\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* A Practical Guide to Using Panel Data

\* Simonetta Longhi and Alita Nandi

\* ISER, University of Essex

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Appendix A**

**British Household Panel Study**

This is a short introduction to the British Household Panel Survey (BHLS), which summarises the main characteristics of the study, also discussed in Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’. For further details see the user guide:

Taylor, Marcia Freed (ed). with John Brice, Nick Buck and Elaine Prentice-Lane (2010) British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex.

You can find the user guide, interactive online documentation, questionnaires, fieldwork and technical documents at <https://www.iser.essex.ac.uk/bhps/documentation/vola/vola.html>.

*The Sample*

The BHPS is a multi-purpose household panel survey which started in 1991 with a sample of approximately 5,000 households drawn from the non-institutionalised resident population of Great Britain (specifically, England, Wales and Scotland south of the Caledonian Canal) in 1990. This sample is often referred to as the original sample or the “Essex” sample. Two samples of around 1,500 households from Scotland and Wales were added in 1999 and a sample of around 2,000 household from Northern Ireland was added in 2001. The samples from Great Britain were drawn from the Postal Address Small Users File of domestic addresses and the Northern Ireland boost sample was drawn from the Valuation and Lands Agency list of domestic addresses. All the Great Britain samples have a clustered and stratified sample design while the Northern Ireland sample has a simple random sample design. For the Great Britain samples, in the first stage postcode sectors were selected and in the second stage approximately 33 addresses were selected from each selected postcode. At each selected address up to three dwelling units were randomly selected and at each selected dwelling unit up to three households were randomly selected. The selection probability was much higher in the additional regional boost samples as compared to the original “Essex” sample.

During the period 1997 to 2001, an additional sample was added to the BHPS to provide data to the UK component of the European Community Household Panel Survey (ECHP). This sample was interviewed in exactly the same way as the rest of the BHPS sample. The data are included in the BHPS files for the relevant waves and can be identified by using the sample origin variable.

*The Survey*

The BHPS started out as an indefinite life panel survey but ended in its current form in 2008 after 18 years of interviews. The surviving members of the BHPS sample became part of Understanding Society: the UK Household Longitudinal Study (UKHLS) from the second wave in 2010. In the BHPS, interviews were mainly conducted face-to-face during the months of September to December of each year. From the third wave onwards approximately 500 households were interviewed by telephone.

All household members of responding households in wave 1 and their descendants are classified as Original Sample Members (OSMs). OSMs are followed wherever they move as long as they reside in the UK (until 2001 the scope of the survey was restricted to Great Britain only). Anyone who moves into a household with an OSM (wave 2 onwards) is classified as Temporary Sample Members (TSMs). TSMs are only interviewed as long as they are co-resident with at least one OSM. Any TSM who becomes the parent of an OSM child becomes a Permanent Sample Member (PSM). The following rules for the PSMs are the same as those of the OSMs.

In the BHPS most information was collected prospectively. The exceptions were employment, marital and fertility histories prior to the start of the survey, which were collected retrospectively.

In the BHPS, data are collected using the following set of survey instruments.

* Household and enumeration grid: information about who lives in the household, their relationships to the household reference person (the person who owns or rents the property or, if more than one, the eldest among them),and some basic information about them such as marital status, age, sex.
* Household questionnaire: information about the residential property, household expenditure, durable goods, access to car and so on. This information is collected from one adult in the household who knows most about these issues (generally the household reference person).
* Adult face-to-face questionnaire: all adults (aged 16 years or above) are asked detailed factual and attitudinal questions about themselves (and sometimes about their children).
* Proxy interview questionnaire: basic factual information is collected about non-responding adults from their spouse, partner or adult child who knows them well. This questionnaire was also used to interview telephone respondents.
* Adult self-completion questionnaire: adult respondents are also asked to fill in a paper questionnaire which may include questions on sensitive topics.
* Youth self-completion questionnaire: from the fourth wave (1994) onwards, young persons in the household between the ages of 11 and 15 were asked to complete a paper self-completion questionnaire about issues particularly relevant to the lives of young persons.

*Data*

BHPS data are provided in Stata, SPSS and TAB formats. The BHPS collects information at different levels (household, adults, youths, individual income source, jobs and so on) and at different points in time. An efficient way to provide these data is as separate files for data from each source and level, and separate sets of files for each wave. Wave-specific files and variables in the BHPS follow a specific naming convention: they have the same root name with a wave-specific letter prefix (a, b, c and so on). For example, the file containing the data from the individual interview is called aindresp if it refers to the first wave, bindresp if it refers to the second wave, and so on. In a similar way, the variable for age is aage in the first wave, bage in the second wave and so on. The variable pid is the only one that does not to have a wave prefix; this is the cross-wave unique person identifier.

Within each wave households have a unique identification number called whid (w is a placeholder for the wave-specific letter prefix and goes from a to r). Household identifiers can be used to match household level information to individual level data. For further details on how to match household and individual level data see Chapter 5 of Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’. Each individual within a household and at each wave can also be identified by their person number, wpno. In each wave these two variables, whid and wpno, together can be used to uniquely identify an individual as well.

The BHPS provides the identifiers for parents, spouses and partners and the relationship between all household members can be identified in the BHPS using the wegoalt files (see Chapter 6 in Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’).

