Andreas Behr,

Egon Bellgardt,

Ulrich Rendtel

(J.W. Goethe-Universität Frankfurt/Main)

Extent and determinants of panel attrition in the European Community Household Panel

Extent and determinants of panel attrition in the European Community Household Panel[*]

Andreas Behr, Egon Bellgardt, Ulrich Rendtel[**]

JEL classification C4, C8

This draft: 10.11.2002

Contents

1. Introduction

In this paper we analyse the panel attrition in the European Community Household Panel (ECHP). The ECHP is a longitudinal survey on households and individuals covering countries of the European Union (EU). Due to its far reaching harmonization, the ECHP offers the opportunity of cross country comparisons within the EU in many different research areas of economics and sociology.

Comparability across countries is the target of the ECHP. While great effort has been devoted to harmonizing the surveys, still they are not completely standardized. The differences mainly have to do with the sampling procedure, the panel's-follow up rules and the field operations.

The existence of panel attrition is often seen as a potential factor that could reduce the value of a panel and it may thus to some extent reduce the level of harmonication of the ECHP. Depending on the missing mechanism attrition may even lead to biased results due to non-random attrition. In our analysis we will in a first step give detailed description of the attrition that happened in the ECHP. In a second step we try to find out what determines attrition by the means of logistic regression analysis.[1]

In the following section we give a very brief description of the ECHP, a survey is provided by Peracchi (2002) containing detailed information about the organization of the survey.

The second section gives thorough definitions of important concepts used to describe the nonparticipation and participation patterns across countries. Based on these definitions we compare the response behaviour across countries over the first five waves of the ECHP. The comparison of response rates for different subpopulations reveals considerable differences according to socio-demographic and field variables.

While there have been attempts to split the response behaviour into sequential decisions, namely the contact and the decision to respond when contact has happened (Nicoletti and Peracchi 2002), we do focus on the outcome of this process.[2] We differentiate only whether an eligible person has been interviewed or not. This is in our view justified as we are finally interested in assessing the attrition and the possible effects of the entire attrition process on empirical results based on the surveys.

Based on the findings of section three we try to explain in section four the differences found in the response behaviour within and between the different countries and within and between different waves by making use of a logit regression model. We estimate response probabilities using socio-demographic as well as well as information of the data collecting process.

The empirical findings we obtain for all countries and all waves of the panel will help to detect similarities as well as differences in the attrition behaviour across the EC countries. In this analysis individual information of persons from the wave prior to the attrition will be used as well as information of the year of attrition if available. The latter contains information like whether the individual has moved or if a new interviewer was assigned. Finally, section 5 contains the conclusions.

---

[1] An overview of econometric methods proposed for the case of non-random attrition is given by Verbeek and Nijman (1995), see also Hausman/Wise (1979). For a general discussion of problems caused by missing data see Little/Rubin (1987).

[2] The attrition within the ECHP was also studied in a paper by Nicoletti and Peracchi (2002) and by Peracchi (2002). The first paper analyses attrition focusing on housholds, while the latter analysis covers the first three waves of the ECHP only.

2. Description of the ECHP

The first wave of the ECHP in 1994 covered about 130,000 individuals above 16 years living in about 60,000 households. In the first wave 12 countries took part, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal Spain and UK. While Austria took part from the second wave on in 1995, Finland started its participation in 1996 and Sweden in 1997.

The ECHP was aimed "in response to the increasing demand in the European Union for comparable information across the Member States on income, work and employment, poverty and social exclusion, housing, health, and many other diverse social indicators concerning living conditions of private households and persons".[3] The most attractive feature of the ECHP for research is its standardisation

In most of the participating countries the survey was newly started, while a couple of countries made use of already existing panel surveys. In Belgium, the Netherlands and Sweden already ongoing surveys where used to gather data, while in three countries, Germany, Luxembourg and the UK a unique situation emerged as for three years two surveys ran parallel. In 1997 the newly started ECHP surveys in these three countries were terminated and the data for the ECHP from that year on are derived from the already existing national surveys. These are the German Social Economic Panel (GSOEP), the Luxembourg's Social Economic Panel (PSELL) and the British Household Panel Survey (BHPS). The User Data Base only covers the ECHP survey in Luxembourg, hence we only regard parallel surveys in Germany and the UK in our analysis by country.

This analysis is based on the 2002 ECHP-User Data Base which contains the following countries and waves:

Fig. 1: Participation of Countries in the ECHP

	1994	1995	1996	1997	1998
Germany	■	■	■		
Germany SOEP	■	■	■	■	■
United-Kingdom	■	■	■		
United-Kingdom BHPS	■	■	■	■	■
Denmark	■	■	■	■	■
Netherlands	■	■	■	■	■
Belgium	■	■	■	■	■
Luxembourg	■	■	■	■	
France	■	■	■	■	■
Ireland	■	■	■	■	■
Italy	■	■	■	■	■
Greece	■	■	■	■	■
Spain	■	■	■	■	■
Portugal	■	■	■	■	■
Austria		■	■	■	■
Finland			■	■	
Sweden				■	

---

[3]   Eurostat (1996), cited after Peracchi (2002), p. 64.

Because we assume effects of the number of waves being more important compared to effects of a given year, all data are ordered by country and wave. This means that data of wave 1 will include mainly data from 1994, but also from 1995 (Austria), 1996 (Finland). Because for Sweden only one wave is available (1997) what renders most of the analysis impossible, we do not include Sweden in the following analysis. When pooling the data over all countries we only consider one of the two parallel surveys running in Germany and UK , while in country-specific analysis we include both available data sets for these countries.

In the following we concentrate on individuals as the relevant unit. A detailed description of participation patterns based on the household as the relevant unit is given by Peracchi (2002).

3. Participation patterns across countries

In this section we present the results of a descriptive analysis of panel attrition. Because there is no clear-out definition of panel attrition, we first give a thorough definition of response categories used throughout this paper.

3.1. Defining relevant categories

The basis of our classification are persons included in the first wave in the ECHP or in the used national panels ("sample persons"). To select these sample persons we make use of the longitudinal link file (containing all persons ever included in the ECHP or in the used national panels) and choose those persons who are included in the first available wave of the sample in question.

Next we segment the sample persons into the following three subgroups with relation to their panel participation status: (1) respondents, (2) non respondents, (3) ineligibles. The segmentation can be accomplished by making use of the wave-specific set of variables describing the interview result status, a variable that characterises the residential status of the sample persons in each wave and the age of the person.

1. A *respondent* is a sample person for which the interview result variable indicates that the personal interview has been fully completed or that the reduced questionnaire has been completed. Apart from new entries after the first wave who are not considered here, this group is identical with the group represented in the personal file (p-file).

2. A **non respondent** is a sample person for which the interview result variable indicates one of the following events:

   a)   lack of cooperation or

   b)   absence or

   c)   other types of non contact or

   d)   other types of non cooperation:
   - the interview result variable indicates that the data of an completed interview is not transmitted to Eurostat,
   - the interview result variable is not applicable and the residential status shows that the person is lost and the person in question is not a child,
   - the interview result variable indicates a not completed interview with missing reason.

3. A sample person is **ineligible** if the interview result variable is not applicable and

a) the residential status variable indicates that the person has died in the wave in question or

b) the person moved outside the EU or moved to an institution or

c) the person is a child (age < 16) in the wave in question or

d) the person is not applicable according to the interview result variable and is not yet considered in the taxonomy above.

Fig. 2: Subgroups of sample persons
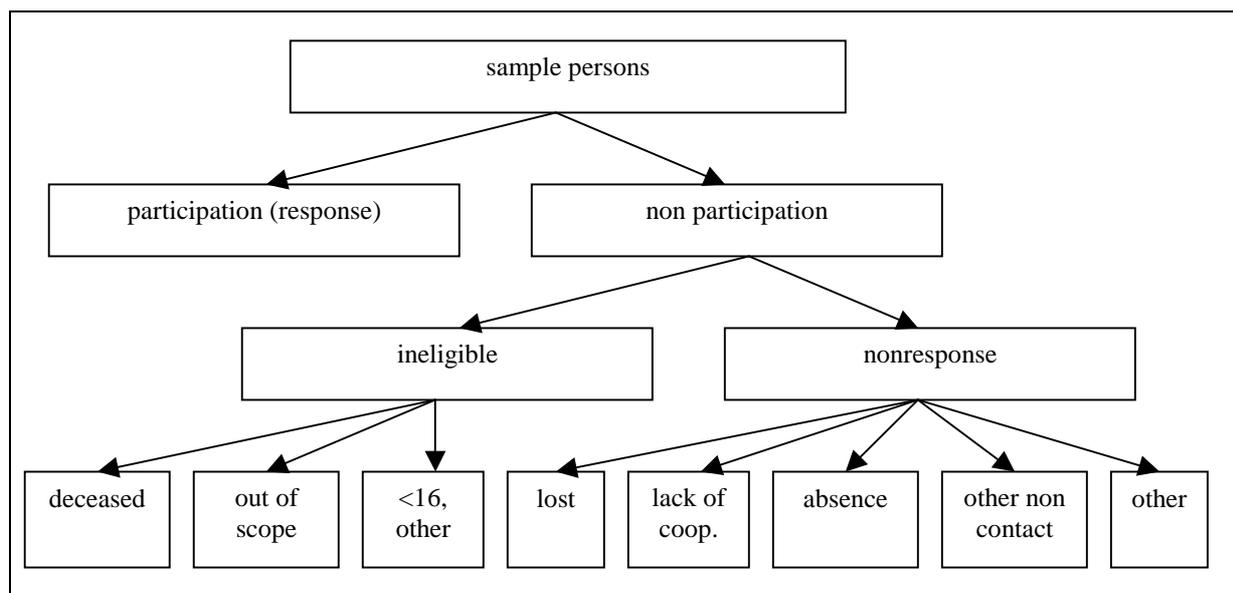
3.2. Empirical evidence on sample size, response and attrition

The ECHP samples in the different European countries are of comparable size, the largest country sample (Italy) is about 7.7 times the sample size of Luxembourg, while the largest country (Germany) has a population more than 200 times the population of Luxembourg. The smaller relative difference in sample sizes are due to very different sample ratios, which are considerably higher for smaller countries. The sampling fraction ranges from 0.64 percent for Luxembourg to only 0.014 percent for Germany (D-ECHP). We find that the sample fraction is negative correlated to the population size ($r = -0.29$). The following table contains information on the population, the adult population of participating countries in the first wave (1994 for all countries except Austria (1995) and Finland (1996)) and the sample size achieved in the first wave. The last column contains the sampling fraction in percent.

Table 1: Sample Size, shares of countries and initial response rate

	Population	Adult population in private households (target population) in mio.	% of total target population	sample size (1000)	resp. in wave 1 (1000).	initial response rate, %	% of total respondents	sampling fraction in %
Germany	80.585	67.173	16.7	12.944	9.490	73.3	5.8	0.014
Germany SOEP	80.585	67.173	16.7	19.302	12.233	63.4	7.4	0.018
United-Kingdom	57.489	45.465	11.3	15.237	10.517	69.0	6.4	0.023
UK-BHPS	57.489	45.465	11.3	17.502	9.028	51.6	5.5	0.020
Finland	5.063	4.033	1.0	11.858	8.173	68.9	5.0	0.203
Denmark	5.216	4.254	1.1	13.18	5.903	44.8	3.6	0.139
Ireland	3.520	2.582	0.6	15.943	9.904	62.1	6.0	0.384
The Netherlands	15.080	12.343	3.1	15.944	9.407	59.0	5.7	0.076
Belgium	9.971	8.018	2.0	11.501	6.710	58.3	4.1	0.084
Luxembourg	0.397	0.319	0.1	3.491	2.046	58.6	1.2	0.642
France	56.088	44.627	11.1	21.012	14.333	68.2	8.7	0.032
Spain	38.783	31.096	7.7	25.3	17.893	70.7	10.8	0.058
Portugal	9.893	7.922	2.0	17.271	11.621	67.3	7.0	0.147
Austria	7.905	6.401	1.6	10.808	7.437	68.8	4.5	0.116
Italy	56.301	46.980	11.7	26.736	17.729	66.3	10.8	0.038
Greece	10.206	8.465	2.1	17.897	12.492	69.8	7.6	0.148
All countries	494.571	402.316	100.0	255.926	164.916	64.4	100.0	0.041

In the following figure we display the response rates in wave 2 up to wave 5 as well as the overall response rate in the latest wave.

