
Chapter 6

Weighting

In order to provide accurate and unbiased national estimates, the English Housing Survey (EHS) is weighted to take account of the over-sampling of the less prevalent tenure groups and differential non-response. This chapter provides details of the weighting methodology used and gives advice on which weights to use when conducting analysis using EHS data.

Overview

6.1 The following weights have been calculated for the 2012-13 EHS data:

- aagfh12: Household weight for full sample interview survey 2012-13
- aagpd1112: Average dwelling weight for cases in 2011-12 and 2012-13 that had both the interview and physical survey conducted (paired cases)
- aagph1112: Average household weight for cases in 2011-12 and 2012-13 that had both the interview and physical survey conducted (paired cases)

6.2 These weights adjust the sample to take account of the over-sampling of the less prevalent tenure groups and reduce the bias from differential non-response.

6.3 The weighting of the 2012-13 EHS data was undertaken by NatCen who managed the survey on behalf of the Department for Communities and Local Government (DCLG).

Summary of weighting methodology

6.4 The weighting strategy uses a sequence of fifteen stages described below. Some stages correct for the disproportionate selection when sampling; others for differential non-response in the interview/physical survey response process.

Stages 1 and 2 apply to both the full household sample and the dwelling sample and adjust for:

- Stage 1: Address selection probabilities
- Stage 2: Address to dwelling relationship

Stages 3 to 6 apply to the full household sample and adjust for:

- Stage 3: Dwelling to household relationship
- Stage 4a: Office refusals - refusal to cooperate prior to the interview
- Stage 4b: Non-contact at the interview survey
- Stage 5: Non-cooperation at the interview survey
- Stage 6: Calibration of weighted estimates from the interview survey to known population totals

Stages 7a to 15 apply to the dwelling sample and adjust for:

- Stage 7a: Office refusals – refusal to cooperate prior to contact at the dwelling
- Stage 7b: Non-contact at the dwelling
- Stage 8: Non-cooperation at the dwelling
- Stage 9: Sub-sampling by tenure for the physical survey
- Stage 10: Non-cooperation at the physical survey
- Stage 11: Calibration of weighted estimates to dwelling controls totals, including an adjustment for new builds
- Stage 12: Creation of household weight by calibrating the weighted estimates to the same population totals as the full household sample
- Stage 13: Creation of a household weight using the stage 11 weights and the household to dwelling relationship
- Stage 14: Final household weight by scaling stage 13 weight to produce the same weighted totals as the stage 12 calibrated weight
- Stage 15: Final dwelling weight which is consistent with the final household weight

Stages 1 to 5, 7a to 10 and 13 to 15 are implemented using a set of SPSS syntax scripts in combination with the specialist SPSS module AnswerTree.

Stages 6, 11 and 12 use the calibration command in Stata.

6.5 These processes are described in more detail below.

Adjusting for the relationship between addresses and dwellings

- 6.6 Stage 1 generates the selection weights for the issued sample of addresses. This has two components: the first is a weight for the address selected for the office sift to the number of PAF; the second corrects for the unequal selection of addresses within each predicted tenure type (within each quarter) from those addresses selected from the PAF for the office sift.
- 6.7 Subsequent stages involve multiplying these initial weights by successive adjustment factors.
- 6.8 The EHS analyses are concerned with dwellings and households rather than addresses, and there is not always a one-to-one relationship between an address, a dwelling, and a household. (For the purposes of the survey, a dwelling is defined as 'a self-contained unit of accommodation where the occupants of that accommodation have sole use of all the rooms and facilities').
- 6.9 Usually there is only one dwelling at each address sampled from the PAF, but addresses are occasionally found to cover more than one dwelling (for example if a house has been converted into self-contained flats) or only part of a dwelling (for example a bedsit which shares facilities with a household at a separate postal address). As only one dwelling is selected at these addresses, weights are required to account of this.
- 6.10 Where an address refers to more/less than one dwelling, each dwelling at that address would have a lower/higher chance of selection for the EHS. Stage 2 adjusts the weight calculated at Stage 1 to take account of the address to dwelling relationship for such cases. Weighting classes are generated defined by region, tenure and type (house, low rise flat, high rise flat). Within each of these weighting classes, the average number of dwellings per address is calculated to generate the address to dwelling weight.

