esds

Economic and Social Data Service

British Household Panel Survey: Waves 1-11, 1991-2002:

Teaching Dataset (Social and Political Attitudes)

ESDS Longitudinal

Author: Jack Kneeshaw
Version: 1.2
Date: 30 September 2010

Introduction to the British Household Panel Survey (BHPS)

The BHPS is carried out by the Institute for Social and Economic Research (ISER) at the University of Essex. The main objective of the survey is to further understanding of social and economic change at the individual and household level in Britain, to identify, model and forecast such changes, their causes and consequences in relation to a range of socio-economic variables. The BHPS is designed as a research resource for a wide range of social science disciplines and to support interdisciplinary research in many areas. The unique value of the survey resides in that fact that:

- it follows the same representative sample of individuals - the panel - over a period of years;

- it is household-based, interviewing every adult member of sampled households;

- it contains sufficient cases for meaningful analysis of certain groups such as the elderly or one parent families;

- it allows for linkage of data both from other surveys and from local area statistics.

The BHPS was designed as an annual survey of each adult (16+) member of a nationally representative sample of more than 5,000 households, making a total of approximately 10,000 individual interviews. The same individuals are re-interviewed in successive waves and, if they split-off from original households, all adult members of their new households are also interviewed. Children are interviewed once they reach the age of 16; there is also a special survey of 11-15 year old household members from wave four. Efficient fieldwork practices, training videos for interviewers and regular contact with the panel members ensures that the sample remains broadly representative of the population as it changes over time.

The sample

The British Household Panel Survey is a longitudinal survey of private households in Great Britain. The initial selection of households for inclusion in the panel survey was made using a two-stage stratified systematic method as a balance between efficiency and cost and is approximately equivalent to the current sample design of the General Household Survey (GHS). This sample design is an approximately equal probability of selection method (espem) design.

The frame used for the selection of sample units was the small users Postcode Address File (PAF) for Great Britain (i.e. excluding Northern Ireland). This is the main frame now used by large government surveys and is, in general, the most acceptable of available frames for the selection of a sample of households in Great Britain. In the first stage of selection 250 postcode sectors were selected as the primary sampling units (PSUs) from an implicitly stratified listing of all sectors on the PAF using a systematic sampling method. In the second stage of selection, delivery points, which

are approximately equivalent to addresses, were sampled from each selected PSU using an analogous systematic procedure.

Access to the main BHPS

New editions of the main BHPS data are deposited at the UK Data Archive approximately 12 months after the end of each wave's fieldwork. Further information about the methodology of the BHPS can be found in the BHPS user documentation on the Economic and Social Data Service (ESDS) website or on the ISER web pages. There is also an FAQ for the BHPS, prepared by the ESDS Longitudinal function (http://www.esds.ac.uk/longitudinal/).

How to obtain the BHPS Social and Political Attitudes teaching dataset

The BHPS teaching dataset can be obtained from the UK Data Archive by ESDS registered users and is subject to UK Access Management Federation (UKAMF) authentication. Further information can be found on the ESDS how to access data web page.

The teaching dataset is available in two proprietary formats: SPSS and STATA. The dataset is also made available in tab-delimited format.

The dataset can also be accessed via the ESDS Nesstar server – http://nesstar.esds.ac.uk/webview/index.jsp. See **Appendix C** for further details.

Data and variables within the dataset

In order to reduce the complexity of the data's structure and to make the data more accessible to inexperienced users, the dataset is in a rectangular format. Users are not required to undertake potentially problematic record/file merges in order to analyse the data longitudinally.

Again, for reasons of simplifying the dataset, the data file contains only wave one respondents (n=9912), and follows them for 11 waves, selecting broadly the same variables at each wave. There are some (fixed) variables which are only included at wave one, and there are a few variables which only appear intermittently, or are introduced later in the panel. The dataset contains a mix of original variables and derived variables (501 in total, see **Appendix A**).

Variable names are the same as those in the original BHPS dataset such that variables are prefixed A for wave one, B for wave two, C for wave three and so on. For

example, the variable EAGE records the respondent's age at the date of the wave five interview.

This data file - the second BHPS sampler - is smaller and more specialised than the first (SN 4901). The data file is based around the theme of social/political attitudes and allows for respondents' views and opinions on a range of issues to be tracked over time.