Turning to the ratio of respondents in the last wave (horizontal bar in the figure) to respondents in wave 1, we find considerable differences across countries. The ECHP is most affected by attrition in Ireland where the remaining share of respondents dropped to 54%. In the UK-ECHP, which only lasted three years, response rates have been the lowest across the EU (about 80%) resulting in only 62% respondents after three years. High response rates were attained in Germany, the UK-BHPS (which started already in 1991), Luxembourg and Portugal. Beside in the UK-ECHP and in Ireland response rates are also below average in Denmark and Spain.

Fig. 3: Response rates across countries for wave 2 to wave 5 and the overall response rate
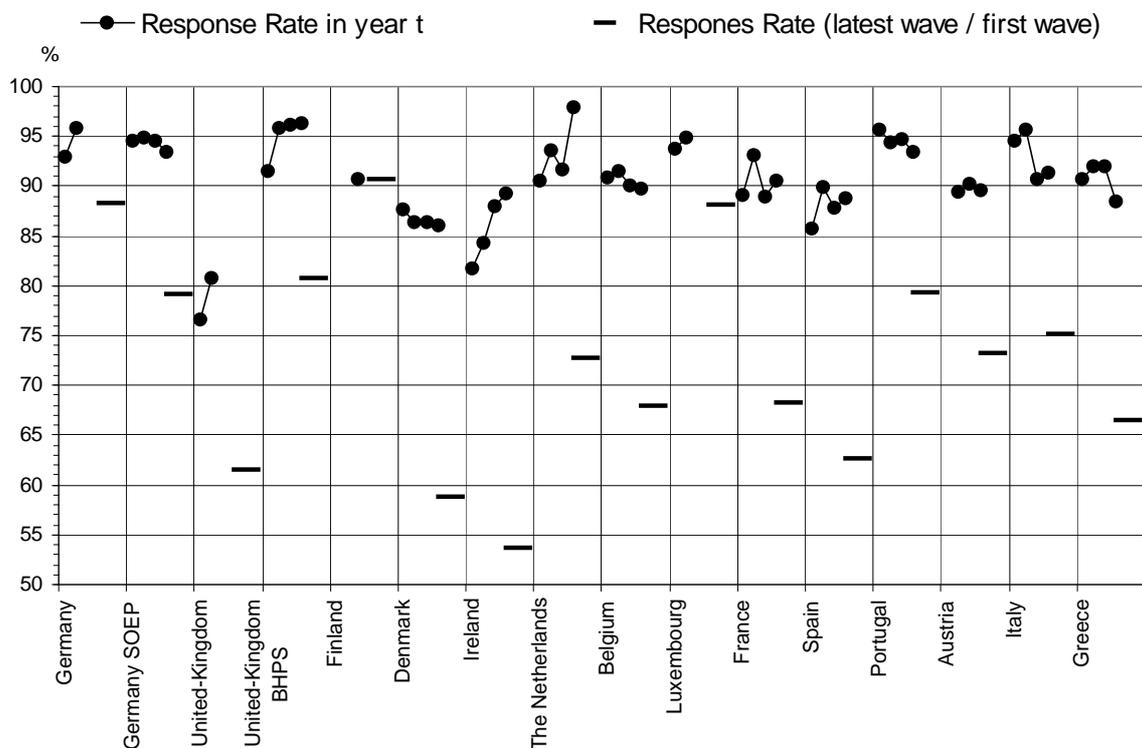
The figure also makes evident, that there is no clear tendency across all countries in the response rates to rise or fall. While we have increasing response rates in the German-ECHP, the UK-ECHP, in Ireland, the Netherlands and Spain, we find slightly decreasing response rates in the German-SOEP, Belgium and Portugal. In the remaining countries there is no clear tendency present.[4]

3.3. Types of participation

To describe the participation patterns we make use of a taxonomy of participation patterns given in table 2.

Our calculations are based on a variable indicating the availability of an interview in a specific wave in the personal file (p-file). It takes the value 1 if a person in the link-file (a list of all persons in the sample) has completed the personal interview; otherwise the variable is 0. Whether or not this person belongs to a household with a completed interview recorded in the register-file (r-file) is ignored.[5]

---

[4]   Household and individual attrition rates during the first three waves of the ECHP are also given by Peracchi (2002).

[5]   Using a different taxonomy Nicoletti/Peracchi (2002) give empirical evidence on attrition patterns by countries.

Table 2: Types of participation patterns, all 5-wave panels

Participation type	No. of persons	Percent
never participating	51,003	25.3
always participating	85,605	42.5
monotone non-participation after wave 1	11,200	5.6
monotone non-participation after wave 2	7,872	3.9
monotone non-participation after wave 3	8,185	4.1
monotone non-participation after wave 4	8,946	4.4
new entry	17,032	8.4
irregular participation, present in last wave	4,788	2.4
irregular participation, not present in last wave	6,957	3.5

We find that 25.3 percent of all people contained in the p-file never participated in the panel. This group of persons mainly consists of children below the age of 16. The four different monotone attrition patterns are shown by 3.9 and 5.6 percent of persons. There is no clear tendency over time. 2.4 percent of all persons considered in this analysis are respondents in the last wave but display an irregular response pattern. This group of persons can be considered as being subject to higher risk of attrition in future waves.

New entry (8.4 %) happens mainly because of children becoming 16 years and therefore leaving the status of ineligibility. The second source of entry is the formation of new households including sample persons and formerly non-sample persons.

When comparing the participation patterns across countries we do not consider never responding persons and consider four different categories only: always participating, new entry, monotone attrition and irregular participation.

Fig. 4: Participation types by countries

☐ allways participating   ▨ new entry   ▨ monotone attrition   ■ irregular participation
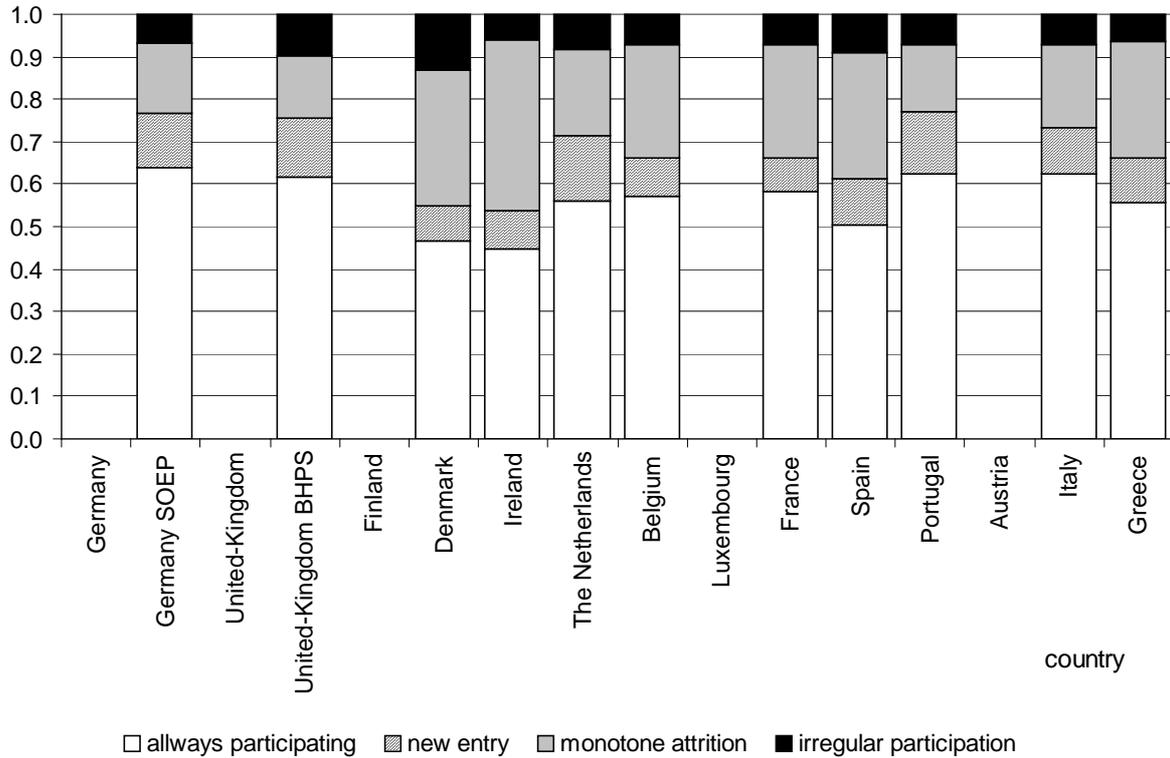
We find the highest share of always participating persons (about 60 %) in the German SOEP, the UK BHPS, in Portugal and Italy. The Netherlands and Portugal face the highest share of new entries, while in Denmark the share of irregular participation is largest.

3.4. Panel stability of the ECHP across countries

The following figure contains the maturation of the ECHP, pooled over all countries.[6] It contains four different categories of sample persons: (temporary) nonrespondents, respondents, children and other ineligibles (without final nonrespondents).

For a specific duration analysis a particular group of persons of the first available wave is defined. This "initial group" includes three types of sample persons:

(1) respondents of wave 1 (R_1),

(2) children of wave 1 (C_1) and

(3) temporary non respondents (TN_1) of wave 1 (non respondents according to the definitions in section 2.1, who are not monotone attritors in wave 1 according to the definition in section 2.2).

In each subsequent wave i it is checked to which of the above three types the persons of the initial group belong or whether they got ineligible in the actual wave i what corresponds to a

---

[6]   We consider only countries in this picture that took part in all five years, hence Germany (SOEP), UK (BHPS), Denmark, Netherlands, Belgium, France, Ireland, Greece, Spain and Portugal.

9

type (4) ineligible I_i. To summarize additionally the following relation is computed for all waves i:

$$X_t = \frac{R_{i(1)} + C_{i(1)} + I_i}{R_1 + C_1 + TN_1}$$

with

$R_{t(1)}$ person belonging to the initial group who is respondent in wave i

$C_{t(1)}$ child belonging to the initial group who is still child in wave i

$I_{t(1)}$ person belonging to the initial group who is ineligible in wave i

X_t may be interpreted as a measure of panel stability.

In 1994 all eligible sample persons are considered. Figure 5 displays the whereabouts of the three categories nonrespondents, who might become respondents or obtain the status of ineligibility, the children growing into the categories of respondents or nonrespondents when reaching the age of 16 and finally the respondents who can become nonrespondents or obtain the status of ineligibility as well as staying respondents.

Fig. 5: Whereabouts of the ECHP sample persons
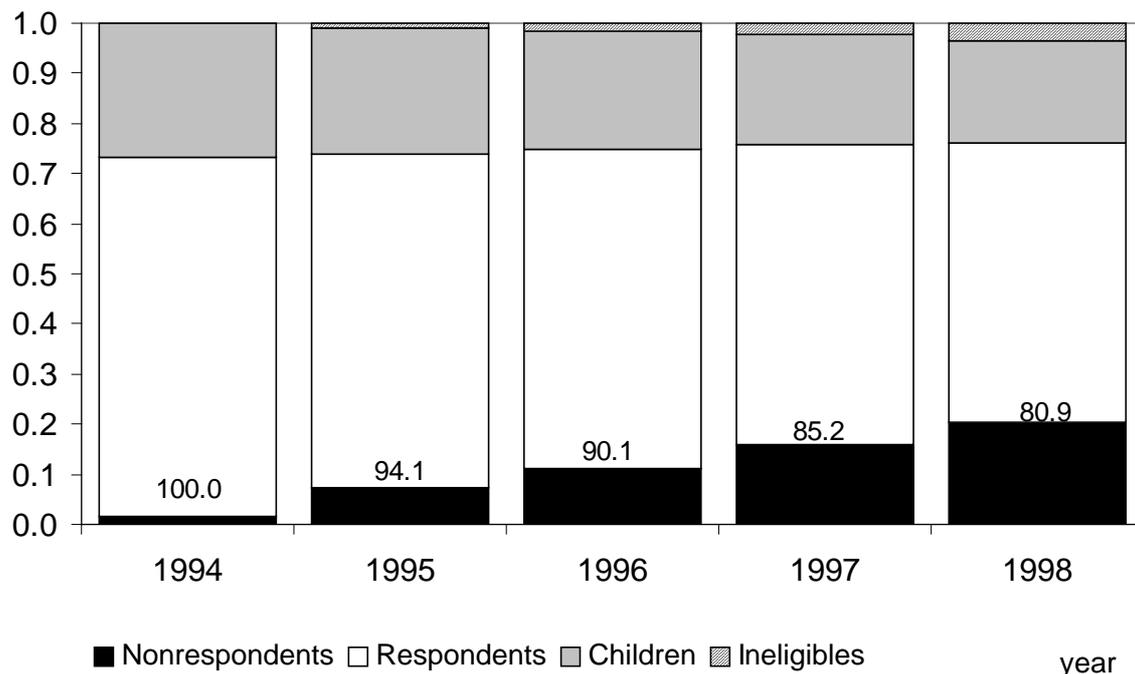
Considering all surveys that took part in all five waves, we find that the ratio of wave 5 respondents to the respondents in the first wave 1994 is 80.9 %. As the survey matures, we find an steadily increasing share of nonrespondents, a small but growing share of ineligibles and steadily decreasing share of children due to reaching the age 16.

