Appendix 3

Additional information on technical details of the datasets used in the thesis

1. The European Community Household Panel (ECHP)

The ECHP-UDB released by Eurostat includes several files, some wave specific, others covering all the waves.

The 'country file' and the 'longitudinal-link file' cover all waves. The first contains information for each wave and country on population figures, purchasing power parities, purchasing power standards, and exchange rates figures. The second contains a record for every person that ever appeared in the ECHP and allows rebuilding the 'longitudinal status' of the person from the beginning to the end of the panel.

There is a series of files available for each wave: the 'register file', the 'relationships file', the 'household file' and the 'personal file'. The 'register file' contains a record for each person currently living in a private household with a completed household interview. The 'relationships file' records the relationship between each pair of persons in the same household.

The main files in terms of analysis are the 'household file' and the 'personal file'. The 'household file' contains one record for each household with a completed household interview. The data in the household file are grouped into seven sections, as follows:

- HG: General information
- HD: Demographic information
- HI: Household income
- HF: Household financial situation
- HA: Accommodation
- HB: Durables
- HL: Children

The 'personal file' contains one record for each adult with a completed personal interview. The information is grouped into thirteen sections as follows:

- PG: General information
- PD: Demographic information
- PE: Employment

- PU: Unemployment
- PS: Search for a job
- PJ: Previous job
- PC: Calendar of activities
- PI: Income
- PT: Training and education
- PH: Health
- PR: Social relations
- PM: Migration
- PK: Satisfaction with various aspects of life

The target population of the ECHP survey includes all the private households throughout the national territory of each country. The sampling frames used in the Member States included the population register, master samples created from the most recent census of population and houses, the postal address registers and the electoral roll.

The sample size of each Member State was defined on the basis of various criteria (theoretical, practical and budgetary). In the first wave (1994) the total community sample was slightly over 60 000 households and 130 000 persons aged 16 years and over. It should be noted that generally, larger countries, because of their greater need for disaggregate results but also because of their greater capacity, received larger samples. Within each country, the sample was distributed proportionally across geographical regions, so as to maximise the precision of the estimates at the national level. Some countries, however (i.e. Spain and Italy), have chosen disproportionate allocations with a view to ensure a minimum sample size for each region of the country (sampling smaller regions at higher rates).

All surveys in the ECHP are based on probability sampling. Most of the surveys are based on a two-stage sampling (i.e. selection of sample areas in the first stage, followed by the selection of a small number of addresses or households at the second stage, within each selected area). However, there are cases where a single stage sample is drawn or, on the other hand, a three-stage sample takes place.

Diverse criteria are used for the stratification of area units before selection. The most common criterion used in the surveys is the geographical region and/or urban-rural classification. Stratification by population size or other social indicators was also used in

some countries. Within explicit strata, areas were selected systematically in some countries, randomly in others (Peracchi 2002).

Table A3.1 gives information on the number of cases available for cross-sectional analysis in each wave released at the time of the analysis.

Table A3.1. Number of observations available for analysis in each file and wave of the ECHP, 1994-1998

Wave	Year of survey	Household file	Personal file
1	1994	71 367	149 306
2	1995	73 715	156 063
3	1996	74 746	157 536
4	1997	68 788	143 935
5	1998	66 097	136 238

Source: Eurostat documentation released with the ECHP-UDB.

The overall participation rates, by country and by wave, are generally high, despite some national variations. In the first wave, there is a high of 91.4% in Italy and a low of 83.1% in Ireland (Nicoletti and Peracchi 2002). More importantly, the attrition rates along waves as well as the response rates vary substantially across countries. Nicoletti and Peracchi, in their analysis of non-participation in the ECHP, have summarised for the first 5 waves the following figures:

Table A3.2. Response rates and attrition rates, by country, along the first 5 waves of the ECHP

Country	Response rate	Attrition rate
Germany	63.8	16.8
Denmark	46.8	31.9
Netherlands	56.1	20.1
Belgium	57.1	26.5
France	58.1	26.6
United Kingdom	61.8	14.8
Ireland	44.7	40.0
Italy	62.4	19.5
Greece	55.5	27.6
Spain	50.4	29.6
Portugal	62.4	16.0

Source: (Nicoletti and Peracchi 2002)

The variability in response rates and attrition rates is important for many reasons, among which I would highlight the selective character of its occurrence. Some researchers have demonstrated, for example, that sample attrition is more incident among the oldest and among the less educated groups of the population (Nicoletti and Peracchi 2002). This can

have implications in the findings one gathers from the available data. In this thesis, though, and given Portugal is the case of interest, these problems may have a lesser impact. Portugal has shown one of the highest response rates and one of the lowest attrition rates along the first 5 waves of the ECHP.

