CHINTEX

Working Paper #11

Work-package 6

Date: 17 April 2003

Andreas Behr

Egon Bellgardt

Ulrich Rendtel

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

The estimation of male earnings under panel attrition. A cross country comparison based on the European Community Household Panel

CHINTEX - The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel – Implications on Data Quality

Financed by the European Commission under contract number IST-1999-11101

The estimation of male earnings under panel attrition.
A cross country comparison based on the European Community Household Panel

by Andreas Behr, Egon Bellgardt, Ulrich Rendtel

Contents

Abstract

The aim of the paper is analyzing the effect of attrition in the European Community Household Panel (ECHP) on earnings equations. By splitting the completely observed sample according to the response behavior of the following wave, we assess empirically the bias of an un-weighted as well as an inverse probability weighted estimator.

Our findings lead us to conclude that the problem of attrition is no matter of great concern when estimating income equations of the Mincerian type based on the ECHP data. While we find in some cases smaller differences in the regression parameters due to attrition, the main findings seem rather unaffected by attrition. Concerning the question of correcting for possible attrition biases through inverse probability weighting, we conclude that the additional variance of the estimated response-probabilities outweighs possible gains from this correcting procedure and does not lead to an overall improvement compared to the un-weighted estimator.

1. Introduction

The aim of this empirical paper is to explore the effect of attrition on the estimation of earnings-equations for male employees across Europe. The income equation we use is of Mincerian type based on the human capital approach of income determination. The data base underlying our research is the European Community Household Panel (ECHP) giving unique opportunity for cross-country comparisons in the European Union on the level of individuals and households. The considerable extent of panel attrition in the ECHP is documented in detail by Behr, Bellgardt and Rendtel (2002). These findings indicate that there might be some concern of attrition influencing the results of empirical socioeconomic analysis. In this paper we take special emphasis on analyzing such possible biases caused by panel attrition. The analysis is carried out through a comparison of estimated earnings equations on full samples and samples of respondents only.

Besides analyzing the effect of attrition when using an un-weighted estimator, we also assess the performance of an inverse probability weighted estimator considered as correcting for attrition. The effects of attrition on these estimators will be analyzed by transferring the observed attrition behavior in the subsequent waves to the wave under consideration, which facilitates the comparison of the estimated income process for attriters and respondents as well as respondents only.

In the following section we give a brief description of the ECHP and the extent of attrition. Section 3 contains some theoretical foundation of the earnings equations and the empirical results ignoring possible biases caused by attrition. In section 4 we model the response probability and assess the biases which occur in the case of un-weighted as well as inverse probability weighted estimation. The unique situation of two parallel surveys in Germany and the United Kingdom allows the comparison of empirical results for identical earnings equations based on two different data sets for the same country and time. These analysis is carried out in section 5. Section 6 concludes.

2. The data base and panel attrition

2.1 The ECHP-UDB[1]

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".[2] The most attractive feature of the ECHP for research is its standardization.

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

1 The analysis is based on the December 2001 release of the ECHP-User Data Base (UDB).

2 Eurostat (1996), cited after Peracchi (2002), p. 64.

Sweden already ongoing surveys were used to create the national subsamples, 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 comparative analysis.