Using the same graphical representation we depict the evolvement of the ECHP across countries. The two parallel national surveys, the SOEP in Germany and the BHPS in the UK are displayed for the relevant time span as well.

10

The following 16 figures show the proportions of the defined types and the development of the relation X_t.

Fig. 6: Whereabouts of the ECHP sample persons across countries

Germany

	1994	1995	1996	1997	1998
	100	94.6	91.7		

Germany SOEP

	1994	1995	1996	1997	1998
	100	96.8	94.0	91.2	87.5

United-Kingdom

	1994	1995	1996	1997	1998
	100	82.5	71.4		

United-Kingdom BHPS

	1994	1995	1996	1997	1998
	100	95.4	94.0	92.0	90.3

Finland

	1994	1995	1996	1997	1998
			100	93.7	

Denmark

	1994	1995	1996	1997	1998
	100	92.5	84.9	79.2	73.4

Ireland

	1994	1995	1996	1997	1998
	100	88.3	79.7	74.1	69.3

The Netherlands

	1994	1995	1996	1997	1998
	100	93.9	90.7	86.1	85.5

Belgium: 1994: 100, 1995: 95.5, 1996: 90.6, 1997: 85.1, 1998: 80.0

Luxembourg: 1994: 100, 1995: 95.8, 1996: 92.6

France: 1994: 100, 1995: 92.4, 1996: 89.1, 1997: 82.9, 1998: 77.9

Spain: 1994: 100, 1995: 90.2, 1996: 85.5, 1997: 79.4, 1998: 74.0

Portugal: 1994: 100, 1995: 96.9, 1996: 93.8, 1997: 91.1, 1998: 87.4

Austria: 1995: 100, 1996: 92.9, 1997: 88.1, 1998: 83.0

Italy: 1994: 100, 1995: 97.9, 1996: 95.9, 1997: 89.6, 1998: 85.3

Greece: 1994: 100, 1995: 95.0, 1996: 89.8, 1997: 84.6, 1998: 78.3

Comparing the figures for the two different surveys for Germany we find that response rates are considerably higher in the SOEP which already started in 1984 compared to the ECHP-sample for Germany. This fact of decreasing attrition rates for a panel over time over the first

waves is also observed for other surveys like the PSID.[7] For many European countries we observe this fact as well, as was shown in section 3.2.

For the UK the difference between the already ongoing BHPS and the newly started ECHP in 1994 to be much more pronounced than in the German case. The decrease in the rate of respondents shows a remarkable difference of 22.6 % after three years. After five years in the UK-BHPS the ratio of respondents in 1998 is still above 90% of that figure in 1994.

Despite the effect of children reaching the age of 16 and hence increasing the number of respondents over time, we observe a considerable shrinkage of respondents within the five waves under consideration.

3.5. Income, age and response behaviour

In this section we analyse the relation between age and response behaviour as well as the relation between OECD-household income per person and response behaviour.[8] For each of the sixteen surveys, thereof two for Germany and two for the United Kingdom, we show in one figure the density distribution for the first wave (thin line and right scale) as well as two response rates. The upper bold line depicts the response rate in wave two , while the lower bold line shows the overall response rate. In this case we have count as response only permanent participation in the survey in all waves.

Fig. 7: Age and response rates across countries[9]

---
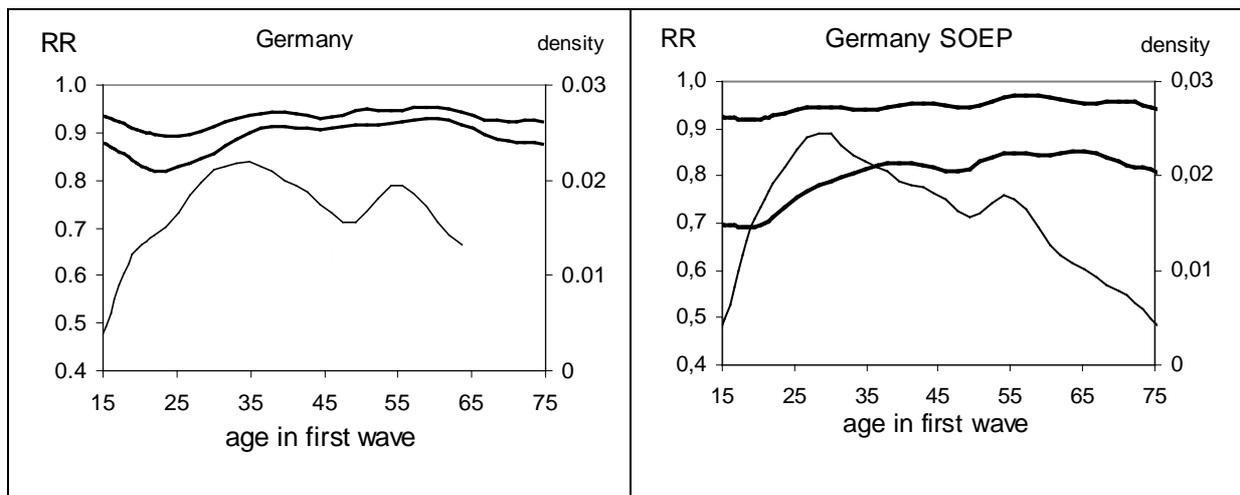
[7]   In the PSID the attrition rates stabilized at about 2 percent after wave three. See e.g. Fitzgerald/Gottschalk/Moffitt (1998).

[8]   For about one third of all households some kind of imputation was made for household income items. See Peracchi (2002) for details about income item-nnresponse and imputation methods in the ECHP.

[9]   The ending of the density for the age at 64 in the figure of the German ECHP sample is due to the fact that for persons above 65 years only a categorial measure is given indicating the belonging to that group.

United-Kingdom

United-Kingdom BHPS

Finland

Denmark

Ireland

The Netherlands

14
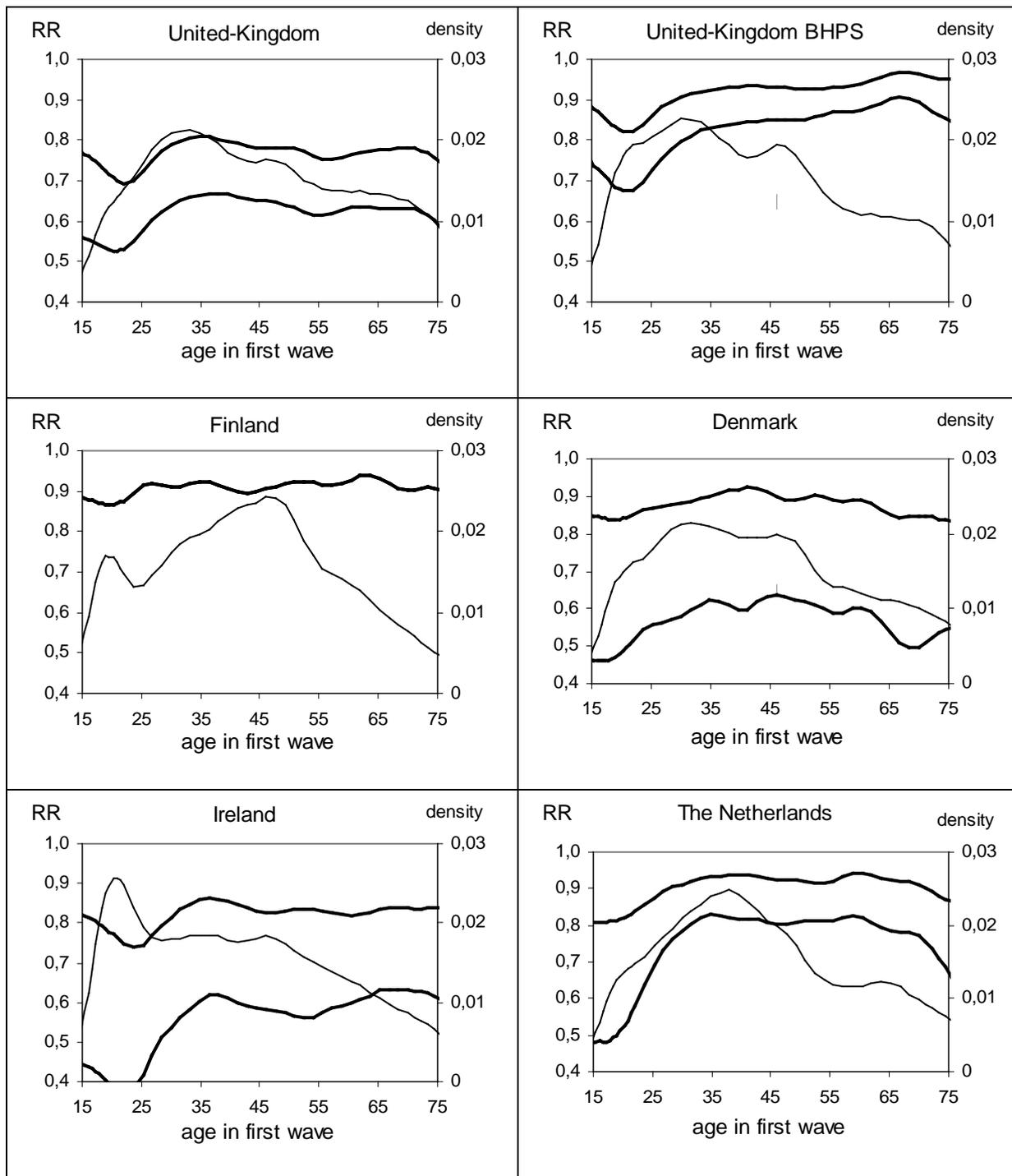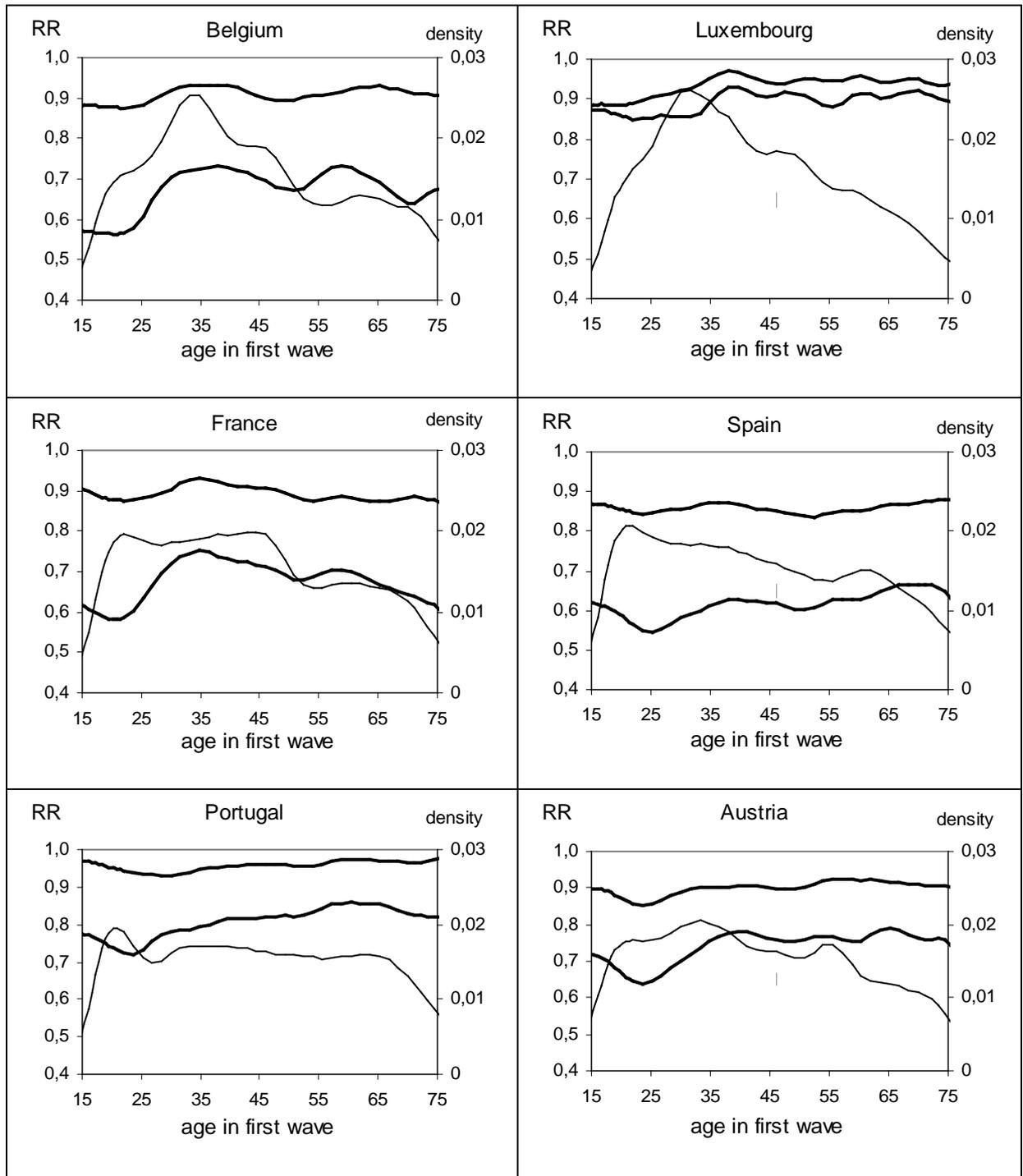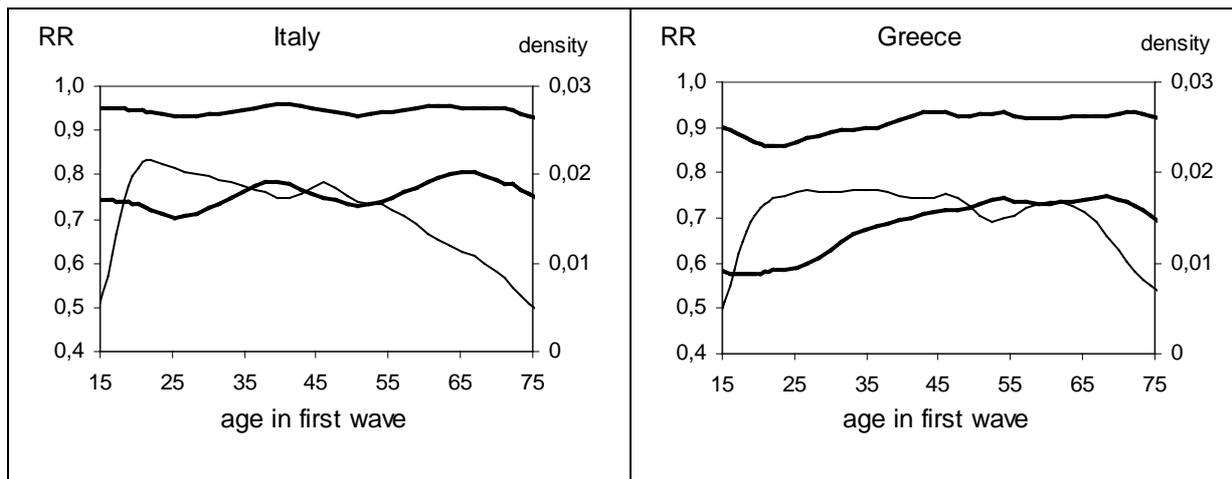
There is no clear relation visible that holds for all countries. But for most of the countries we find that the response rates for younger persons is somewhat smaller compared to the response rates for ages between 30 and 64. For the people aged above 64 we find no clear tendency, also for some countries the older persons might have a slightly lower response rate.

