

Introducing the British Household Panel Survey using SPSS



Lee Bentley
UK Data Service

Aims

- What is the BHPS?
- Why use panel / longitudinal data to study attitudes
- How to organise the data for analysis
 - This will include a practical exercise

What we can't do

- Provide a discussion of the intricacies of the BHPS
- Offer advice on issues related to longitudinal modelling
 - This can be very complicated!
- Teach you how to use SPSS beyond the specific topics in the practical

Links

- BHPS website:
 - <https://www.iser.essex.ac.uk/bhps>
- Working paper: The British Household Panel Survey: Introduction to a longitudinal data resource:
 - http://www.longitudinal.stir.ac.uk/wp/lda_2006_2.pdf

What is the BHPS?

- Household panel survey
 - Interviews all adult (16+) household members
 - A youth questionnaire introduced in 1994 (age 11-15)
- Representative of the population of GB (south of the Caledonian canal) in 1990
- Booster samples in Scotland, Wales and Northern Ireland, have since been added
- Individuals rather than households are followed
- Ran between 1991 and 2008. Is now part of the Understanding Society sample.

How is it collected?

- Prospective survey with retrospective elements
- Most of the data is collected using face-to-face interviews, some by telephone (around 500 households only), some by proxy
- There is no refresher sample
- Can use geographic level data by linking BHPS data to these data using geographic locators (need special permission for some levels of geography)

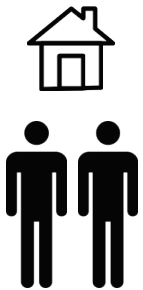
Sample design

- Initially a stratified random sample of households
 - 1991 included ~5,500 households and ~10,000 individuals
 - This represented a ~65% successful contact rate
- Individuals are followed and re-interviewed annually
 - The survey is comprised of:
 - Original Sample Members (OSMs)
 - Temporary Sample Members (TSMs)
 - Permanent Sample Members (PSMs)

Example of following rules

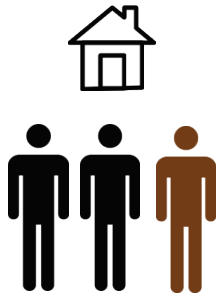
Wave 1:

1 Household:
two
individuals:
both OSMs



Wave 2:

1 Household:
Three
individuals: 2
OSMs; 1 TSM



Wave 3:

2 Household:
4 individuals:
2 OSMs; 2
TSMs



Wave 4:

2 Household:
3 individuals:
2 OSMs; 1
TSM



Wave 5:

2 Household:
4 individuals:
3 OSMs (inc.
child); 1 PSM



The sample over time

- The sample is reduced by...
 - Non-response / attrition: refusal and non-contact
 - Ineligibility: for example, moving beyond the geographical scope of the survey, or deaths
- It is increased by...
 - New OSMs due to birth
 - TSMs who may move in and out of the sample
 - Additional samples: 'ECHP', waves 7–11; Scottish boost, wave 9→; Welsh boost, wave 9→; Northern Ireland boost, wave 11→.

The sample over time

- The longitudinal sample is reduced in two ways:
 - Attrition: refusal and non-contact
 - Becoming ineligible: i.e. death or moving abroad
- By Wave 9: 19,653 individuals ever associated with this sample, including
 - 1,342 new births to sample members
 - 4,290 other new entrants
 - Over time, approximately the same number of persons interviewed each year (but number ever in panel increases over time)

Main themes

- Data are collected around a number of main themes:
 - household composition
 - housing conditions
 - residential mobility
 - education and training
 - health and the usage of health services
 - labour market behaviour
 - socio-economic values
 - income from employment, benefits and pensions

Taking advantage of repeated measures

- Is the BHPS suitable for what you want to do?
 - Cross-sectional for individuals, family and households: only one interview is required although waves may be pooled
 - Panels: longitudinal sample of individuals over some period of interest
 - Transitions: movement into and out of different states.
 - At least 2 points needed.
 - May be a pair of consecutive years or could be based on 2 pooled samples.
 - Spells and histories: interested in the time spent in different states.