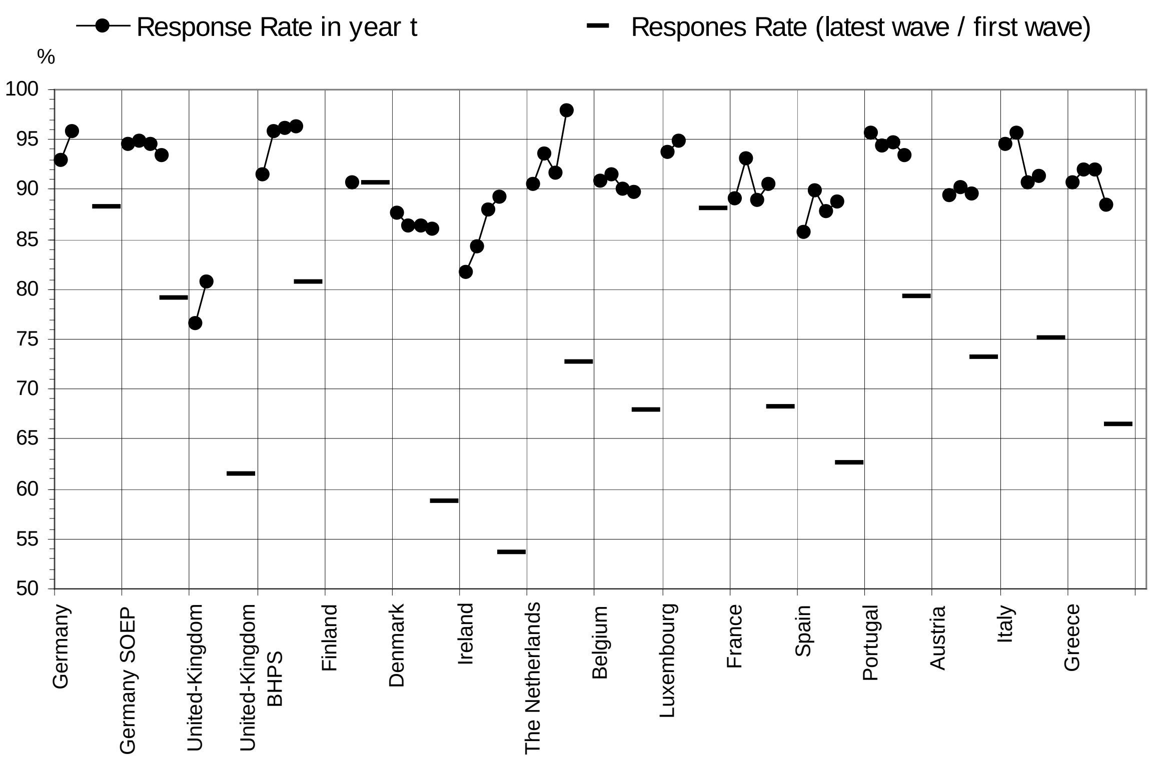
Because we assume effects of the number of waves being more important compared to effects of a given year, all data are ordered by country and wave. This means that data of wave 1 will include mainly data from 1994, but also from 1995 (Austria), 1996 (Finland). Because for Sweden only one wave is available (1997) what renders most of the analysis impossible, we do not include Sweden in the following analysis.

2.2. The participation in the ECHP

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). At the individual level the attrition in the ECHP is studied by Nicoletti/Peracci (2002) and Behr/Bellgardt/Rendtel (2002).

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.

Fig. 1: Response rates across countries for wave 2 to wave 5 and the overall response rate



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.

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.

The following figure contains the maturation of the ECHP, pooled over all countries.[3] 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, who are not monotone attriters

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 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*

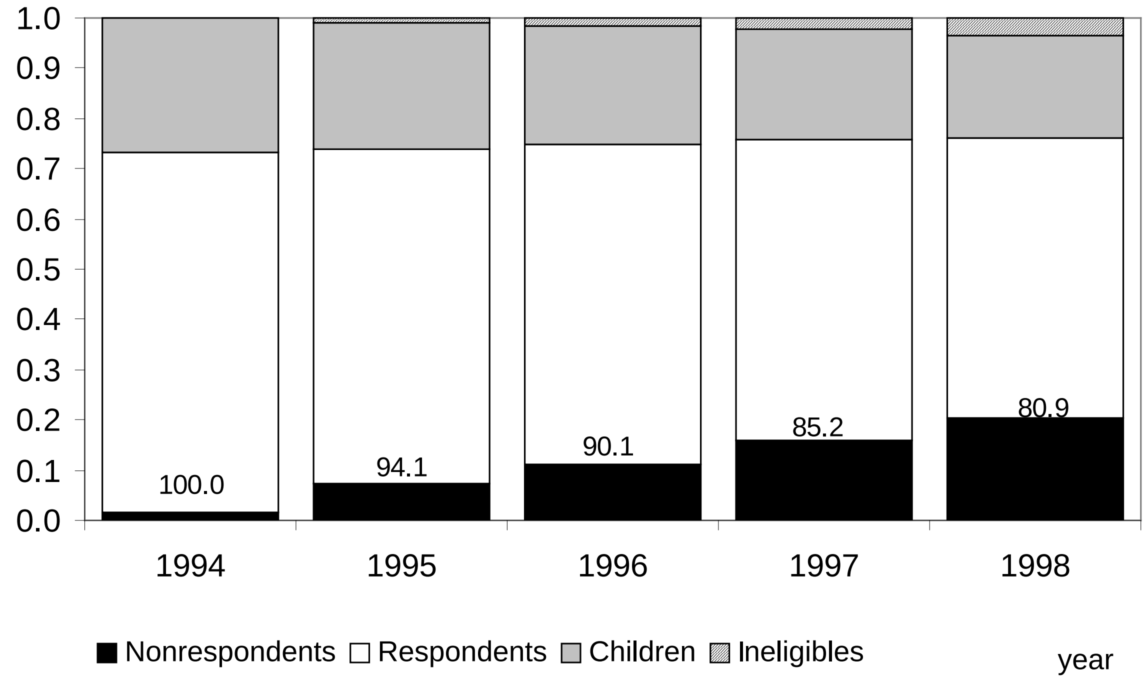
$I_{t(1)}$ person belonging to the initial group who is ineligible in wave *i*