Using the same method we look next at the OECD-income. Again within the figure for one country we give the income distribution (kernel density estimates) in wave 1 as well as two response rates. The response rate in wave two and the overall response rate. To ease comparability we express all local currencies in Euro.

Fig. 8: OECD-income and response rates across countries

16
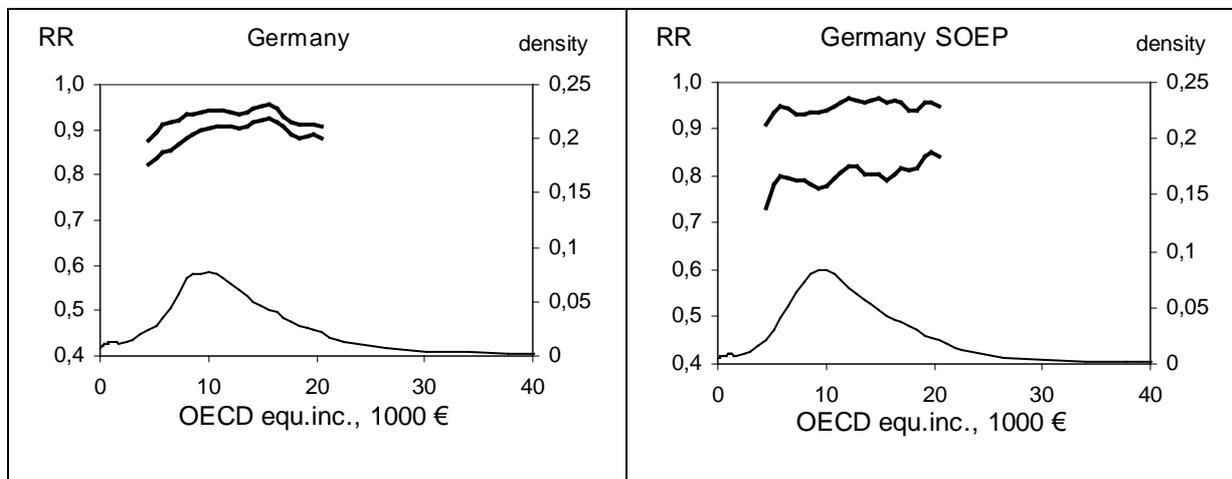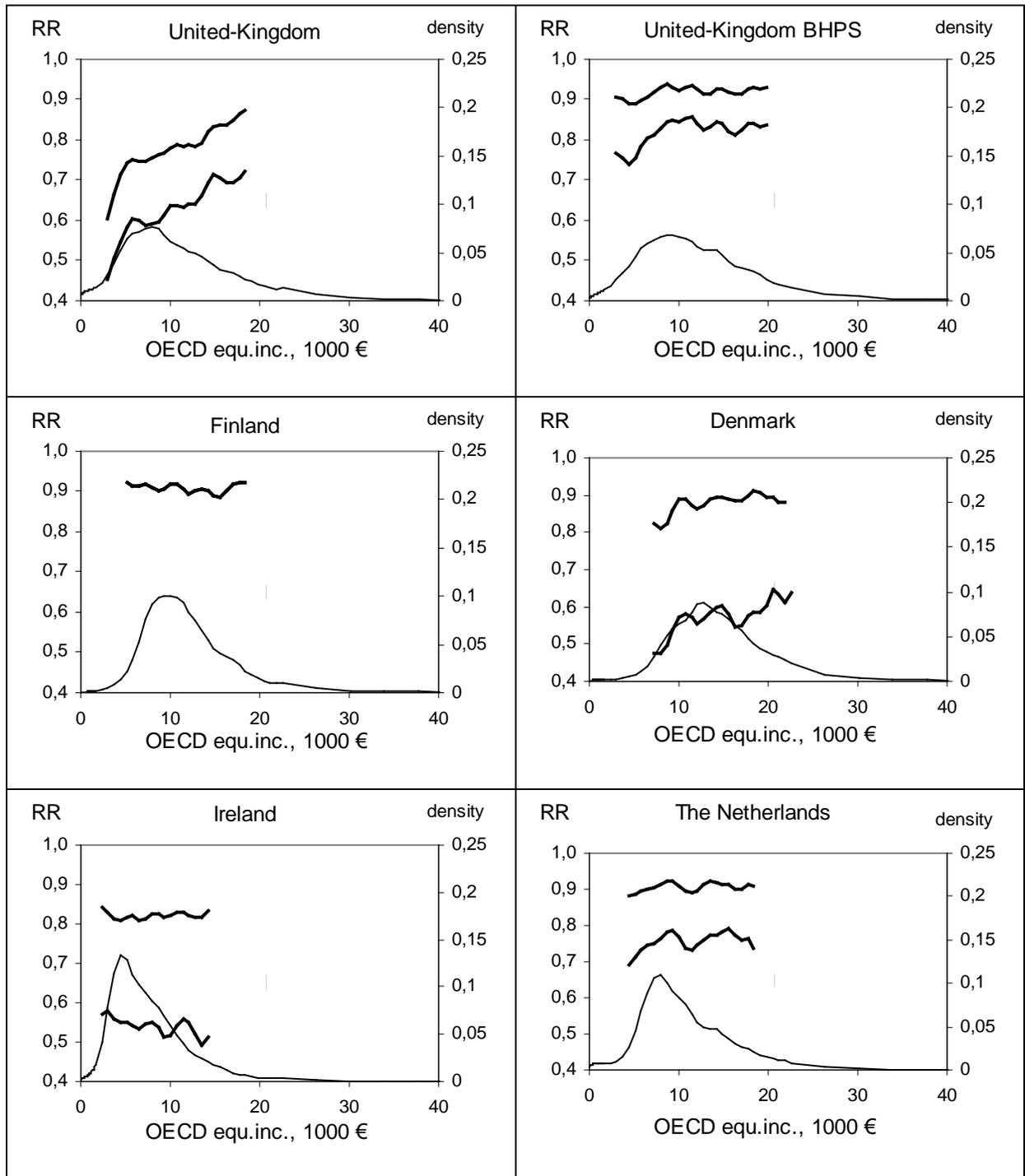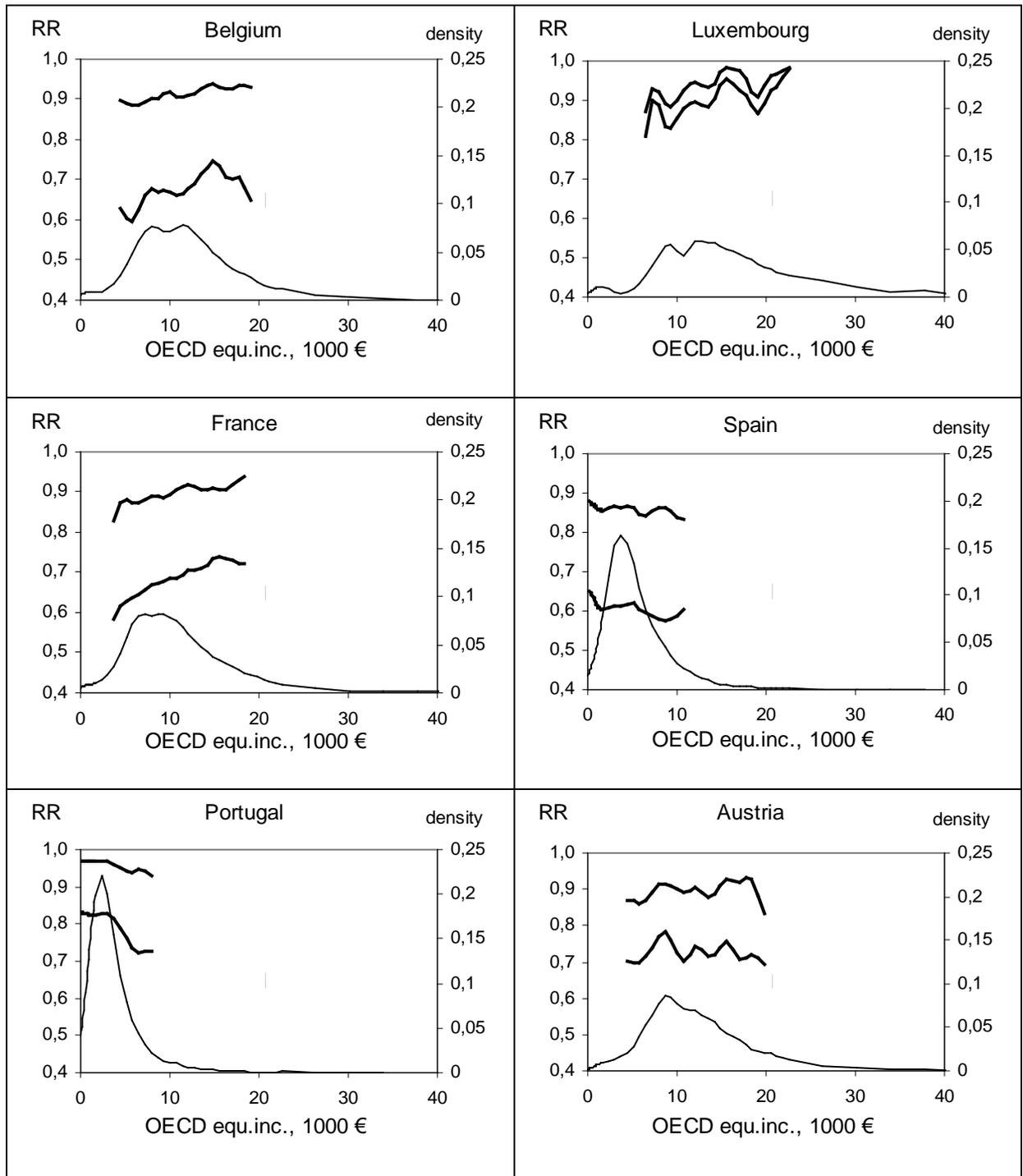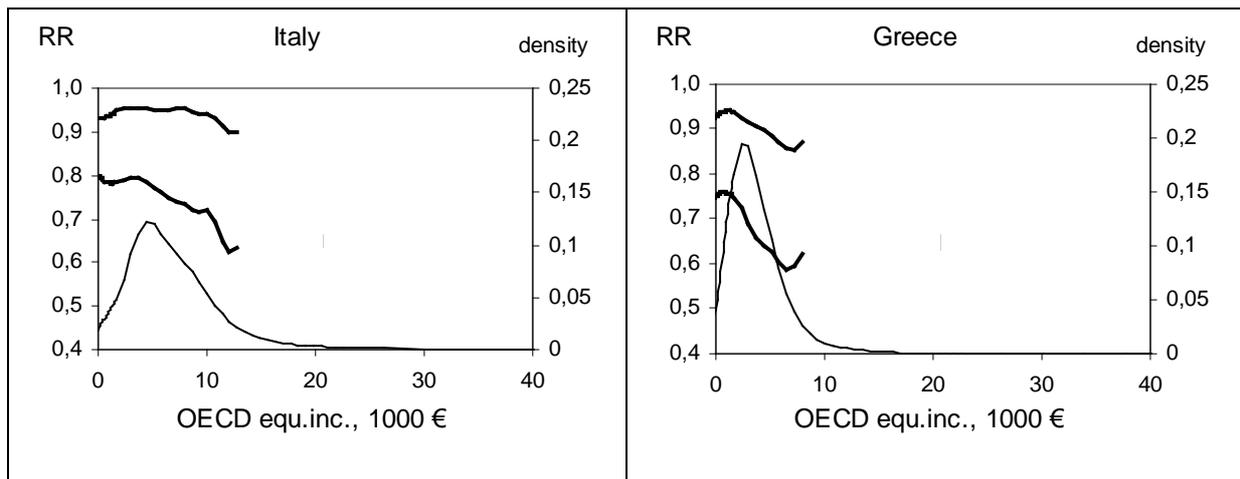
When comparing the shape of the curve representing the response rates across countries, it is evident that there is no identical relation present in all countries. While for some countries there is a negative relation between income and response rate visible, e.g. in Portugal, Italy and Greece, we find the opposite in several countries. Especially for the UK, Belgium, Luxembourg and France there is a clear positive dependence between income and response rate visible.