Weighting the full household sample

- 6.11 Stage 3 makes an adjustment for the selection of a single household at dwellings that contain more than one household. This weight is calculated using the same approach as the address to dwelling weight (see 6.8), as the average number of households per dwelling within weighting classes defined by region, tenure and type (house, low rise flat, high rise flat).
- 6.12 Refusal to co-operate prior to interview, non-contact at interview and refusal to co-operate at interview do not happen completely at random and the factors associated with each of these three processes may differ. Stages 4a,

4, and 5 each use available information about each case as predictor variables in a model to partition the sample into groups in order to describe as much variation in the response variable as possible. These models are constructed using the CHAID¹ algorithm in the SPSS AnswerTree software. Typical predictor variables for Stage 4a are geographical area; predominant tenure, dwelling age and dwelling type in the area; urban/rural classification. For subsequent stages, information collected by the interviewer is also used.

- 6.13 Response weights are calculated for each of the groups produced by AnswerTree at each of Stage 4a, 4 and 5.
- 6.14 The previous five stages account for the sampling and response probabilities. Applying the weight from Stage 5 to the household-level data would provide a survey estimate of the total number of households in England. However, this will differ from the true value because of sampling error, under-coverage of the frame and inadequacies in capturing the non-response mechanisms. Estimates for subgroups such as tenures will differ from their true values for the same reason. These differences in the survey estimates can be reduced by adjusting (calibrating) the weights so that the total final weights match external control totals.
- 6.15 The control totals used for Stage 6 were based on ONS population projections by sex and age group by geographical area, and tenure (owner occupied, social sector and private rented sector) from the Labour Force Survey as at 1 October of the survey year.

Weighting the dwelling sample

- 6.16 The approach to weighting the paired cases which make up the dwelling sample is dwelling-based rather than household-based so, in contrast to the weighting of the full household sample described above, all dwellings identified as vacant at which a physical survey was achieved are included in the sample to be weighted.
- 6.17 This process is more complex than that for the full household sample, partly because of the need to create internally consistent household and dwelling weights, and also because of the need to combine two years' dwelling sample data to obtain a large enough sample for analysis.
- 6.18 In outline, the approach is to first weight up the dwellings data for the current year to estimated dwelling controls by tenure based on DCLG published data on dwelling stock, then to adjust this weighting so that the number of weighted households that result from it is consistent, within tenure and geographical area, with the weighted full household sample.

¹ Chi-squared Automatic Interaction Detector

-
- 6.19 The weighting process starts by using Stages 1 and 2, described above, to take account of the initial sampling fractions and the address to dwelling relationship.
- 6.20 Stage 7a then compensates for office refusals at the interview survey stage. This process uses AnswerTree as in Stage 4a but with the inclusion of dwellings identified as vacant.
- 6.21 Stages 7b and 8 adjust for interviewer contact at the dwelling and for co-operation with the interview survey. All vacant dwellings are deemed to have been 'contacted' provided they have been located by the interviewer, and to have 'cooperated' with the interview survey phase.
- 6.22 The dwelling sample is required to contain a disproportionate number of rented properties. This is achieved by sub-sampling the issued sample during the interview survey once information about tenure, including tenure of vacant properties, has been collected. Only dwellings selected at this stage are eligible for the physical survey. Sub-sampling rates for the physical survey can also be varied between quarters, as was done for 2011-12. Stage 9 calculates adjustments to the weights to take account of this sub-sampling.
- 6.23 Stage 10 adjusts for response to the physical survey, within weighting groups created by AnswerTree. For this stage, data collected during the interview survey are also used to help determine the weighting groups; the vacant cases are treated separately.
- 6.24 As with the interview survey weighting, the initial weighting stages of the physical survey (Stages 1, 2, and 7a - 10) attempt to account for sampling and response probabilities and so the total weight of the data gives a survey estimate of the total number of dwellings in the population. However, this will differ somewhat from the true value because of factors such as sampling error, under-coverage of the frame and inadequacies in capturing the non-response mechanisms. These differences in the survey estimate can be reduced by adjusting (calibrating) the weights to match chosen control totals.
- 6.25 The control totals used for stage 11 were based on DCLG dwelling estimates by tenure and geographical area, as at 1 October of the survey year. The weights of all cases with a construction date of 1990 onwards were then adjusted (re-calibrated to the same DCLG dwelling estimates) using the number of new dwellings built since the time of sampling, separately for areas with a high/low rate of new build, and private/social sector housing (excluding local authority housing).
- 6.26 Stage 12 creates a household weight for the paired sample, by calibrating the stage 10 weights to ONS population projections by sex and age group by geographical area, and tenure (owner occupied, social sector and private