Any questions?

- Next I will talk about the utility of the BHPS and attitudinal data from my own work.
- I'll then discuss some practical issues about merging and matching in the BHPS.
- Finally, we can look at some practical examples together using SPSS.

Thinking about social attitudes and the BHPS

- An example from my own research:
- I study the effect of certain attitudes, such as interpersonal trust, on social participation.

My research

- There are good theoretical reasons to suppose that more trusting people will participate more than the less trusting.
- Thinking about this in a cross-sectional analysis we might think about trust, x , having some impact on participating, Y .



However...

- We might also hypothesise that the relationship goes in the opposite direction.



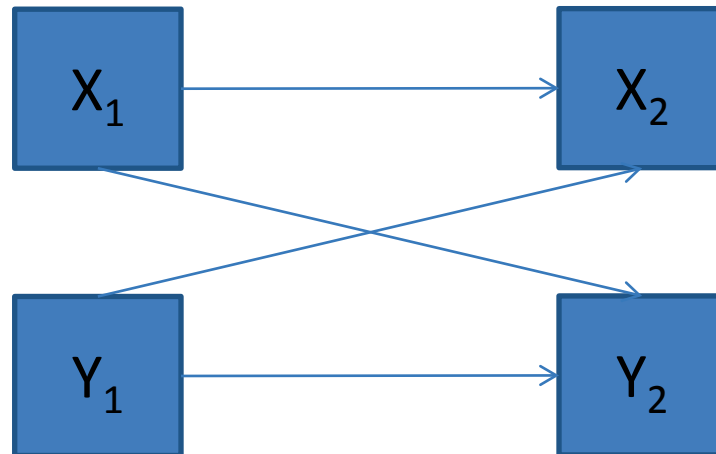
- More realistically it might be a reciprocal relationship.



- While we might be able to model the second relationship in cross-sectional data.
- However, it doesn't necessarily allow us to think very much about causality.

The utility of a panel design

- Having access to data at multiple time points allows us to make slightly stronger claims.
- We still can't determine causality but we can at least look at those things which occur before others.
- Taking trust, X , and social participation, Y , again:



What else might have an impact on participation in neighbourhood groups

- Clearly a this model is very simplistic
- We can also include other variables
 - Gender, class, age, etc.
- In my analysis I also think about attitudes about relationships in the neighbourhood
 - I estimate a factor model

Variables included in the factor

The friendships and associations I have with other people in my neighbourhood mean a lot to me