3.6. Comparing response rate for subsamples

Whether panel attrition causes problems in empirical analysis, e.g. biases in regression analysis, depends on the attrition process. To find out whether several subsamples of the panel are affected by attrition in a similar manner, we compare response rates according to variables that are prominent in socio-economic empirical work. The following figure contains the response rates for males and females within and across countries. For each country the figure contains three different kinds of information: the frequency distribution in wave 1, the response rate in wave two (open dots) and the rate of always responding persons (overall response rate in the following) which is, albeit not exactly, the product of the four wave to wave response rates (filled dots). When comparing the response rates for all waves it has to be kept in mind, that the number of waves differs for countries, although in most countries we have five waves. The comparison of the two different response rates allows to assess the short and long term attrition effects for different subsamples.

Fig. 9: Response rates by sex across countries
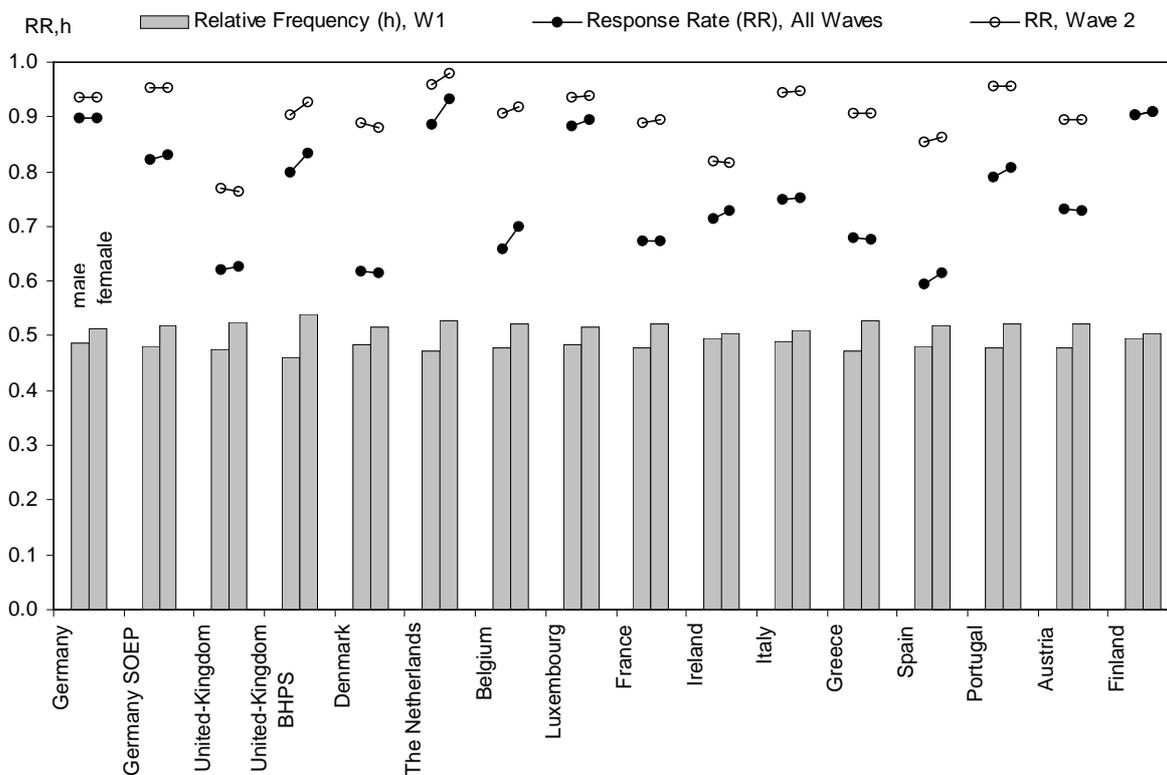
Fig. 10: Response rates and change of interviewer

20
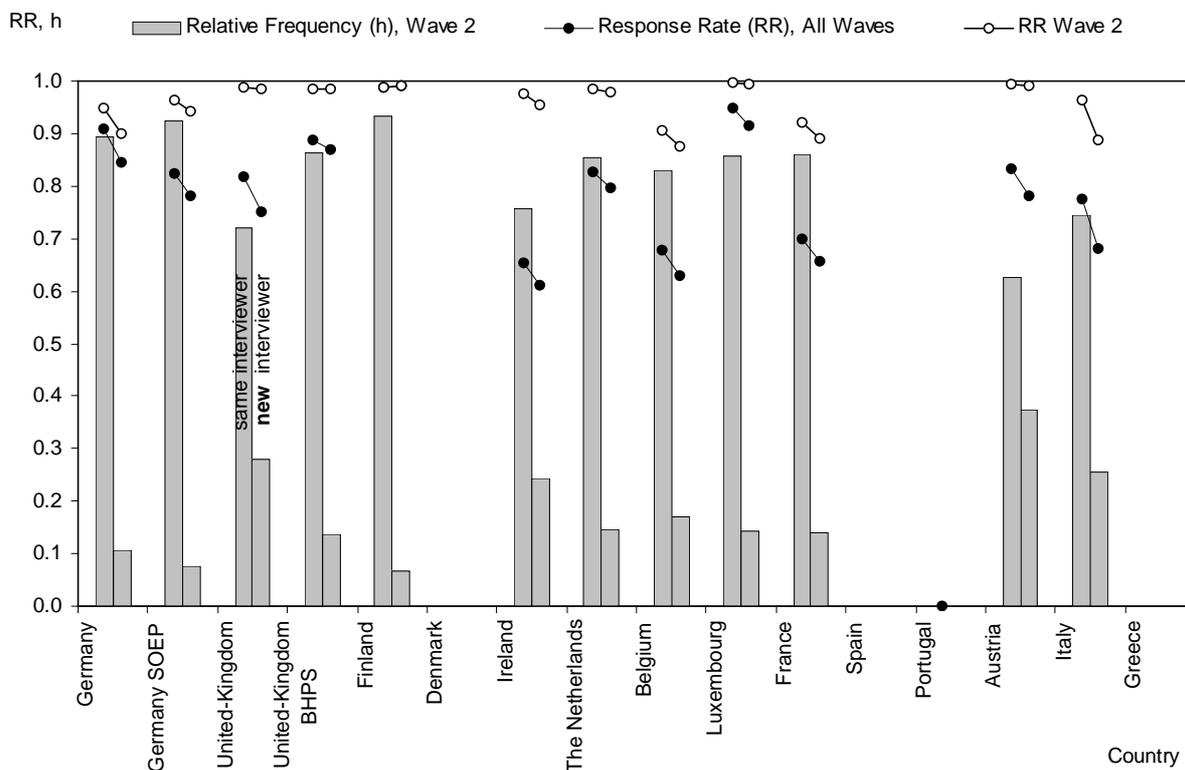
For most countries we find that the response rate for females slightly exceeds the response rates for males. While in some countries there is almost no difference visible, Denmark, Greece and Austria are the only countries showing higher response rates for males. For most countries the differences in the response rates for females and males increase over time as should be expected if female respond rates are slightly higher in all single waves.

The change of an interviewer is known to have the effect of considerably increasing the risk of attrition in panel surveys.[10] In figure 7 we graphically depict the effect of a change of interviewer on the response rate in different countries. As can be seen, the User Data Base of the ECHP does not contain information about interviewer change for Denmark, Spain and Portugal.[11] For the remaining countries, except Greece, there is a strong negative effect of interviewer change on the response rate visible. Comparing the effect of an interviewer change in wave 2 on the response rate in the same wave and on the overall response rate for all waves, we find that the difference is more pronounced in the overall response rate for most of the countries. Hence beside the immediate effect of an interviewer change, the future response rate is negatively effected as well.

