnt||'id = '||matches.spid||',
match'||matchcnt||'score = '||matches.score||' WHERE visitid = '||destination.visitid;
        matchcnt := matchcnt + 1;
      END LOOP;
    END IF;
   END IF;
  END IF;

   IF pcounter = nprint THEN
    nprocessed := (SELECT count(*) FROM mene.visit_match WHERE matchcode IS NOT NULL);
    RAISE NOTICE '     Visits processed: % (% of % = % pct done)', counter, nprocessed,
ntoprocess, (nprocessed/ntoprocess);
    pcounter := 0;
   END IF;

   pcounter := pcounter + 1;
   counter := counter + 1;

 END LOOP;

 IF NOT TEST THEN
   DROP TABLE IF EXISTS MENE.NearSites;
 END IF;

END;
$$ LANGUAGE plpgsql;
```