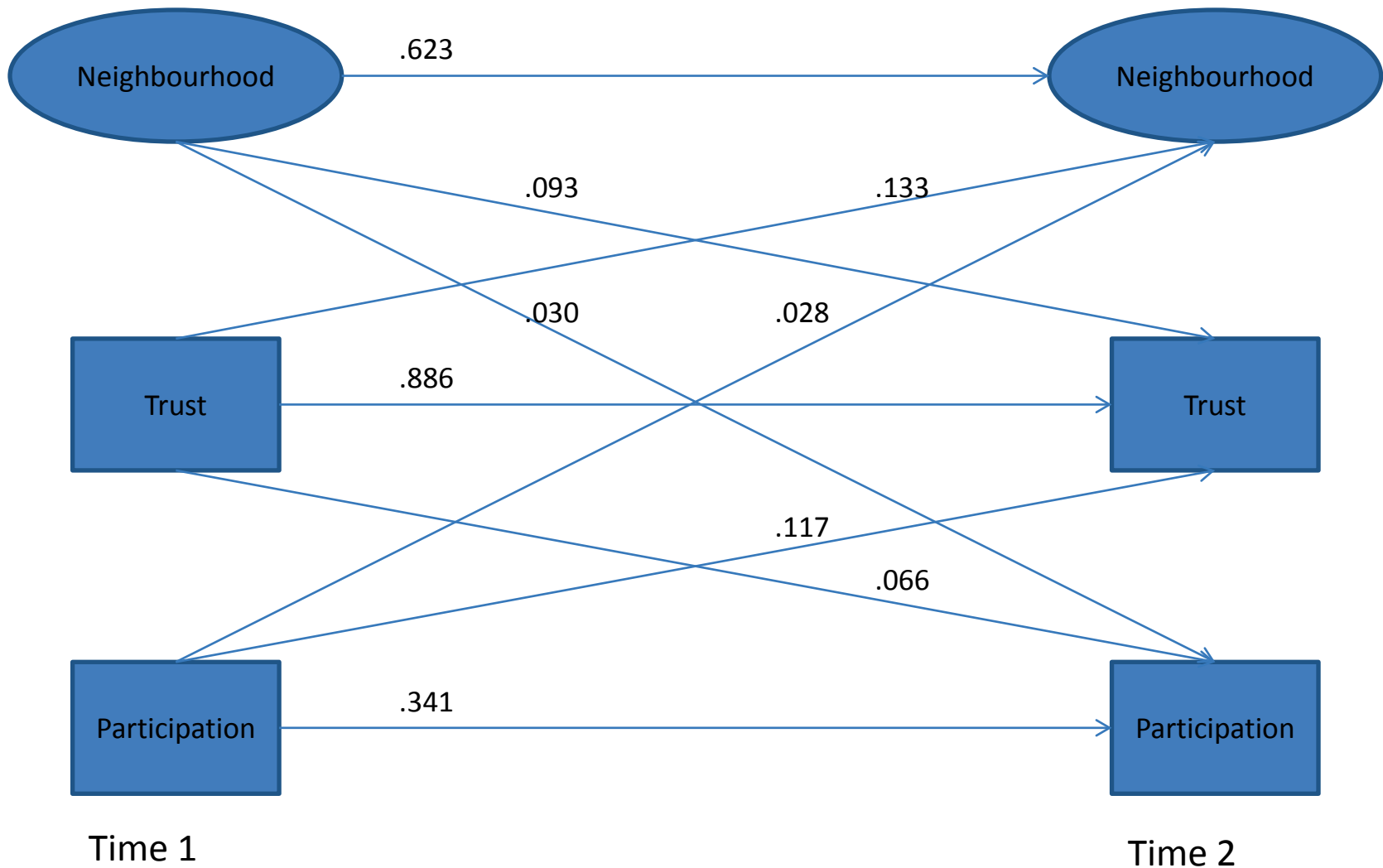
I feel like I belong to this neighbourhood

If I needed advice about something I could go to someone in my neighbourhood

I like to think of myself as similar to the people who live in this neighbourhood

I regularly stop and talk with people in my neighbourhood

What do I find...



BHPS: structure and linking

Data structure

- Each wave contains multiple files that you can link depending on what you want to do

File	Description	File	Description
wHHSAMP	Household coversheet	wINCOME	Income information
wHHRESP	Substantive household data	wYOUTH	Young persons' questionnaire (from wave 4)
wINDSAMP	Individual coversheet	wEGOALT	Relationship of HH members
wINDALL	Enumeration data for all individuals in HH	xWAVEID	Response status at each wave
wINDRESP	Substantive adult data	xWLSTEN	Sample membership status
wJOBHIST	Annual job history	xWAVEDAT	Fixed data about individuals

- Additional retrospective and derived data is also available

How do we link files?