Fig. 11: Response rates and move of a household
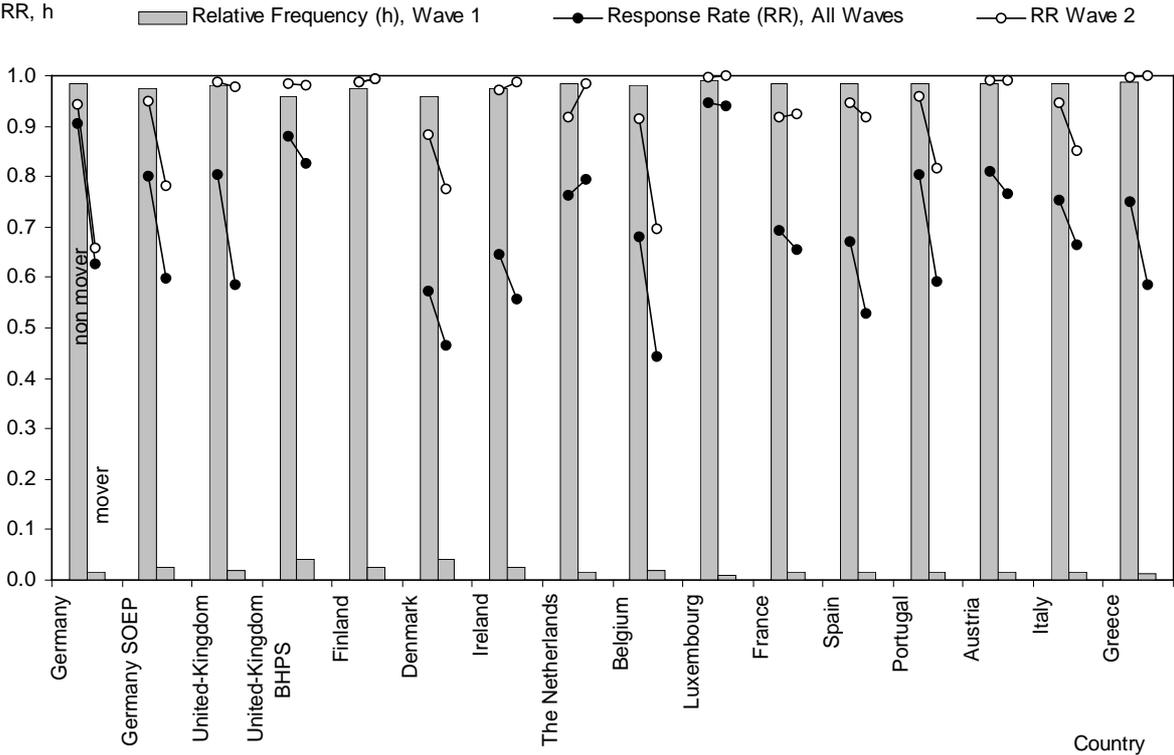
Like the effect of interviewer change, the move of a household was found in other survey to lower the response rate considerably, too. In Figure 11 we compare the response rates for persons in households which moved in wave two and find strong evidence for a strong

---

[10]   See e.g. Rendtel (1995).

[11]   The variable interviwer change for Germany is not contained in the UDB. The corresponding information was delievered by the German Statistical Office.

negative effect on the response rate in almost all countries. Remarkably is the strong long term effect of a move on the attrition rate. The overall response rate differs to a greater degree for movers and non movers than the immediate effect.

The difference in the response rates is strongest in the German-ECHP, German-SOEP, UK-ECHP, Belgium and Portugal. Only in Finland, The Netherlands, Luxembourg, Finland and France moving has apparently no negative effect on the response behaviour.

Comparing the response behaviour for individuals with different marital status, we find for most countries that the response probability for married people exceeds the response rate of people in a different marital status. Especially separated people exhibit lower response rates compared to married people. The effect of being widowed or never married has no uniform effect across all countries. Again we find that the effect of different marital status in wave 1 has increased effects on the overall response rate compared to the effect on wave 2 response behaviour only.

Fig. 12: Response rates and marital status

Marital status in wave 1: 1 Married; 2 Separated; 3 Divorced; 4 Widowed; 5 Never married.
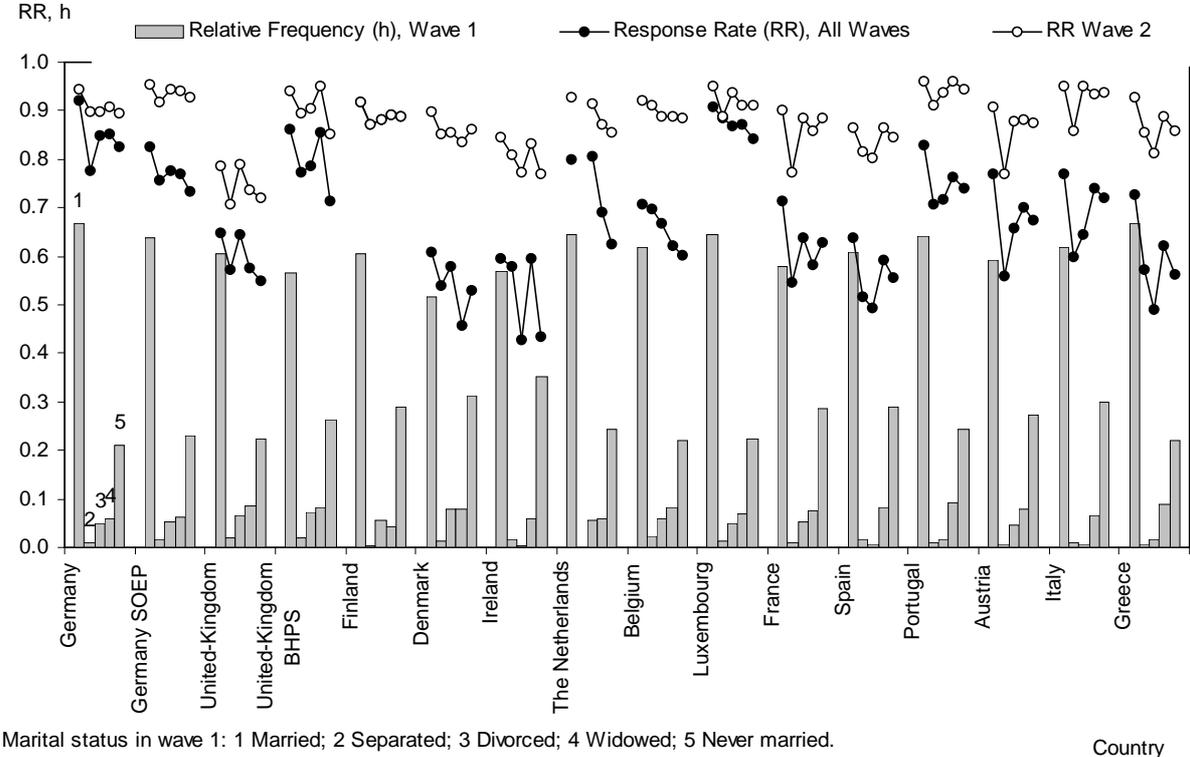
Turning to the response rate for persons according to their main source of personal income, we find no clear pattern across all countries. In the northern countries people earning wages or salaries tend to have slightly higher response rates compared to persons having other main sources of personal income. But the opposite is true for southern countries like Spain, Portugal and Greece. Again the differences in the response rate over all waves are exaggerated compared to the differences in wave 2 only.

Fig. 13: Response rates and main source of personal income
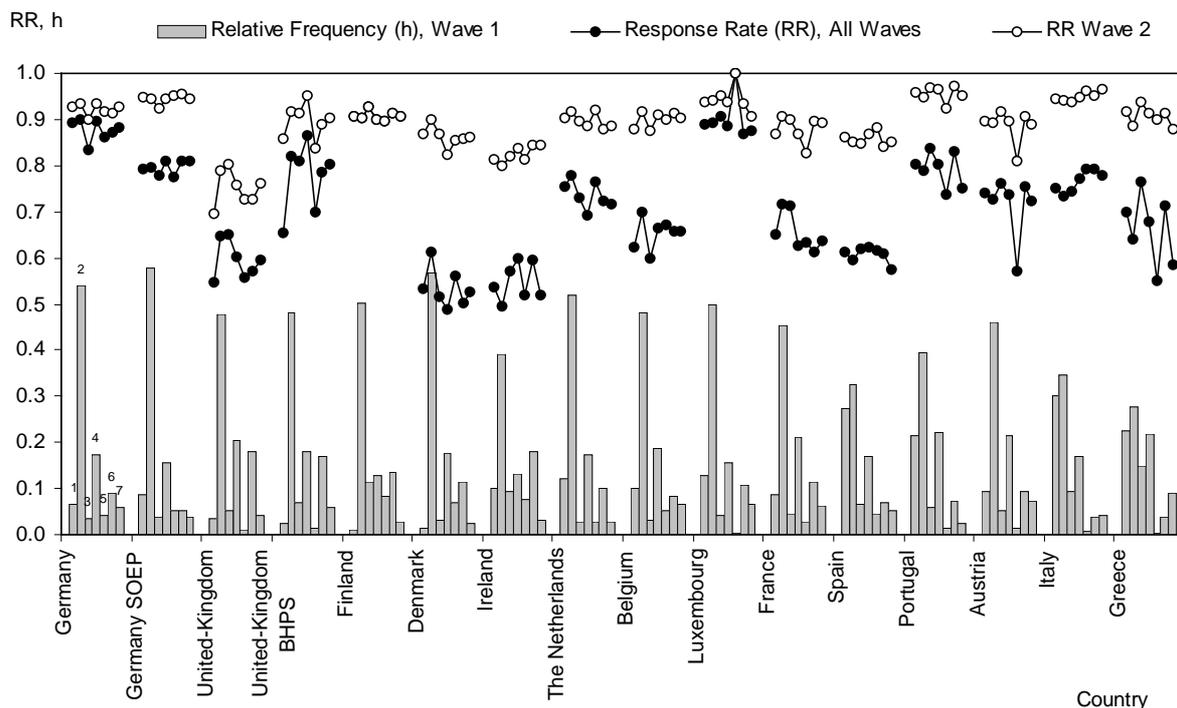
Main source of personal income: 1 Person has no income from any source; 2 Wages and salaries; 3 Income from self-employment or farming; 4 Pensions; 5 Unemployment / redundancy benefits; 6 Any other social benefits or grants; 7 Private income.

A very interesting pattern is found when comparing the response rates according to the highest level of education. In seven surveys under consideration, German ECHP, German SOEP, UK ECHP, Denmark, the Netherlands, Belgium and France we find a strong positive correlation between education and response rates. The higher the level of education, the higher the response rate. In the UK BHPS, Finland, Ireland and Luxembourg there is no clear relation between the level of education and the response rate. In five countries, except Austria, southern countries, we find the opposite correlation that was found for the northern countries. Because the level of education is quite time invariant, it could have been expected that the differences are more pronounced when comparing the response rate over all waves .

Fig. 14: Response rates and level of education

Highest level of education: 1 Recognised third level education (ISCED 5-7); 2 Second stage of secondary level education (ISCED 3); 3 Less than second stage of secondary education (ISCED 0-2).
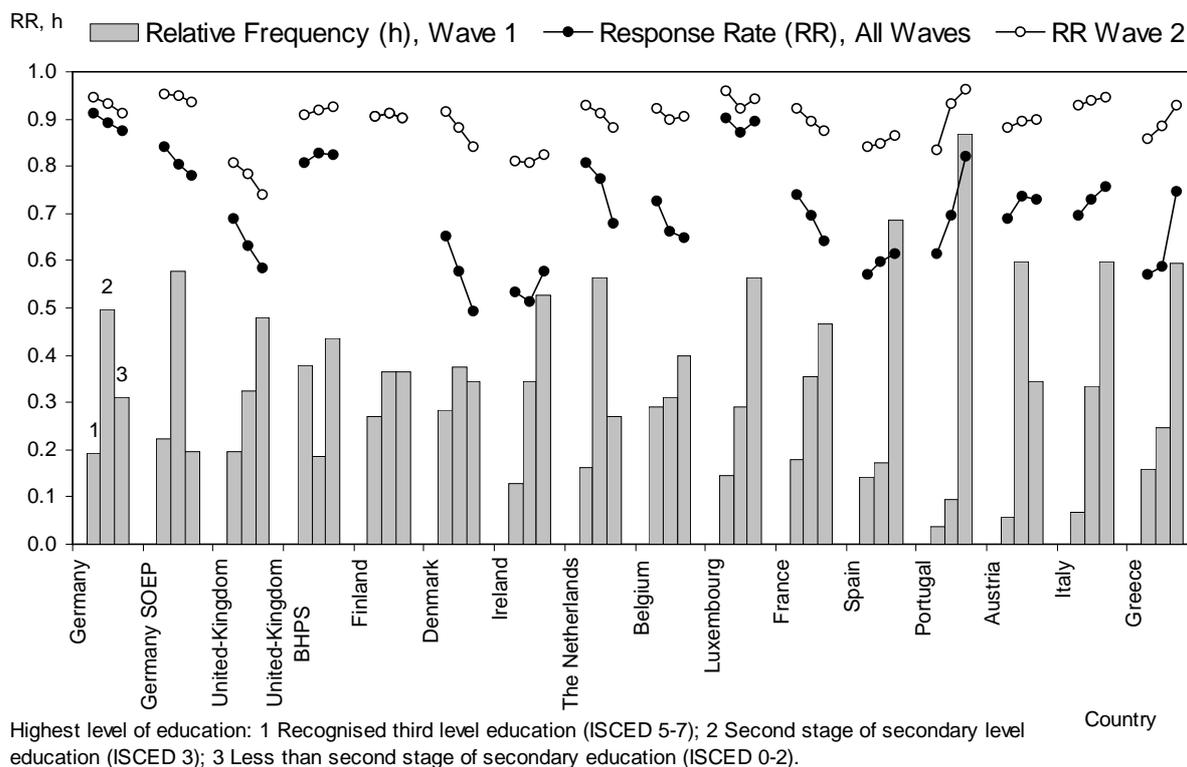
Finally we compare response rates according to the main activity status.