X_t may be interpreted as a measure of panel stability.

In 1994 all eligible sample persons are considered. Figure 2 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.

[3] 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.

Fig. 2: 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 a steadily increasing share of nonrespondents, a small but growing share of ineligibles and steadily decreasing share of children due to reaching the age 16.

3. Earnings equations, theoretical considerations and empirical results ignoring attrition

The analysis of attrition is based on the estimation of earnings equations. In the context of panel attrition earnings equations are a standard application, used e.g. by Fitzgerald et al. (1998) for the PSID. Hence the analysis of earnings equations facilitates a comparative judgment of the empirical findings. In the following we discuss very briefly the two most important approaches to explain income inequalities and the distribution of income theoretically.

3.1. The human capital approach

There is a growing recognition of the importance of investment in people as an underlying principle in theoretical and empirical analysis of income distributions.[4] In the present paper we restrict our analysis to males and to earnings of labor in the context of the human capital approach.[5]

4 See e.g. the overview article by Mincer (1970), and Mincer (1958, 1997).

5 Historical functional approaches of the Ricardian type or Smith's compensatory principle to explain income differentials are not taken up in this analysis. First, according to the blurring of social class identification nowadays the variance of labor incomes is the dominant component of total income inequality. Second, there is overwhelming evidence that occupations in which work is more unpleasant and unstable command

The schooling model can be seen as the most simple form taking into consideration individual investment behavior as the main source of the heterogeneity in labor incomes. First variants of these models assume absence of environmental inequalities and attribute the variance of incomes to the length of training. In a competitive equilibrium, the distribution of earnings is such that the present value of future earnings, discounted at the market rate of interest are equalized at the time training begins.

Variation of earnings over the life cycle is an important source of income inequality. In general age profiles are found to slope upward over a large part of the life cycle.[6] This empirically well confirmed result is strongly supported by our results of section 3.3.

3.2. Signaling, screening and sorting

The observation that workers with higher levels of education and more work experience tend to have higher wages is usually explained within the framework of the human capital theory. This explanation states that the time spent in school and on the job increases worker's productivity directly and hence increases wages.

But one question discussed rather controversially is, whether inequalities in earnings are due mainly to schooling or inequalities in abilities. Since higher levels of education are generally positively correlated with several characteristics rewarded by employers like lower propensity to quit or to be absent and general healthiness, longer schooling simultaneously signals these characteristics.[7] Students expecting employers to favor longer schooling indicating positive characteristics in turn will choose the length of schooling to "signal" their ability and characteristics.[8]

The main conclusion is that the presence of signaling or screening might produce distortions that have to be taken into account, when estimating rates of return to schooling and to job experience.[9]

The theoretical considerations made evident that investments in human capital should be seen as the most important factor influencing the earnings across persons. The variables most often

---

lower, not higher wages. Nevertheless Smith's compensatory principle, earnings differential that tend to equalize net advantages and disadvantages of work, might be taken into account when explaining differences in labor incomes in occupations requiring similar training costs. See Mincer (1970).

6 Because of finite lifetimes later investment produce returns over a shorter period of time. If investments in human capital are profitable, their postponement reduces the present value of the net gains. As a consequence of human capital accumulation later investment is more costly because of higher foregone earnings. The only exception to these considerations can be seen in the case when productivity in learning grows as fast or even faster than productivity in earning.