The data from the ECHP included in the analysis refer to wave 5, 1998, the most recent series available at the time of the empirical analysis[1].

However, throughout the analysis, there were moments when it was considered relevant to include a time-variant approach to clarify some dynamics. This was done using cross-sectional comparisons for different points in time, and involved using data from the first wave of the ECHP (1994). It was considered that a 5 years interval could shed some light on some general trends of interest for the thesis. The use of data for 2 distinct years was confined to the cross-national comparison part of the empirical analysis. The national samples used for cross-national comparisons were limited to the cases present in wave 5 that were also present in wave 1. The countries considered included only those present in wave 1. Luxembourg was excluded given there were no data available for that country in wave 5.

The final sample sizes of the different ECHP countries included in the analysis are those presented in table A3.3.

Table A3.3. National samples sizes for ECHP countries[a] included in the analysis, by type of file and by population

Country	Household file (all sample)	Personal file (all sample)	Household file (elderly sample)	Personal file (elderly sample)
Germany	4 917	9 344	825	1 058
Denmark	1 967	3 220	369	463
Netherlands	3 785	6 575	709	932
Belgium	2 454	4 411	554	703
France	4 991	9 430	1 101	1 419
United Kingdom	4 249	7 045	960	1 155
Ireland	2 449	5 124	608	757
Italy	5 334	12 891	1 388	1 782
Greece	3 732	8 138	1 200	1 508
Spain	4 582	10 575	1 442	1 911
Portugal	3 960	8 852	1 338	1 716
Total	**39 971**	**85 605**	**10 494**	**13 404**

Notes:
a) The samples for Germany and the UK refer to the original ECHPdata.

---

[1] Eurostat issued the first release of the ECHP-UDB, covering waves 1 and 2, in December 1998, three years after completion of fieldwork for wave 2. The second release, covering the first three waves, was issued in December 1999. The third one, covering waves 1 to 4, was released in June 2001. The fourth one, used in this research, covering waves 1 to 5, was released in February 2002.

Table A3.4. below displays figures for the relative weight of the target population (individuals aged 65 or more) in the final ECHP national samples and in the respective national populations.

Table A3.4. Relative weight of the target population in the ECHP national samples analysed and in the respective national populations [1], in 1994 and 1998

		65 to 79 years		80 or more years	
		1994	1998	1994	1998
Germany	ECHP	14.2	18.3	2.9	4.8
	Population	13.4	14.4	4.8	4.4
Denmark	ECHP	13.6	16.0	2.6	5.1
	Population	13.9	13.4	4.7	4.7
Netherlands	ECHP	14.2	17.2	1.7	4.1
	Population	12.4	12.6	3.7	3.9
Belgium	ECHP	14.0	17.5	2.2	4.0
	Population	14.4	15.7	4.5	4.4
France	ECHP	13.3	17.0	3.2	4.9
	Population	13.3	14.7	5.1	4.7
UK	ECHP	16.3	19.1	3.6	6.3
	Population	14.8	14.8	4.8	4.8
Ireland	ECHP	11.6	14.1	1.8	3.6
	Population	12.0	11.6	3.2	3.3
Italy	ECHP	12.3	16.8	3.6	5.7
	Population	14.5	15.8	4.5	4.7
Greece	ECHP	14.4	19.0	2.1	4.1
	Population	14.1	15.4	3.8	3.7
Spain	ECHP	13.8	17.7	2.3	4.3
	Population	13.9	15.0	4.0	4.3
Portugal	ECHP	13.5	16.9	1.9	4.1
	Population	14.2	15.0	3.6	3.9

Source: ECHP, waves 1 and 5 (own calculations on total cases retained for analysis); EUROSTAT population estimates for years 1994 and 1998
Note: [1] shares in age group adjusted to same base population as in ECHP: individuals 16 or more years of age

The analysis of data from the ECHP takes both the individual and the person as units of analysis, and at times combines these. In order to combine the person and the household, some variables are used for matching files. Each person is identified by a unique identification number (PID) and by a country code. Also each household is identified by a unique identification number (HID) and by a country code. Matching these variables across files allowed the building of a final dataset combining information for each person with information about his/her household, in the two waves selected for analysis.

In terms of the variables/dimensions of information available in the ECHP, the analysis carried out in this thesis has considered the following:

- Household demographics

- Household income

- Housing conditions

- Status of dependent children in the household

- Personal demographics

- Employment history

- Personal income

- Health status and needs

- Social relations and personal networks

- Relationships between persons in households.