Fig. 15: Response rates and main activity status

Main activity status (self defined, regrouped): 1 normally working (15+ hours / week); 2 unemployed; 3 inactive.
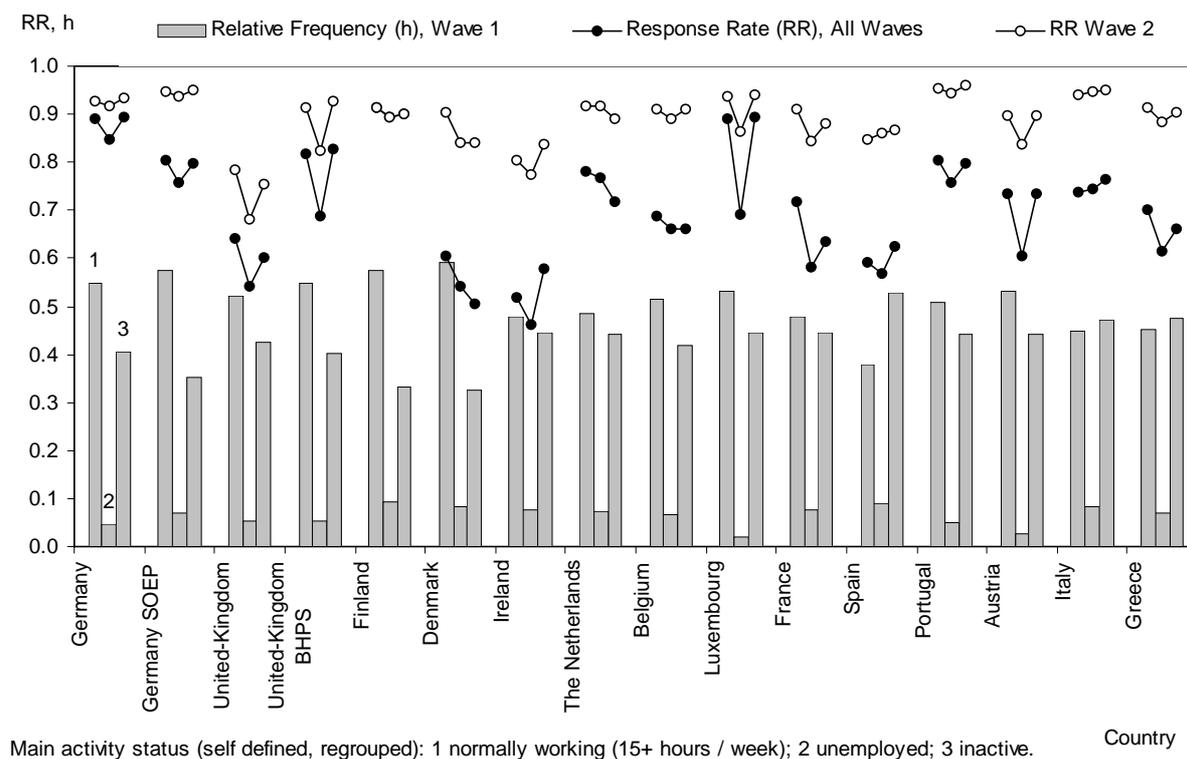
For the majority of countries, with the exception of Finland, Denmark, the Netherlands, Belgium and Italy, we find a similar pattern of response rates. Unemployed persons tend to have considerably lower response rates. This effect, despite the fact that unemployment is not hopefully not time-invariant, the differences are more evident when comparing the response rate over all waves.

To conclude the comparison of response rates for subgroups, we find some variables to be strongly correlated with nonresponse in this bivariate analysis. The strongest negative effects were found for the field variable change of interviewer and for persons moving during the year of the intended interview. We also found divorce and unemployment considerably decreasing the response rate. With considerable differences across countries, following a north-south pattern, we found higher response rates for highly educated persons in northern countries, while in southern countries the opposite tendency was found. When turning to model the response probability of individuals, this bivariate findings shall be used as a guide for a parsimonious logit model in the following section.

4. Modelling the response probability

In this section we estimate response probabilities using a logit model.

We use the following variables in the logit analysis:

- OECD-income: income, as well as the income raised to the power of two and three
- age: we allow for three different categories <30, 30-64, >65
- move of a household in the wave under analysis
- marital status: we allow for three categories, married, divorced or separated, widowed or never married
- sex
- highest level of education: we consider three categories, first, second, third level
- main activity status: we allow for three categories, employed, unemployed, inactive
- interviewer change

As the reference category we consider an employed male, aged between 31 and 64, married, with second level education, who neither moved nor experienced and interviewer change in the period under analysis.[12]

In the following we analyse the determinants of the response behaviour at three different levels of aggregation. First we estimate a logistic response function for each country and each wave separately. This means that we allow the different covariates to exert a different influence on the response behaviour in each country as well as in each wave. In our second analysis we pool the data across the available waves within each country. While we allow for different intercepts in each wave through the use of wave dummies, we restrict the covariates to exert the same influence on the response behaviour across waves.

Finally we pool all data across countries and across waves. This restricts the covariates even further to an identical partial influence in all countries in all waves. By including country and

---

wave dummies we allow for a different overall response rate in each country and in each wave.

4.1. An attrition analysis by country and by wave

Table 3: Logit results by country and by wave

Country	Wave	n	i	i²	i³	<30	>64	HH has moved	separated or divorced	widowed or never married	Female	third level	less than second stage	unemployed	inactive	Interviewer change
Germany	1	8,965	+	−	+			−	−	−			−			−
	2	8,245	+	−	+			−	−	−						−
Germany SOEP	1	9,039	+	−	+		−	−							+	−
	2	8,480	+	−	+	−		−	−							
	3	8,371						−	−	−						−
	4	7,395			+			−		−	+					−
United-Kingdom	1	8,069			+	−				−		+				
	2	6,013									+					−
United-Kingdom BHPS	1	7,955	−	+	−						+					
	2	7,526	−	+							+					
	3	7,324			+						+					−
	4	7,054	+		+						+				+	
Finland	3	7,426	−	+	−				+	−	+					
Denmark	1	5,705		−				−	−			−		−		.
	2	4,948						−	−	−				−		.
	3	4,358						−								.
	4	3,895												−		.
Ireland	1	7,432	−	+	−				+	−						−
	2	7,046	+	−	+				−			+				−
	3	6,005	−			−			−						+	−
	4	5,343				−			−							−
The Netherlands	1	8,343			+	−					+		−			
	2	7,722		+		−			+		+	+	−			
	3	7,200	−	+	−	−			+		+	+		+	+	−
	4	6,672			+	−			+		+	+	−			
Belgium	1	2,665	−	+				−	−	−						−
	2	5,136			+	−		−				+				−
	3	4,663				−	−	−				+				−
	4	4,228	+	−	+			−						+		−
Luxembourg	1	1,806	+	−	+	−		+	−	−	+	+	−		+	−
	2	1,689	+		−	−	+	+	+	−	+	+	−	+	+	
France	1	11,405	+	−	+			−				+	−			−
	2	10,308	−	+	−	+	−		+		+	−	−	+	+	−
	3	9,305		+	−	−	+		+		+	+		+		−
	4	8,637	+	−	+				+	−	+					
Spain	1	16,000	−												+	.
	2	14,009	−	+	−		+		+							.
	3	12,331	+	−	+							−		+	+	.
	4	11,396					+			−					+	.
Portugal	1	10,966	−						−	−		−	+			.
	2	10,010					+		−	−	+					.
	3	9,507	−	+					−			−			−	.
	4	9,136		−	+	−			−		+		+	−		.
Austria	2	6,585						−	−		+					
	3	5,909	−	+	−	−			+					+		−
	4	5,387				−			+						+	
Italy	1	16,064	+	−	+		−	−	−	−						−
	2	15,118	−	+		−		−	−				−			−
	3	14,596	−	+	−		+	−	−							−
	4	13,257	−	+	−			−							+	−
Greece	1	11,154	−			−		+	+	−			−			.
	2	10,043		−	+	−	+	+	+	−	+		+	−	−	.
	3	9,263		−	+		+	+	−	−	−	+	+		+	.
	4	8,151	−	+	−							+	+			.

To ease the readability and to allow for an overview of the many different results of the logistic regressions, we reduced the information given in table 3.[13] The table contains for each country and each wave the number of observation (*n*) and the information whether the covariate was according to its *t*-value exerting a significant influence on the response behaviour. We note a significant positive (+) as well as a significant negative (-) influence, while we did not use any sign when the parameter was insignificant.[14] The set of variables included was identical in all countries and all waves with the exception that the variable indicating a change of the interviewer was not available in the ECHP-User Data Base for Denmark, Spain, Portugal and Greece.

Turning to the results we find that the OECD-income, which we included allowing for a cubic influence, is significant in most regressions. While often the linear, the quadratic as well as the cubic effects are significant, we have no identical result in all waves. The linear and the cubic influence are most often positive, the quadratic influence is of mixed sign.

Compared to the reference category (30-64) in about half of the regressions younger persons tend to respond less, while in several regressions there is a significant higher response for elderly people.

Rather unexpectedly we find a very mixed picture for the influence of a move of the household on the response behaviour. In Germany (both surveys), Denmark, Belgium, Portugal and Italy we find a significant lower response probability for moving households. The opposite is true for the Netherlands, Luxembourg, France and Greece.

The marital status was reflected using three different categories. Compared to the reference category (married) we find that separated or divorced persons have no unambiguous tendency in their response behaviour. Widowed or never married persons show in about half of the regressions a significant lower response probability while in no regression we find an increase in the response probability.[15]

In about half of the countries under review we find a significant higher response rate for females, while a significant lower response probability is only found for wave three in Greece.

For the level of education we find a significant influence in some countries only. People who gained the highest level of education tend to respond with significant higher probability in the Netherlands, Belgium, Luxembourg and Greece.[16] Persons with less than second stage education tend in Denmark, the Netherlands and Luxembourg to respond with lower probability.[17]

Contrary to our descriptive findings in the bivariate case we find that in several regressions unemployed as well as inactive persons have an increased response probability. But in about 60% of the regressions for countries by wave there is no significant influence found for the main activity status.

---

[13] The detailed results are given in the appendix.

[14] The logit model we estimate includes no variance component for individuals, hence implicitly we assume independence of subsequent observations for individuals.

[15] The finding of higher response probabilities for married persons corresponds to the findings of Lillard and Panis (1998) for PSID.

[16] This tendency was also found in the PSID, see e.g. Lillard and Panis (1998).

[17] Fitzgerald/Gottschalk/Moffitt (1998) found the same pattern of decreasing risk of attrition with higher educational levels in the PSID.

For all countries the variable indicating the change of the interviewer is available in the User Data Base, we find a significant increase in the risk of attrition for persons who experience a new interviewer. This result confirms strongly the findings of Rendtel (1995) for the German SOEP.