- This depends on what you want to do:
 - In cross-sectional analyses you can use either PID or wHID and wPNO
 - PNO is the unique person identifier
 - wHID is the household identifier
 - wPNO is a person number within the household
 - Individuals across waves using PID
 - Households across waves: cannot be done
 - Spouses using wHGSPNO, SPPID
- Here we will focus on linking individuals across waves

Merging data

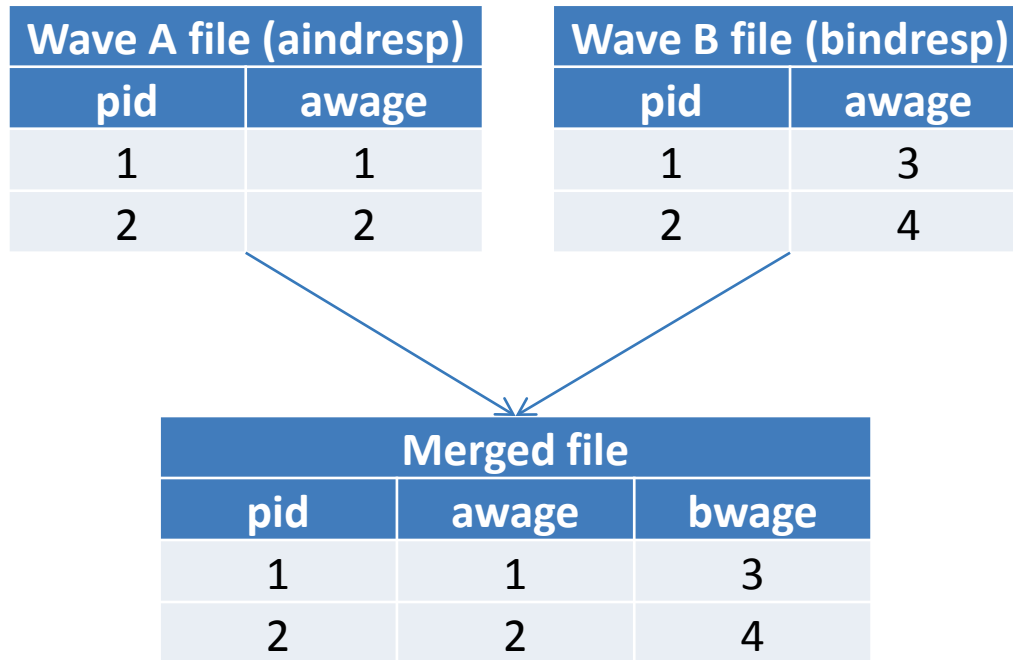
- There are two options when we wish to analyse data over time.
 - The first is to merge our data into a ‘wide’ format
 - The second is to merge the data into a ‘long’ format
- Which we choose will depend on the particular research question we are interested in answering
 - The long format is often used in specialised analyses such as fixed and random effects models, multilevel models, and duration models

The 'wide' format

- One row per case / individual
- Observations for different time points are in different columns
- Variable name identifies time via a prefix

pid	awage (Wage at W1)	bwage (Wage at W2)	cwage (Wage at W3)
10001	7.2	7.5	7.7
10002	6.3	missing	6.3
10003	5.4	5.4	missing
...			