rented sector) from the Labour Force Survey as at 1 October of the survey year (same as Stage 6).

- 6.27 Using the weights for occupied dwellings at stage 11, stage 13 derives a corresponding household weight using the household to dwelling relationship. As in stage 3, this household weight is smoothed within classes defined by region, tenure and type (house, low rise flat, high rise flat).
- 6.28 Stage 14 derives the final household weight by applying a scaling factor to the stage 13 household weight so that it produces the same weighted totals as the calibrated household weights from stage 12.
- 6.29 Stage 15 derives the final dwelling weight by applying the same (stage 14) scaling factor to the stage 11 dwelling weight for both occupied and vacant dwellings, ensuring that the final dwelling and household weights are consistent.

Calculating two year weights

- 6.30 Because of the smaller annual sample sizes involved, analysis of the dwelling sample is normally carried out using 2 years' weighted data. This section sets out how the combined weights were calculated.
- 6.31 Steps 1, 2 and 3 for the two year weights repeat stages 6, 12 and 11 respectively of the annual weighting separately for the 2011-12 and 2012-13 EHS datasets – they are individually re-calibrated to control totals as at 1 April 2012, the centre of the two survey periods.
- 6.32 At step 4, the individual year datasets with the re-calibrated household weights (from step 2) are merged together and a two-year household weight is created by dividing each year's weight by 2, so that each dataset has equal influence.
- 6.33 At step 5a, the individual year datasets with the re-calibrated dwelling weights (from step 3) are merged together and a dwelling weight is created by dividing each year's weight by 2, so that each dataset has equal influence.
- 6.34 At step 5b, the combined household weight (from step 4) is adjusted to take into account the new dwellings built since the time of sampling and then scaled to the original household totals.
- 6.35 Using the dwelling weights at step 5a, step 6 derives a household weight using the household to dwelling relationship of the combined dataset (by repeating the same process as stage 13 of the annual weighting).
- 6.36 Step 7 derives the final combined household weight by applying a scaling factor to the step 6 household weight so that it produces the same weighted

totals as the household weights from step5b (similar to stage 14 of the annual weighting).

- 6.37 Step 8 derives the final dwelling weight by applying the same (step 7) scaling factor to the dwelling weights at step 5a, for both occupied and vacant dwellings, ensuring that the final dwelling and household weights are consistent (similar to stage 15 of the annual weighting).

Application of weights during analysis

- 6.38 The EHS comes with its own weights covering the household and dwelling sample of cases. The weight for the household sample can be found in the file `general12.sav` and is called `agfh12`. This should be used for any analysis for which the aim is to provide estimates of households, based on the interview survey data.
- 6.39 The weights covering the dwelling sample of cases can be found in the file `general 11+12.sav`. `Aagpd1112` should be used for any analysis in which the aim is to provide estimates of dwellings and that includes physical survey data (e.g. percentage of non-decent dwellings) while `aagph1112` should be used for any analysis in which the aim is to provide estimates of households and that includes interview survey data (e.g. percentage of households in non-decent dwellings). The weights can only be used on full 2-year datasets. They cannot be used on the data split into separate years. The recommended application of weights is summarised in Table 6.1.

Table 6.1: Application of weights during analysis

Weight	Description	Base
<code>aagfh12</code>	Household weight for full sample interview survey 2012-13	13,652
<code>aagpd1112</code>	Average dwelling weight for cases in 2011-12 and 2012-13 that had both the interview and physical survey conducted (paired cases)	12,763
<code>aagph1112</code>	Average household weight for cases in 2011-12 and 2012-13 that had both the interview and physical survey conducted (paired cases)	12,269