When comparing the signs depicted in the table for one country across years we find that most often the influence is the same across waves. This indicates that the pooling of the data across waves within countries is much more in accordance with the data then pooling across countries where often opposing influences of the covariates were found.

4.2. An attrition analysis by country pooled across waves

Table 4: Logit results by country pooled across waves

Country	n	Income (i)			Age		HH has moved	Marital Status		Female	Highest level of educatio		Main activity status		Interviewer change
		i	i^2	i^3	<30	>64		separated or devorced	widowed or never married		third level	less than second stage	unemployed	inactive	
Germany	17,210	+	−	+				−	−			−			−
Germany SOEP	33,285	+			−	−		−	−		+			+	−
United-Kingdom	14,082		−	+					−		+			+	−
United-Kingdom BHPS	29,859	+		+						+				+	
Finland	7,426	−	+	−				+	−	+					
Denmark	18,906			−				−	−		+	−	−		.
Ireland	25,826	−	+		−	+	−		−		+	+			−
The Netherlands	29,937		+		−		+		−	+	+	−			−
Belgium	16,692			+	−			−		+	+				−
Luxembourg	3,495	+	−	+	−	+	+	+	−	+	+	−	−	+	−
France	39,655	+	−	+	−						+		−		−
Spain	53,736					+			−				+	+	.
Portugal	39,619	−					−		−	+	−	+	−	−	.
Austria	17,881				−										−
Italy	59,035	−	+	−				−	−						−
Greece	38,611	−			−				−		+				.

In the analysis by country where we pooled the data across all available waves, we find a very mixed result for the influence of income, using a polynomial of order three, across countries.[18] While only in Denmark, Spain and Austria there is no significant influence for any of the three income terms present, the sign of the parameter vary across countries. We find the linear income term to be four times significant negative and also four times significant positive. Just the same holds for the quadratic term. For the cubic term we have varying signs across countries, too.

For seven out of the twelve surveys under consideration we find that younger persons aged below thirty tend to have significant lower response rates compared to the reference class. The response behaviour of the elderly people shows greater variation across countries. Only in six out of the twelve surveys older people have a significant different response behaviour than the

---

[18] The logit model we estimate includes fixed year effects.

reference group. In Germany (SOEP), Denmark and France persons aged above 64 have lower response probabilities, while in Ireland, Luxembourg and Spain older persons have significant higher response probabilities.

The move of a household has the expected significant partial influence on the response probability in seven surveys, while in the Netherlands and Luxembourg persons moving in the sample period even tend to have increased response probabilities. This result clearly contradicts the intuition.

Concerning the marital status we find widowed or never married persons to have significant lower response probabilities compared to married persons. The effect of living separated or being divorced has no uniform effect across countries. The effect is for some countries significant only and of varying sign.

In no country we find a significant negative partial effect for females, but in six countries females tend to have a significant higher response probability compared to males.

For education the effect resembles even more across countries. Persons with third level education tend to have a partially higher response rate compared to persons with second level education. The only exception is Portugal, here persons having the highest level of education have a partially lower response probability. For the category less than second level education the evidence differs across countries.

The main activity status influences the response probability significantly in about half of the countries. While unemployed persons tend to have lower response probabilities, inactive persons tend to have higher response probabilities compared to employed persons. The exceptions are Spain, here unemployed persons have a higher response probability and Portugal where we find lower response probabilities for inactive persons.

The variable indicating the change of interviewer is remorsefully not available in four surveys. With the exceptions, United Kingdom-BHPS and Finland, where we have no significant effect, the change of interviewer decreases significantly the response probability. This finding is in accordance with Rendtel (1995).

4.3. An attrition analysis pooled across countries and waves

Finally we estimate a most restrictive model for the complete data set pooled across countries as well as across waves. In the model country- as well as wave-dummies are included. Hence, we allow for different overall response probabilities in each country and for each wave. The set of covariates is restricted to have the same influence on the response probability in all countries in all waves.

Table 5: Logit results pooled across countries and waves

Country	n	Income (i)			Age		HH has moved	Marital Status		Female	Highest level of educatio		Main activity status		Interviewer change
		i	i²	i³	<30	>64		separated or devorced	widowed or never married		third level	less th. 2nd st. of sec. educ.	unemployed	inactive	
Countries with available information on interviewer change	294,383	−		−	−		−	−	−	+	+		+	+	−
Countries with missing information on interviewer change	150,872				−		−	−	−	+					.
All countries *)	445,255	−			−		−	−	−	+	+			+	−

*) The mean of the variable interviewer change was used for countries with no information for this variable available.

The table contains the summarized information from three logit models. The first contains all countries where the information of a change of interviewer is available in the ECHP-User Data Base. We find a significant negative effect of an interviewer change on the response rate in this model. When pooling the data for countries having no information on interviewer change (Denmark, Spain, Portugal, Greece) most of the remaining variables have identical signs. Young persons, persons moving in the sample period and separated or divorced persons as well as widowed or never married persons have significant lower response rates. Females in both subpopulations have a higher probability to respond.

Table 6 contains the estimates of the logistic regression and some model information. We find income to be significant in all terms of the cubic polynomial, but the variable has no extraordinary explanatory power when considering the rather low t-values. Younger persons (age below 30) tend to have a significant lower response probability (the odds ratio is decreased by about 22%) while elderly persons have no significant different response rate than the reference category of persons aged between 30 and 64 years.

According to the t-values the move of a person as well as the change of the interviewer have the most strongest negative effect on the response probability. Also the numerical impact on the odds ratio is substantial: A move of a household reduces the odds ratio by a factor of $\exp(-1.092) = 0.33$. A change of the interviewer results in a decrease by a factor of $\exp(-0.629) = 0.53$.

Compared to married persons, persons living separated or persons who are widowed or have never been married tend to have lower response probabilities.

Females as well as highly educated persons tend to respond more, while persons with less than second stage education show partially no significant different response behaviour than persons having a second stage education.

Considering the different categories of the main activity status, we find that unemployed as well as inactive persons tend to have a lower tendency to respond as the reference category of employed persons. However, the the estimated effects are small.

Table 6: Detailed logit results pooled across countries and waves

	Countries with available information on interviewer	Countries with missing information on interviewer	All countries
Intercept	2.9945	6.2026	6.3708
	(78.8)	(48.93)	(53.8)
Income	-0.0060	-0.0050	-0.0047
	(-2.84)	(-0.85)	(-2.63)
Income2	3.500E-05	1.042E-04	2.965E-05
	(1.64)	(0.73)	(1.75)
Income3	4.850E-08	-5.869E-07	-3.098E-08
	(1.73)	(-0.83)	(-1.15)
Age lower 30	-0.2189	-0.1679	-0.2011
	(-7.7)	(-4.06)	(-8.6)
Age higher 64	-0.0398	0.0589	-0.0069
	(-1.22)	(1.28)	(-0.26)
HH has moved	-1.0919	-0.8947	-1.0243
	(-23.67)	(-12.99)	(-26.77)
Mar. stat. separated	-0.1810	-0.2487	-0.2074
or devorced	(-3.85)	(-3.97)	(-5.53)
Mar. stat. - widowed	-0.1895	-0.2181	-0.2016
or never married	(-7.65)	(-6.24)	(-9.99)
Female	0.0989	0.0879	0.0945
	(4.93)	(3.03)	(5.73)
Education - third level	0.2135	0.0795	0.1646
	(6.81)	(1.77)	(6.46)
less than 2nd stage	-0.0332	-0.0067	-0.0232
of secondary	(-1.49)	(-0.18)	(-1.22)
Main activity status -	0.0814	-0.0978	0.0218
unemployed	(2.09)	(-1.84)	(0.7)
Main activity status -	0.0888	0.0241	0.0701
inactive	(3.56)	(0.64)	(3.39)
Interviewer change	-0.6291		-0.6547
	(-29.12)		(-30.73)
CHI	8962.14	5826.69	14674.58
p	0.00	0.00	0.00
LRI	0.09	0.12	0.10
R2MZ	0.25	0.40	0.31
CHI,HL	188.77	194.60	75.60
p,HL	0.00	0.00	0.00
n	294,383	150,872	445,255

Turning to the model information, we get a rather mixed picture. The overall chi-square-test of model significance clearly rejects the hypothesis of no combined explanatory power of the model. McFadden's Likelihood-Ratio-Criterion has a rather low value of 0.09, the R^2 suggested by McKelvey and Zavoina (1975) has a slightly higher value 0.25. Both measures, which are defined between the range of 0 and 1, 1 in the case of perfect model fit, are indicating an unsatisfying model fit. This is supported by the Chi-Square test proposed by Hosmer and Lemeshow (1980), which indicates according to the p-value a deficient model fit, too.

Of course, the highly restrictive model is merely a way to condense the information, but the different findings in the analysis by country indicated already that the pooling across all

countries taking no account of the different country-characteristics is not indicated when looking for a satisfying explanation of the response behaviour.

5. Conclusion

The aim of the paper was to analyse the similarities and differences of the panel attrition in the ECHP. The fact that after five waves in some countries the response rate has dropped to about 50%, lead to concerns about the representativeness of the remaining participating persons.

The main result is that the extent and determinants of panel attrition show high variability across countries as well as for different waves within one country. This last finding is in accordance with Neukirch (2002) who found strong differences when comparing the response behaviour between different waves for the Finnish ECHP survey. Differences were also found when comparing the attrition behaviour across different surveys running parallel in the same countries, as was the case for Germany and Great Britain.

Nevertheless our analysis of the response behaviour of different subpopulations revealed, that the difference in response rates is strongest whether households moved in the sample period and whether the interviewer in the sample period changed. Compared to these two influences all other characteristics are of minor influence. Albeit there is the tendency that young persons have a slightly higher probability not to respond.

Strictly opposite results were found for the influence of age and education across countries, following a north-south pattern.

The unavailability of important information in the year of attrition precludes the analysis of the immediate effect of changes in personal circumstances, but we found evidence that such changes have a longer lasting influence on attrition behaviour.

With respect to harmonisation of surveys, which is a crucial point of the ECHP, it can be concluded that the high variation of attrition behaviour between the ECHP subsamples has the potential of introducing a new source of variation in the results. In order to compensate attrition and thus maintaining comparability it may be more adequate to use a weighing strategy that varies over the subsamples, i.e. to use a weighing strategy that is not harmonized.

Whether the attrition we analysed will seriously influence the results of empirical socio-economic research based on the ECHP is left as an open question for future research.

Reference

Fitzgerald, John/Gottschalk, Peter/Moffitt, Robert (1998), An Analysis of Sample Attrition in Panel Data. The Michigan Panel Study of Income Dynamics, The Journal of Human Resources, Vol. 33, pp. 251-299.

Hausman, Jerry A./Wise, David A. (1979), Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, Econometrica, Vol. 47, pp. 455-473.

Hosmer, D. W./Lemeshow, S. (1980), A Goodness-of-Fit Test for Multiple Logistic Regression Model, Communications in Statistics, A10, pp. 1043-1069.

Lillard, Lee A./Panis, Constantijn W.A. (1998), Panel Attrition from the Panel Study of Income Dynamics, The Journal of

Little, Roderick J.A./Rubin, Donald (1987), Statistical Analysis With Missing Data.

McKelvey, R./Zavoina, W. (1975), A Statistical Model for the Analysis of ordinal level dependent variables, Journal of Mathematical Sociology, Vol. 4, pp. 103120.

Neukirch, Th.: Nonignorable attrition and selectivity biases in the Finnish subsample of the ECHP, Working Paper, URL: www.destatis.de/chintex.

Nicoleti, Cheti/Peracchi, Franco (2002), A Cross-Country Comparison of Survey Nonparticipation in the ECHP, Working Paper.

Peracchi, Franco (2002), The European Community Houshold Panel: A Review, Empirical Economics, Vol. 27, pp. 63-90.

Rendtel, Ulrich (1995), Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität, Campus Verlag, Frankfurt/M., New York.

Rendtel, Ulrich (2002), Attrition in Household Panels: A Survey, Working Paper, URL: www.destatis.de/chintex.

Verbeek, Marno/Nijman, Theo (1995), Incomplete Panels and Selection Bias, in: Mátyás, L./Sevestre, P. (ed.) (1995): The Econometrics of Panel Data, Dordrecht.