Merging data: the 'wide' format

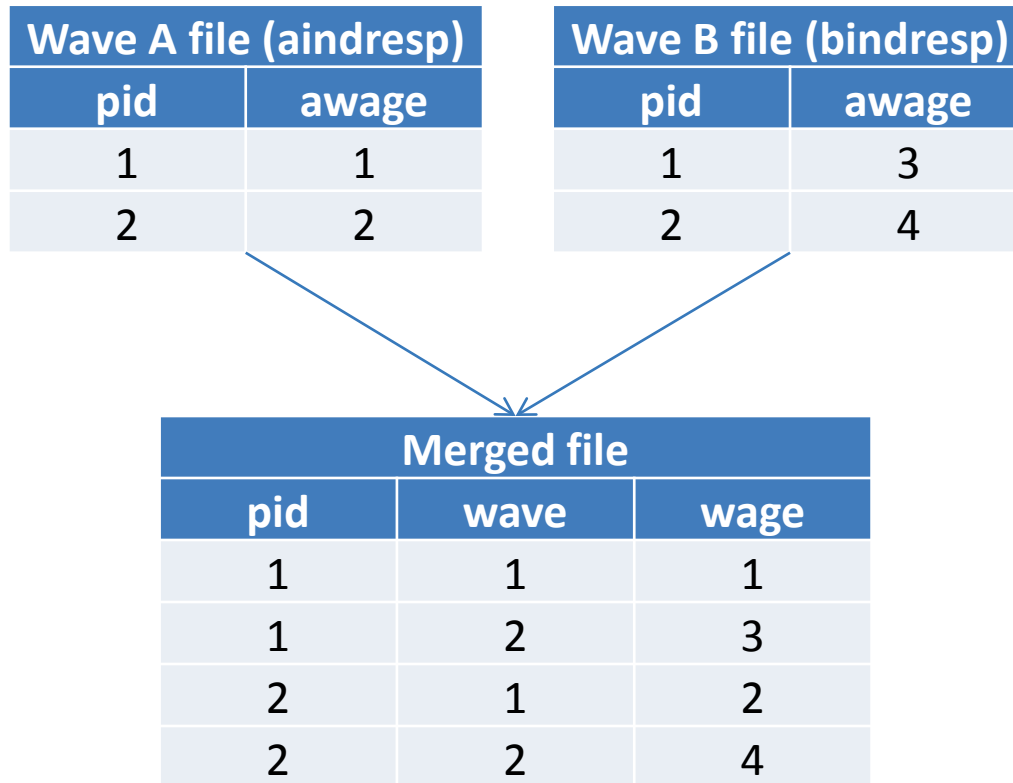


The 'long' format

- Multiple rows per case / individual
- Time points and individuals identified by individual and wave / time identifier: PID, wave

pid	wave	wage
10001	1	7.2
10001	2	7.5
10001	3	7.7
10002	1	6.3
10002	3	6.3
10003	1	5.4
10003	2	5.4
...		

Merging data: the 'long' format



Merging data in SPSS using the interface

- If we want to undertake a 'wide' merge we have to match our data against a common identifying variable
 - To match people across waves we can use the PID
- Before doing this you first need to sort the data on this variable so it matches.

Merging data in SPSS using the interface

Sorting data:

The screenshot shows the SPSS Data Editor interface with the 'Data' menu open and 'Sort Cases...' selected. The 'Sort Cases' dialog box is open, showing a list of variables on the left and 'Sort by: PID (A)' in the center. The 'Sort Order' is set to 'Ascending'. The 'Match cases on key variables in sorted files' checkbox is checked. The 'Both files provide cases' radio button is selected. The 'Indicate case source as variable' checkbox is unchecked, with 'source01' in the text field. The legend at the bottom indicates that (*) denotes an active dataset and (+) denotes bindresp.sav[DataSet1].

Merging data:

The screenshot shows the SPSS Data Editor interface with the 'Data' menu open and 'Merge Files' > 'Add Variables...' selected. The 'Add Variables to bindresp.sav[DataSet1]' dialog box is open, showing a list of variables in the 'New Active Dataset' field. The 'Match cases on key variables in sorted files' checkbox is checked. The 'Both files provide cases' radio button is selected. The 'Indicate case source as variable' checkbox is unchecked, with 'source01' in the text field. The legend at the bottom indicates that (*) denotes an active dataset and (+) denotes bindresp.sav[DataSet1].

Weighting

- As in many surveys it is necessary to weight the BHPS in order to ensure it is representative of the population
- Weights adjust for:
 - Unequal selection probabilities
 - Non-response at the household level
 - Non-response of individuals within households

Weighting

- We can weight the data in two ways
 - Cross-sectional weights are used within waves
 - Longitudinal weights are applied for over time analysis
 - Use the weight from the data of the last wave you are looking at, i.e. if you're looking at waves Q and R use the weights from R

Cross-sectional

Weight	Description
wXHWGHT	Household
wXRWGHT	Individuals respondents
wXEWGHT	Enumerated individuals

Longitudinal

Weight	Description
wLRWGHT	Individual respondents
wLEWGHT	Enumerated individuals