7 See e.g. Weiss (1995).

8 If employers demand a certain level of schooling to "screen" their applicants hoping to find out about unobservable characteristics and abilities the combination of signaling and screening leads to a sorting of workers. In sorting models, a correlation of schooling with differences among workers, which were present before the choice of schooling is made, is assumed. Firms making inferences about workers productivities that vary according to schooling and characteristics by observing schooling choices and students respond to this inference process by choosing longer schooling. See Weiss (1995).

9 While there are several attempts to discriminate between these two competing theories, in this analysis we do not attempt to disentangle the income effect of measured education, but rather view the results as descriptive evidence of a cross-country comparison. But see e.g. Neumark/Taubman (1995) and Kroch/Sjoblom (1994) finding evidence in favour of the human capital model based on data for the U.S. and Groot/Oosterbeek (1994) based on data for the Netherlands.

used in these kinds of earnings equations are the amount of schooling, the highest educational level reached and after schooling job experience or job tenure.[10]

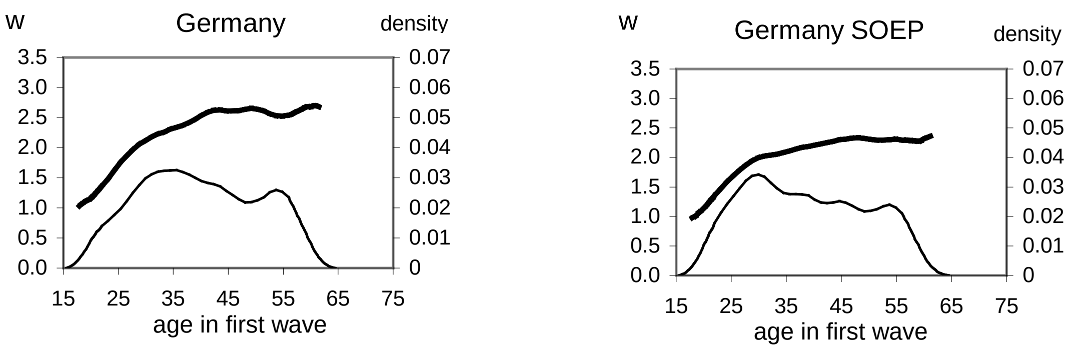
Recognizing the importance of institutionally determined inequalities of opportunity, a pragmatic statistical approach is followed in this analysis. In this spirit, demographic factors as sex, age and education have will be included in the multivariate regression analysis of individual earnings of active mails.[11]

Since we have no information of the exact job experience of the persons in the ECHP we use the age as indicating job experience. If all persons had left school after the same number of years of schooling for working continuously, age minus eighteen years would equal the job experience.[12] The highest educational level reached by persons is indicated in the ECHP-UDB by a variable having three different categories: less than second stage education, second stage education and recognized third level education. In the next section we start our empirical analysis making use of non-parametric methods to give descriptive evidence of the relation between wage and age as well as between age and educational levels. The regression approach to Mincerian earnings equations is followed in section 3.4.

3.3. Descriptive evidence

The following figures display simultaneously the age distribution within the national subsamples of the ECHP as well as the average earnings at each age given by a kernel regression (bandwidth 4.5 years).

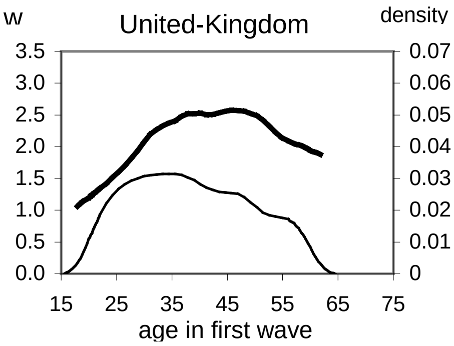
Fig. 3: Earnings-profiles across countries