In the ECHP-UDB files, weights are available for households and persons. These weights, according to Eurostat documentation, are calculated taking into account the sample design and characteristics of persons and households[2]. The weights are calibrated to reflect the structure of the population.

In the ECHP there are two main types of weights: the base weight (at individual level only), used for longitudinal analysis (not required for the analysis carried out along the thesis); and the cross-sectional weight (at both household and individual level) for use in cross-sectional analyses.

The weighting factor for the purpose of pooling of countries in cross-sectional analyses was normalized, which means that the differences in frequencies between weighted and non-weighted variables are very small. In any case, and whenever cross-national analysis was involved, a weighting factor was used for all the calculations. Each table of results introduced in the following chapters identifies if the data were weighted or not.

Harmonization of data was an issue that involved in particular the analysis of income data for the purposes of cross-national comparisons.

The ECHP database includes information on PPP's (Purchasing Power Parities) as defined by Eurostat. Given that, in each wave of the survey, the data recorded referred to income of the previous year, the harmonization of income is done by dividing the data in each income variable of interest by the corresponding national value for the PPP's in the previous year. For example, income from wave 1 (1994) is harmonized according to the PPP's defined for 1993. Income data was recorded in national currencies. In the case of

---

[2] Eurostat releases a series of ECHP documents together with the ECHP-UDB files, designed to describe the main technical details of the survey. PAN 165 is the document describing the weighting procedure that has been implemented for calculating individuals' and households' weights.

Italy, data was recorded in thousand of lire, which means that harmonization of income for the Italian sub-sample involved a previous multiplication by 1000 of the original data in the dataset.

2. The Portuguese Family Budget Survey (FBS)

Like all its European equivalents, the FBS is confined to the population residing in private households. As to geographical coverage, the survey covers the entire population residing in private households in the national territory.

The FBS sample is obtained from a master sample drawn for the purpose of common use in different surveys, and defined by INE on the basis of the most recent census data.

The FBS makes use of probability sampling and of a two-stage design. First, a stratified sample of suitable area units is selected, typically with probabilities proportional to size after stratification by region, socio-economic status of the reference person (head of household) and household type or size. The second stage consists of the selection, within each sample area, of households and addresses for inclusion in the survey.

The total number of observations available for analysis, by file, in the FBS 2000, are the following: 28 311 individuals and 10 020 households.

The use of the FBS in the analysis was restricted to the in-depth analysis of the Portuguese case and focusing exclusively on financial dynamics.

The analysis of FBS data focused on the sub-sample of elderly, defined as all those aged 65 or more. The total number of sampled individuals is 6 217 in 4 447 households. The final database built for the analysis includes variables with information on individual characteristics of the elderly but also information on other members of the households, namely detailed socio-economic and demographic characteristics of the head of household.

The FBS database includes weighting variables, both for households and for individuals, to adjust results to the size, distribution and characteristics of the population. Given it was not possible to obtain, from INE, a detailed description of the weighting procedures used, it was decided not to weight FBS data.

Procedures of harmonisation of data do not apply to the FBS data.

Table A3.5. below displays information on the relative distribution of different age groups in the sampled households of the FBS and the respective estimated weights in the Portuguese population in 2000.

Table A3.5. Relative weight of the target population in the FBS sampled households and in the Portuguese population, in 2000

Age group	N in sample	% in sample	% in total population [1]
< 5 years	1057	3.7	5.4
5 – 9 years	1399	4.9	5.2
10 – 14 years	1763	6.2	5.4
15 – 19 years	2097	7.4	6.7
20 – 24 years	1928	6.8	7.8
25 – 29 years	1478	5.2	7.9
30 – 34 years	1564	5.5	7.1
35 – 39 years	1850	6.5	7.3
40 – 44 years	1962	6.9	7.1
45 – 49 years	1813	6.4	6.6
50 – 54 years	1747	6.2	6.3
55 – 59 years	1648	5.8	5.5
60 – 64 years	1788	6.3	5.3
65 – 69 years	1867	6.6	5.1
70 – 74 years	1756	6.2	4.5
≥ 75 years	2594	9.2	6.7
Total	28311	100.0	100.0

Source: FBS, 2000 and INE
Note: [1] the shares of the total population by age group are estimates for 2000 released by INE and available online at www.ine.pt

3. Eurobarometer Surveys (EB)

The EB survey series takes as target population individuals residing in private households in the space of the European Union. It selects sampled individuals by multi-stage random route sampling of persons 15 years old and older.