Missing values within the dataset

Many variables within the teaching dataset have values of –10 through -1. These are referred to as 'missing values'. Missing values are used to represent the variety of situations where respondents did not provide data in response to questions or where a variable could not be computed.

Missing value conventions in the BHPS teaching/sampler dataset:

-1 represents a respondent response of 'Don't know'.
-2 represents a respondent refusal.
-3/-4 are reserved for situations arising for particular questions where invalid data are given for other reasons, or data that do not fit into the frame of the main variable (e.g. self-employed person made a loss, last payment was a refund).
-7 indicates that the respondent was interviewed by proxy or by telephone and therefore the relevant question was not asked, or the derived variable could not be computed.
-8 represents data missing because not applicable to that respondent, or because of routing from some previous question.
-9 represents data missing in error, with no other explanation, or derived variables which could not be computed.
-10 indicates that the respondent was absent for this wave (i.e. was not interviewed).

For example:

Variable ARACE: Ethnic group membership

LiteralQuestion - Could you look at this card please and tell me which of these groups you consider you belong to?

Value	Categories	N		
1	White	9503		96.1%
2	Black-Carib	67	0.7%	
3	Black-African	45	0.5%	
4	Black-Other	26	0.3%	
5	Indian	107	1.1%	
6	Pakistani	44	0.4%	
7	Bangladeshi	8	0.1%	
8	Chinese	10	0.1%	
9	Other ethnic grp	83	0.8%	
-10	Respondent absent this wave	0		
-9	Missing or wild	10		
-8	Inapplicable	0		
-7	Proxy respondent	0		
-2	Refused	9		
-1	Don't know	0		

It is often useful to run frequencies on the variables as a first stage in any analysis to examine the distribution of responses and the proportion of –10 through -1 answers. Within this dataset you should generally exclude missing values in your analysis because the question is not relevant to these cases. However, this is not always the case – you may, for example, have a specific interest in the 'Don't know' responses.

Missing values have been dealt with slightly differently within the two versions of the teaching datasets.

Missing values in the SPSS teaching dataset (bhps_sampler2.por)
The SPSS teaching dataset has all –10 through –1 responses set as missing values.

Missing values in the STATA teaching dataset (bhps_sampler2.dta)
The STATA teaching dataset includes all –10 through –1 responses as valid responses. STATA has only one code for missing data denoted by a dot (.). You can turn the –10 through –1 values (or any value) into missing by using 'mvdecode'. For example to set –8 to missing for the variable DOBY you would type the following:

mvdecode doby, mv(-8)

Weighting the data

Both cross-sectional and longitudinal weights are included in the dataset. See main BHPS documentation, Volume A, for detailed information on applying these weights.

APPENDIX A

BHPS SOCIAL AND POLITICAL ATTITUDES TEACHING DATASET: CODEBOOK

A full data dictionary, listing all the variables, values and their labels, is provided in the documentation file, bhps_sampler2_UKDA_Data_Dictionary.rtf, that accompanies this study.

APPENDIX B

RELATED RESOURCES

Online user guide for the BHPS:
http://www.esds.ac.uk/longitudinal/access/bhps/L33196.asp

ESDS Longitudinal resources, including other teaching/sampler datasets:
http://www.esds.ac.uk/longitudinal/resources/introduction.asp

Learning and teaching resources:
http://www.esds.ac.uk/findingdata/learning.asp

APPENDIX C

BRITISH HOUSEHOLD PANEL SURVEY (BHPS):
NESSTAR SAMPLERS USER GUIDE

INTRODUCTION

Via the online Nesstar software tool, registered Economic and Social Data Service (ESDS) users can now analyse, visualise, subset and download selected data from the British Household Panel Survey (BHPS). Frequencies can also be run on the data by users who are not registered with ESDS.

The three BHPS 'samplers' are the first files with a longitudinal component to be added to the ESDS Nesstar Catalogue. This unique resource was created in order to provide potential users of this large and complex dataset with a taste of how the BHPS might be used. As with other Nesstar datasets, the sampler is in an accessible rectangular format and is therefore ideal both for exploratory research and as a tool for teaching on research methods courses.