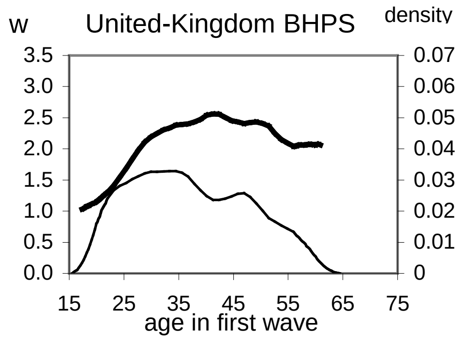


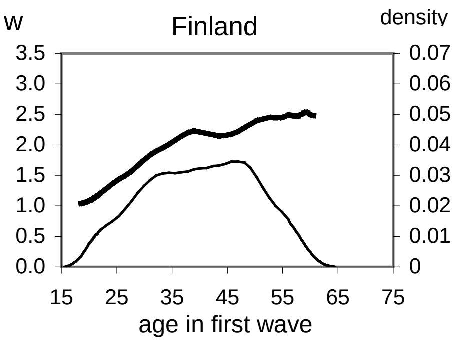
10 While after schooling job experience is the sum of time worked altogether, the job tenure measures the time the person is working within the same firm. Hence tenure might be closer related to firm-specific knowledge compared to job experience.

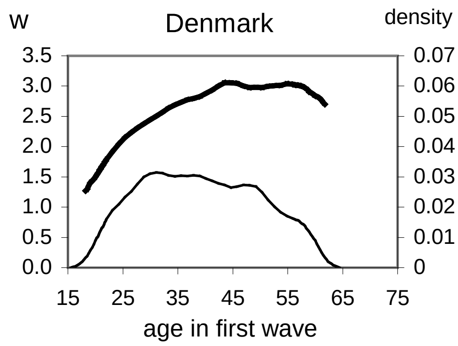
11 The specification in the empirical section leaves out the important information of post-school investment.

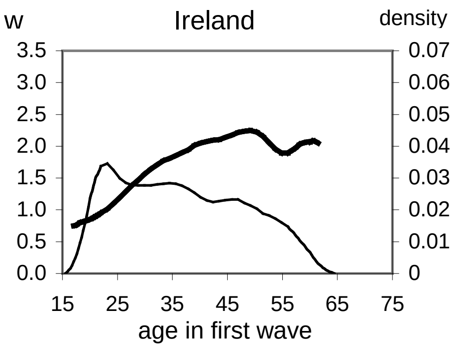
12 Since we include only active mails in our analysis, the years spent for child-bearing will not have to be taken into account.