EB surveys do not involve representative samples of countries. Given that the main goal of the program is to identify profiles of attitudes and opinions, statistical representativeness is not a core technical requirement. The majority of the national samples include around 1000 individuals. The exceptions are Luxembourg with a sample of approximately 600 respondents, Northern Ireland with approximately 300 respondents and Germany with around 2000 respondents (1000 in the East of Germany and 1000 in the West of Germany).

Information about response rates in Eurobarometer is not published.

The EB surveys are not panel sets therefore the analysis does not involve a longitudinal approach to data. However, and given that values and social norms have a clearer meaning if analysed at the societal level and not at the individual level, by examining data at different

points in time it can be introduced in the discussion some ideas on values change/resilience.

The information recorded in the EB surveys concerns individuals, this being the unit of analysis. The information available is used for the discussion on the normative dimension of welfare arrangements and, when possible, tries to relate that to several individual characteristics. In section 3 of this chapter some more in-depth discussion about the variables used for the analysis will be resumed.

Starting with EB 32 (1989), for each of the participating countries, a comparison between the sample and a proper universe description is carried out for internal weighting procedures. The universe description is made available by the national research institutes in charge of the survey and/or by Eurostat. On that basis, a national weighting procedure, using marginal and intercellular weighting, is applied (criteria used include sex, age, region NUTS II and size of locality). Population size weights correct for the fact that most countries have identical sample sizes, no matter how large or small their populations are.

For purposes of pooling of countries for analysis, all EB survey datasets contain variables with weighting factors, defined according to different interests in terms of analysis (GESIS). Given the purpose of the analysis carried out within this thesis, I have used the following weighting factors:

- WEIGHT RESULT FOR TARGET: it reproduces the real number of cases for each country and can be used when the national samples are analysed separately.
- EURO WEIGHT 15: it includes the adjustments of each national sample in proportion to its share in the total population of the European Union (at the time of the analysis only comprising 15 countries). It is used when the total population of the 15 EU countries is to be analysed as a whole.

As for harmonisation of data, the EB series does not require any harmonisation procedure. This is explained especially by the fact that income variables are made available as aggregate results, expressed as relative position to distribution quartiles. This is, as discussed below, one of the limitations of the EB surveys data.

4. Measurement of variables and data completeness

In this section the reader finds some information on how some specific variables were measured in the original surveys. These are variables that show central in the analysis put forward in the thesis but that are also potential source of data problems when conducting a survey. Those variables include: income; health status and education.

Table A.3.6. below summarises how these variables were measured in the original surveys and how data was manipulated before inclusion in the final datasets used for the analysis.

Table A.3.6. Measurement of variables in the original surveys and organisation of data in the final datasets

Variables	Dataset(s)	Original measurement of variables	Procedures for obtaining data in datasets
Personal income	ECHP FBS	Interviewed individuals are asked to provide exact amount for a series of sources of income ranging from paid employment, to self employment, social benefits and private income.	Amounts of income declared in each type of income are added up to compute final personal income. Procedures of income imputation were developed to reduce missing data.
Household income	ECHP FBS	Not asked to interviewed individuals.	Amounts of household income are obtained from the data available for the personal income of the members of the household. Imputation procedures were developed to minimise missing data. The teams managing the databases run both procedures.
	EB	Individuals are asked to state the monthly income of their households relative to a list of income-intervals.	Data is harmonised relative to the quartile distribution. Final data available for analysis include the relative position of each respondent's household income to the quartiles of the national income distributions.
Long-standing limiting health problem	ECHP EB	Individuals are asked if they have any health problem that limits their daily activities. If	Original answers are presented in the final datasets.

		answering yes, individuals are asked to qualify the extent of the hampering condition: if moderate or if severe.	
General health status	ECHP	Individuals are asked to rate their own health status, being presented with the following alternatives: very good; good; fair; bad.	Original answers are presented in the final datasets.
Education	ECHP	In each national survey individuals are asked about their highest level of formal education completed, according to the national typologies.	A final typology is derived from the national datasets, standardising responses to a three categories typology: recognised 3rd level education; second stage of secondary level education; less than second stage of secondary education.
	FBS	Individuals are asked to identify the highest level of formal education completed in a 10 categories typology.	Original answers are presented in the final datasets.
	EB	Individuals are asked about the number of years of schooling completed.	Data is presented in a three categories typology: up to 15 years; 16 to 19 years; 20 or more years.

In terms of data completeness, it should be stated that neither the ECHP nor the FBS surveys are affected by large amounts of missing data, this in particular for the income variables. This is related to the imputation procedures developed by the managing teams of each survey. It was not possible to assess the reliability of those imputation procedures. The EB series is systematically affected by large numbers of missing data for the income variable, across datasets with an average of 25% of cases.