Each sampler contains wave one respondents, and follows them for between eleven and thirteen waves, selecting broadly the same variables at each wave. The first sampler focuses on the theme of work, family and health, but it can also be used to examine social mobility and income mobility. The dataset itself contains a number of background variables indicating parental social class, and some opinion variables relating to the work and family theme. The second BHPS sampler file is based around the theme of social and political attitudes and the third is based around the theme of time use, leisure and social memberships.

This user guide provides worked examples from the first sampler.

1. BEFORE YOU START …

1.1. What is Nesstar?

Nesstar is a web-based system for the dissemination of data. The program allows users to search for, locate, browse, analyse and download a wide variety of statistical data within a web browser. There is no need for specialised software – data held on a Nesstar server can be viewed by anyone with a web browser.

1.2 How do I access Nesstar?

The UK Data Archive, in its role as service provider of the ESDS, hosts a Nesstar server populated with a wide variety of survey datasets. The ESDS Nesstar catalogue can be accessed via http://nesstar.esds.ac.uk/webview/index.jsp.

2. GETTING STARTED

2.1 ESDS Nesstar

Please take a moment to familiarise yourself with the Nesstar interface.

Before you should be two main frames. The left frame has a light blue background. In the right frame who should see the welcome to the ESDS Nesstar Catalogue message.

2.1.1 The left frame

At the top of the frame are two folders (e.g. ⊞) called Research Datasets and Teaching Datasets – click on the Teaching Datasets folder to open up its contents.
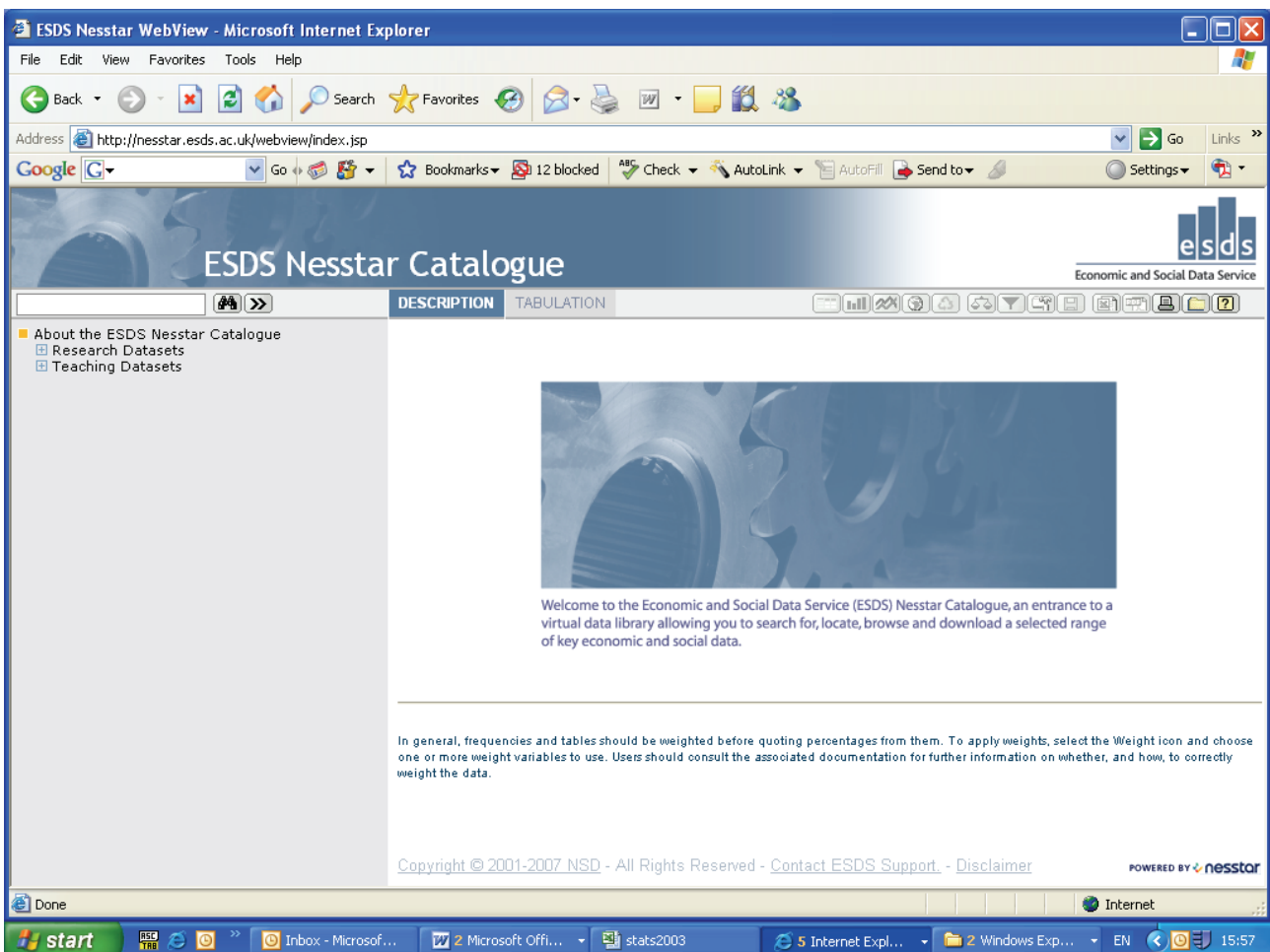
The left frame now contains a list of sub-folders (e.g. ⊞), each with the name of a particular survey series.