In the BHPS, different types of missing data can be identified by their values. The values -1, -2, -7, -8 and -9 represent don’t know, refusal, not asked in proxy or telephone interview, valid skips or not applicable, and inconsistent or implausible values, respectively. The variable wsampst shows the sample status (OSM, TSM, PSM) and the variables wmemorig and whhorig show the sample origin (Essex sample, Scottish, Welsh or Northern Irish boost samples, ECHP sample).

All information collected during the adult individual interview including self-completion, proxy and telephone interviews are provided in the windresp files. Household questionnaire data are stored in whhresp files, and youth questionnaire in the wyouth files. The information that interviewers collected at the doorstep (from the enumeration and household grid) are stored in the windall files; these are the only source of data for non-respondents and children.

The file xwavedat is an individual level file that contains fixed information about individuals (note the absence of the wave prefix in the file name). Time-invariant information is generally collected when respondents are interviewed for the first time (such as date of birth, sex, parents’ education and occupation when the respondent was 14 years old) or at some specific waves (such as whether lived with siblings as a child, which was asked in wave 13). These variables are available in the respective wave-specific files, but are also reported in the individual level file xwavedat. A list of files and their contents is provided in Table 1 at the end of this document.

*Identifying other household members*

The BHPS provides variables which show the cross-wave identifier and person number within the household of the spouse or partner and parents. Other family members can be identified using the wegoalt files. The method for doing this is described in Chapter 6 of Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’.

*Sample Weights*

The BHPS weights are composite weights which account for both the unequal selection probabilities (due to the regional over-samples) and non-response. Both cross-sectional and longitudinal weights are provided separately for each year. Different sets of weights are available to use for analysis using the original Essex sample, all four samples combined, or only regional samples separately. Separate household, enumerated and respondent cross-section weights are also available. Longitudinal weights are only available for OSMs who responded continuously since the beginning of the sample, and all TSMs have zero longitudinal weights. For guidance on using weights and their naming convention see section V of the user guide, Taylor et al (2010). The individual and household weights are provided in the individual and household respondent files, while strata and clustering variables are available in the file whhsamp. For further details on how, why and when to use weights and correctly estimate standard errors when the sample design is not a simple random sample see Chapter 7 of Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’.

*History Files*

The marriage, fertility, employment and job histories, prior to the start of the survey were collected in the second and third wave for the original sample and later on for the regional boost samples. Additionally, each year information on all labour market spells and jobs held since the last interviews are also collected. The history files are multi-level files where each row represents one spell for one individual and include information on the type of spell, start and end dates and whether the spell is still ongoing. For more details see Part III of Longhi and Nandi (2014) ‘A Practical Guide to Using Panel Data’.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1: Description of BHPS data files** | | | |
| **Contains** | **File Name** | **Responses provided by** | **Each row of observation is uniquely identified by** |
| *Substantive information* |  |  |  |
| Information on the household and enumeration grid. | windall | Any adult in the household | pid or  whid wpno |
| Responses from the household questionnaire | whhresp | Knowledgeable adult in the householda | whid |
| Responses from adult individual interviews (face-to-face, telephone, self-completion and proxy). Also includes interviewer remarks about the interview process. | windresp | Adults (16+ year olds) AND interviewers | pid or  whid wpno |
| Responses from adult (face-to-face) individual interviews. |  | Adults (16+ year olds) |  |
| Information of all income sources since last interview, one row for each income source of each individual | wincome |  | pid wfiseq or  bhid wpno wfiseq |
| Information of all jobs since last interview and this, one row for each job of each individual | wjobhist, wjobhistd(p,q,r) |  | pid wjspno or  bhid wpno wjspno |
| History of marriages, cohabitation, employment status and job histories before start of the survey. One row for each marriage, cohabitation, employment status and job spells of each individual | wmarriag(b,k,l)  wcohabit(b,k,l)  wlifemst(b,k,l)  wlifejob(c) |  | pid wmarno or  whid wpno wmarno  pid wlcsno or  bhid wpno wlcsno  pid wleshno or  bhid wpno wleshno  pid wleshno or  bhid wpno wleshno |
| Information about each natural, adopted/step children of the adult respondent | wchidad(b,k,l)  wchildnt(b,k,l)  wchild(l,m,q) |  | pid wlacno or  whid wpno wlacno  pid wlncno or  whid wpno wlncno  pid wljseq or  whid wpno wljseq |
| Responses from the youth questionnaire | wyouth | 11-15 year olds in the household | pid or  whid wpno |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1: Description of BHPS data files (continued)** | | | |
| **Contains** | **File Name** | **Responses provided by** | **Each row of observation is uniquely identified by** |
| *Sampling information & Paradata* | |  |  |
| Sampling information and information from the ARF (includes information on non-responding households) | whhsamp | Survey organisation and interviewer | whid |
| Household location and interview outcome information about every person in fielded households | windsamp | Interviewer | pid wfinloc |
| Interviewer information | xivdata | Survey organisation | ivno |
| *Derived files* |  |  |  |
| Relationship between every pair of household members | wegoalt |  | pid wapid or  whid wpno wapno |
| Fixed information about everyone in every enumerated households | xwavedat |  | pid |
| Information about the most recent status of time-varying data | xwlsten |  | pid |
| Interview outcome in every wave | xwaveid |  | pid |
| aIn most cases this was the person who rented or owned the accommodation (or the eldest if more than one). In the BHPS such a person is referred to as the Household reference person. | | | |