2.1.2 The right frame

When any action is performed in the left frame of the interface, the results are displayed in the right frame. We will see this when we come to browse the metadata and analyse the data.

3. THE BHPS SAMPLER DATASETS

3.1 Becoming familiar with the British Household Panel Survey sampler datasets

At http://nesstar.esds.ac.uk/webview/index.jsp, click on the Teaching Datasets folder and open up its contents. The list includes three BHPS sampler files, click on the icon (e.g. ▦) for the Work, Family and Health file.

You will be presented with two further sub-folders or 'trees': 'Metadata' and 'Variable Description'.

If you click on the Metadata folder you will open up two new folders: 'Study Description' and 'Data Files Documentation'. You should also notice that in the right frame, the survey's 'Abstract' has appeared.

3.2 The Variable Description tree

Open up the Variable Description folder, which is below the Study Description folder. The Variable Description tree houses a list of sub-folders. Each sub-folder contains a set of variables, grouped according to whether the variable is a 'classification' indicator or by the wave in which the variable appears.
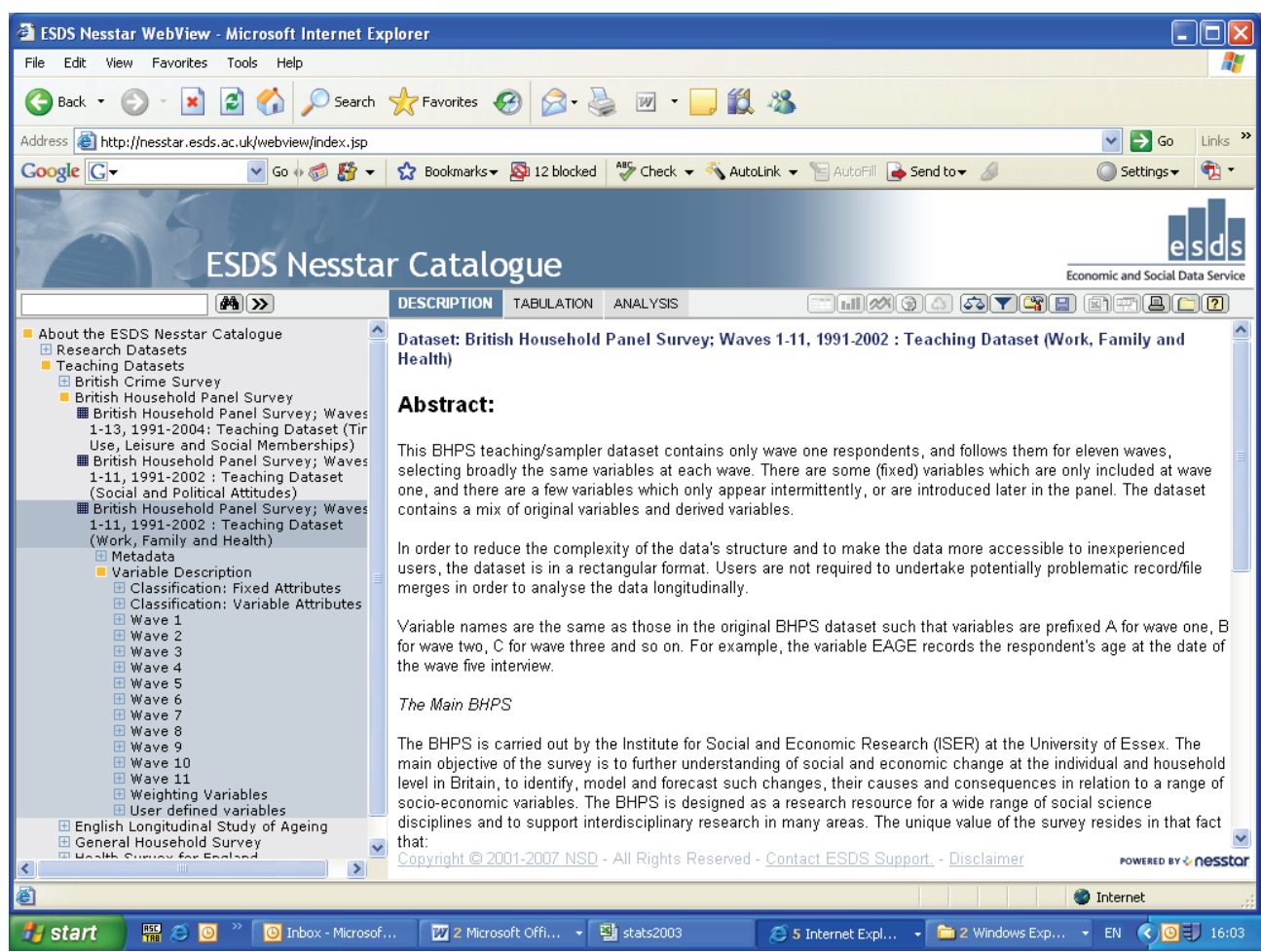
Click on the 'Wave 11' folder and then on the 'Opinions (Role of Father/Mother) sub-folder to view a list of the variables categorised under that topic. Now click on the text 'Pre-school child suffers if mother works'. Displayed in the right frame you should now see the actual question asked in the survey and the possible responses to the question together with the coded value for each response and the respondent totals and percentages.

4. EXPLORATORY ANALYSIS

4.1 Cross-tabulations

In addition to requesting frequencies, Nesstar allows registered ESDS users to conduct cross-tabulations.

To conduct a cross-tab you first need to click on the 'Tabulation' tab at the top of the right frame. In the right frame, a template for a table will then appear with instructions on how to populate the table. Follow these instructions by moving to the list of variable group folders in the left frame and opening the 'Wave 1' folder and then the 'Health and Health Behaviour' folder. Click on the variable 'Smoker'.
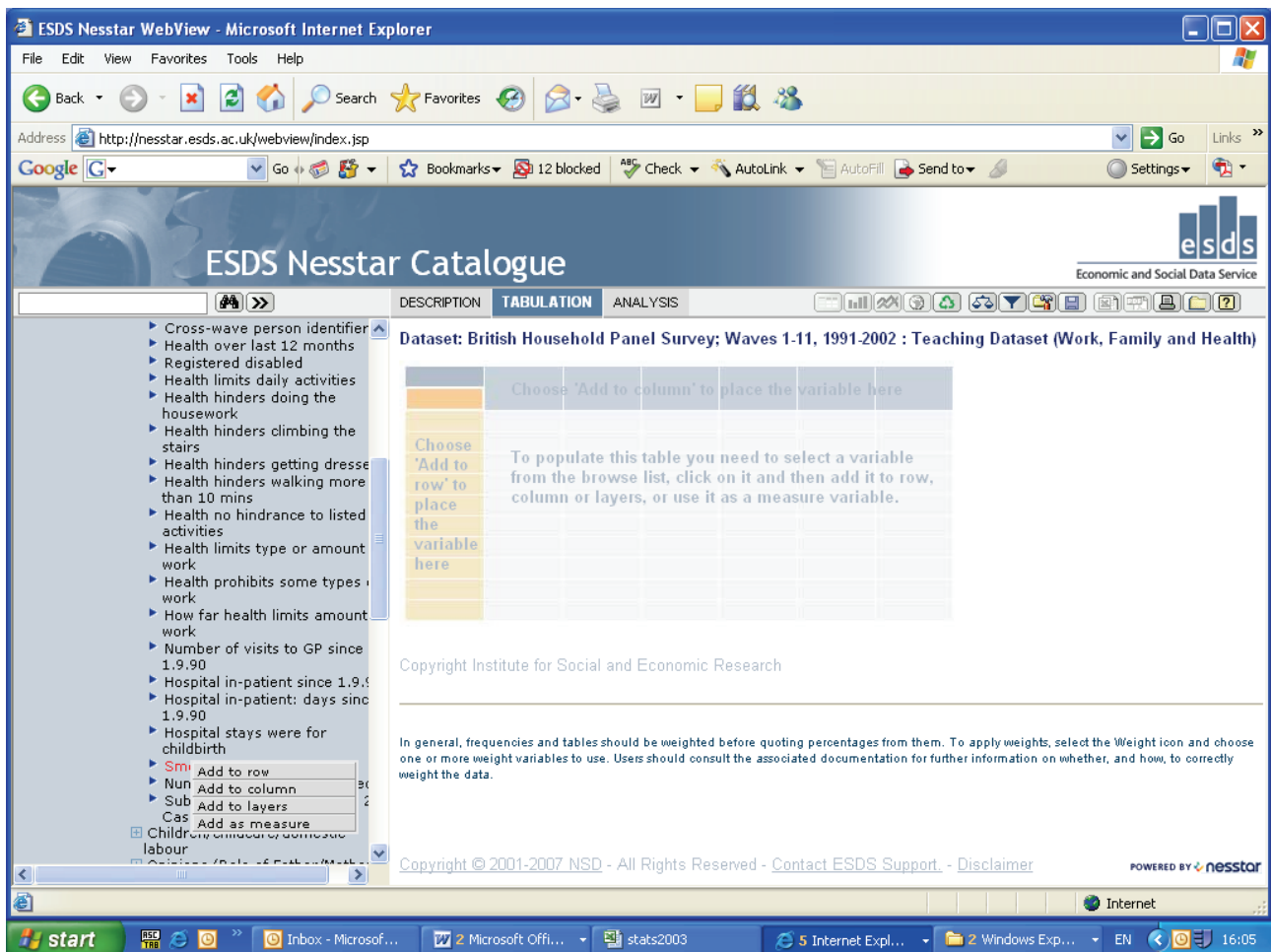
The variable should now be highlighted in red and a drop-down box should appear alongside it. Click on 'Add to row'.

At this point, you will be prompted to enter your institution (or UK Data Archive) username and password (see earlier section on 'How to obtain the BHPS <subject area> teaching dataset'). Having entered your details, a frequency table for the wave 1 variable 'Smoker' should be displayed in the right frame.

To create a two-way table, select a second variable in precisely the same manner as you chose the first. For example, highlight 'Smoker' from the Wave 11/Health and Health Behaviour folder and this time 'Add to column'. The results for the two-way table should appear in the right frame. Responses for the *first* variable selected (Wave 1 'Smoker') can be read across the rows and responses for the *second* variable (Wave 11 'Smoker') can be read down the columns. This order is important to remember as corresponding variables in different waves will all have the same label (i.e. the variable that relates to how people are 'managing financially these days' is called 'Financial situation' in all eleven waves).

4.2 Weighting data

The Nesstar software allows users to weight their results. To apply a weight, the user should click on the 'Weight' button (![icon], at the top of the right frame) and then select the appropriate weight by clicking it across to the 'Selected Variables' box. Click the 'Ok' button to view the results with the selected weight applied.
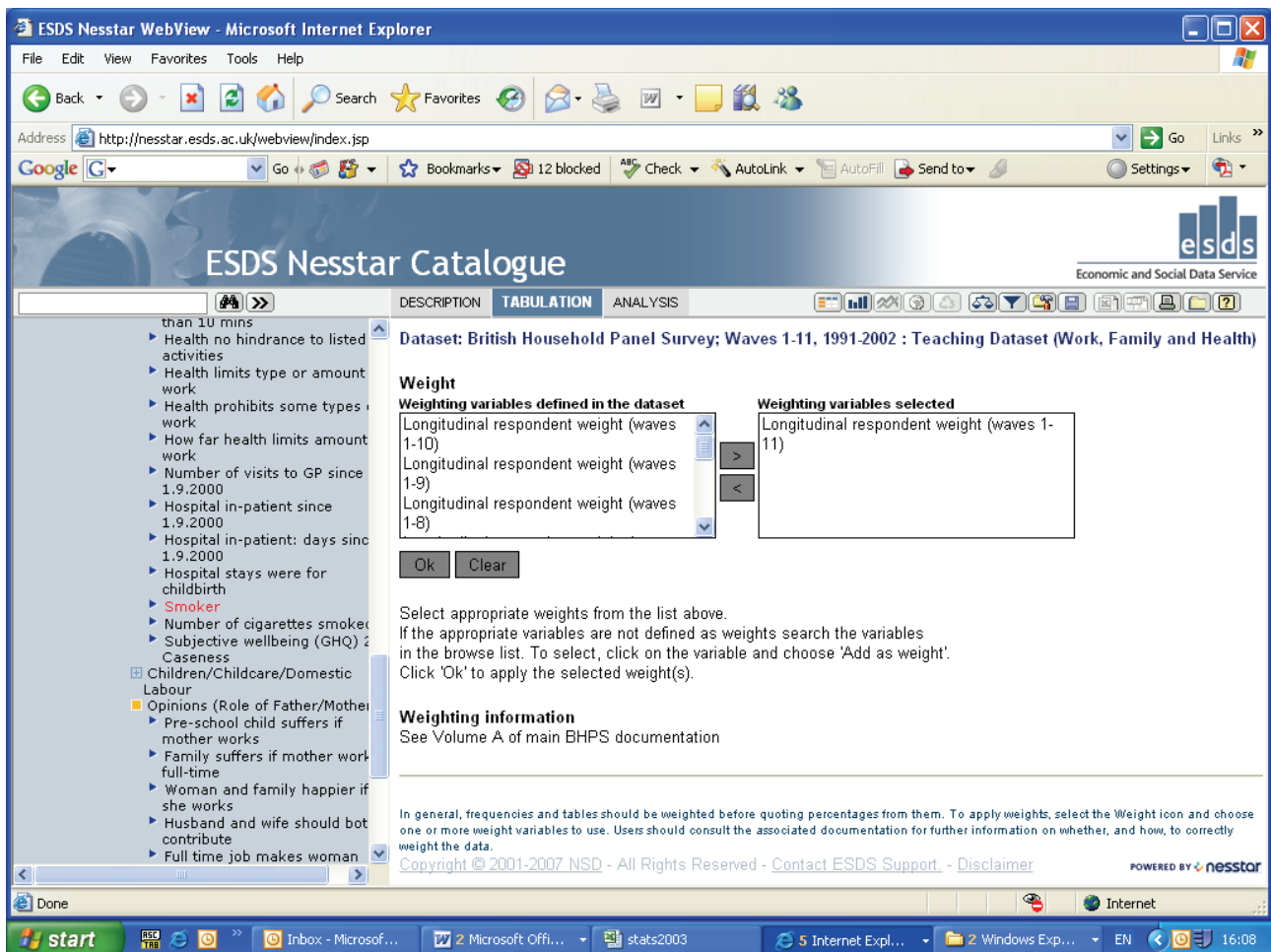
4.3 Subsetting data

Nesstar users may also subset the data that they analyse. For example, to select male respondents only for analysis, click on the 'Subset' button (▼). Employing the same strategy used to select variables for frequencies and cross-tabs, select the variable 'Sex' from the Classification: Fixed Attributes folder and click on 'Add to subset'. Next, highlight '1 Male' in the right-hand box and then click on the 'Add' button. Click the 'Ok' button to view the results for male respondents only.

5. DOWNLOADING DATA

Having browsed and explored the BHPS sampler datasets within the Nesstar environment, users may download the dataset in order to undertake further analysis in their preferred data analysis software package.

To download the dataset, click on the 'Download' button (🖫), select your preferred data format and then click 'Download'.
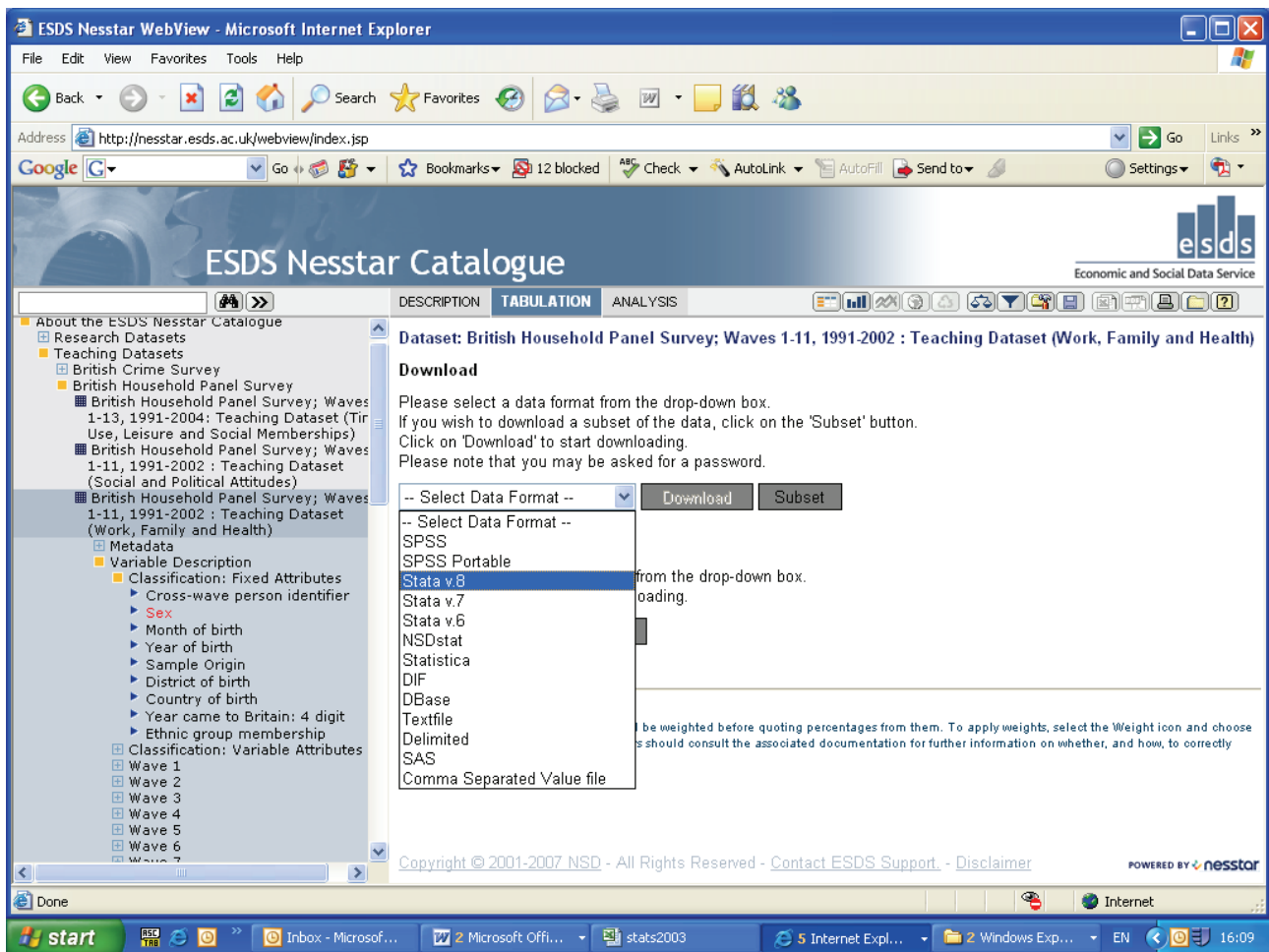
6. MORE ABOUT NESSTAR

A more complete guide to using Nesstar can be found at http://www.nesstar.com/download/documentation.html. Remember that to run cross-tabs and undertake other exploratory analysis beyond requesting frequencies, you will need to be registered with the ESDS (http://www.esds.ac.uk/).

esds

Economic and Social Data Service