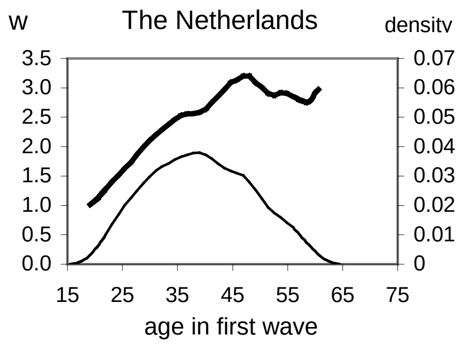

w
United-Kingdom
density
age in first wave

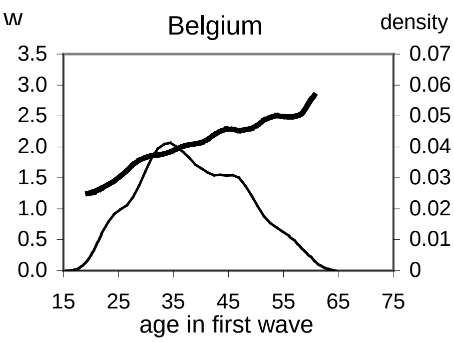

w
United-Kingdom BHPS
density
age in first wave

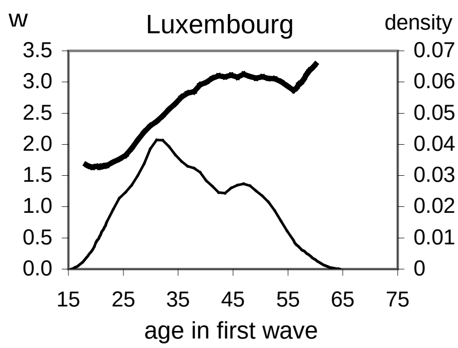

w
Finland
density
age in first wave

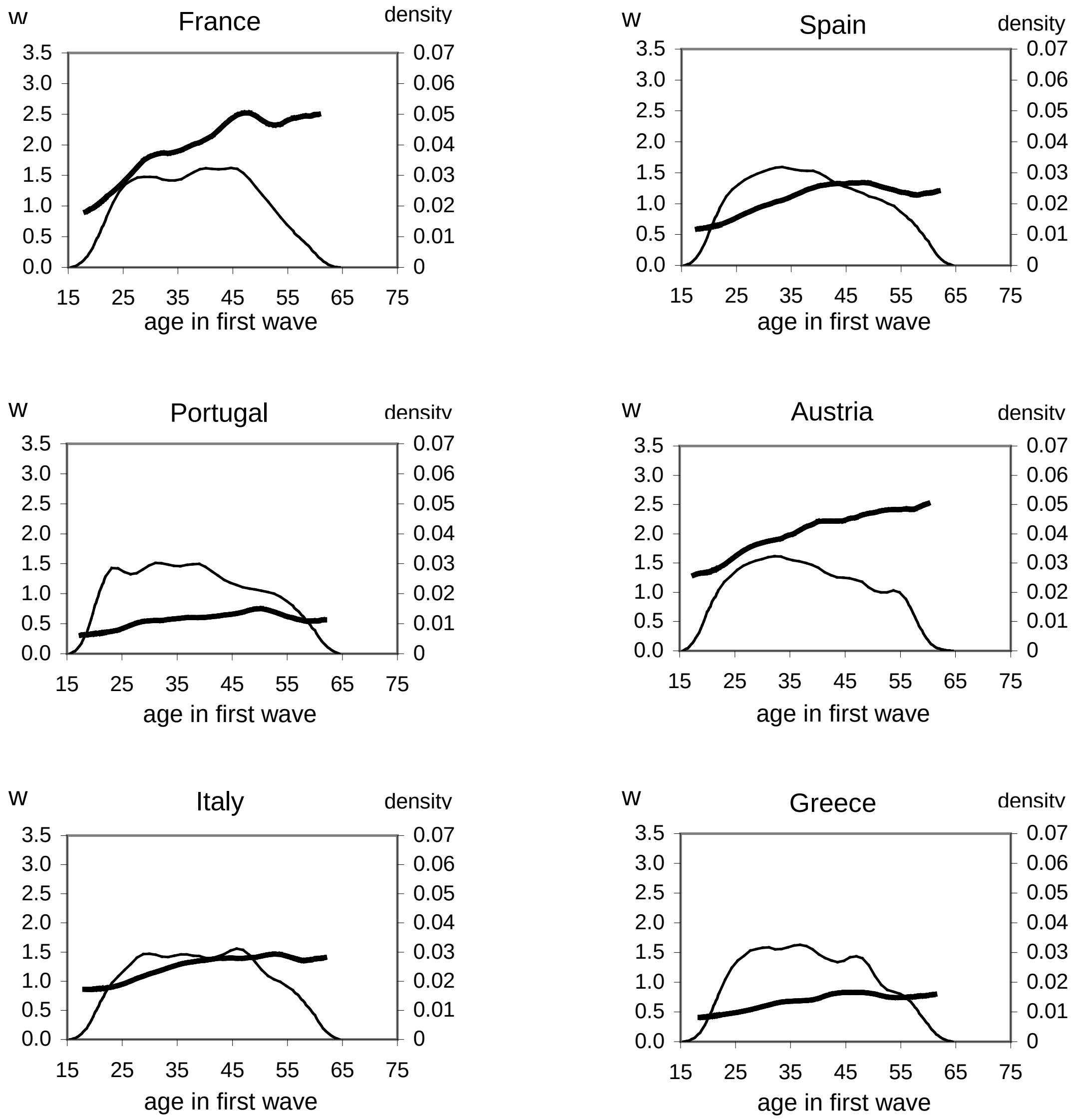

w
Denmark
density
age in first wave


w
Ireland
density
age in first wave


w
The Netherlands
density
age in first wave


w
Belgium
density
age in first wave


w
Luxembourg
density
age in first wave



Comparing the wage profiles for sixteen ECHP-samples we find considerable differences, in the level of earnings as well as the shape of the earnings profile. While we observe almost linear increasing profiles for Portugal, Italy and Greece, we find noticeable concave profiles in Germany, the UK and Luxembourg. In the Netherlands, there is a strong decline in the earnings profile at about 46 years.

In the figure 4 the frequency distribution of the three levels of education as well as the average earnings of the three categories is displayed.[13]

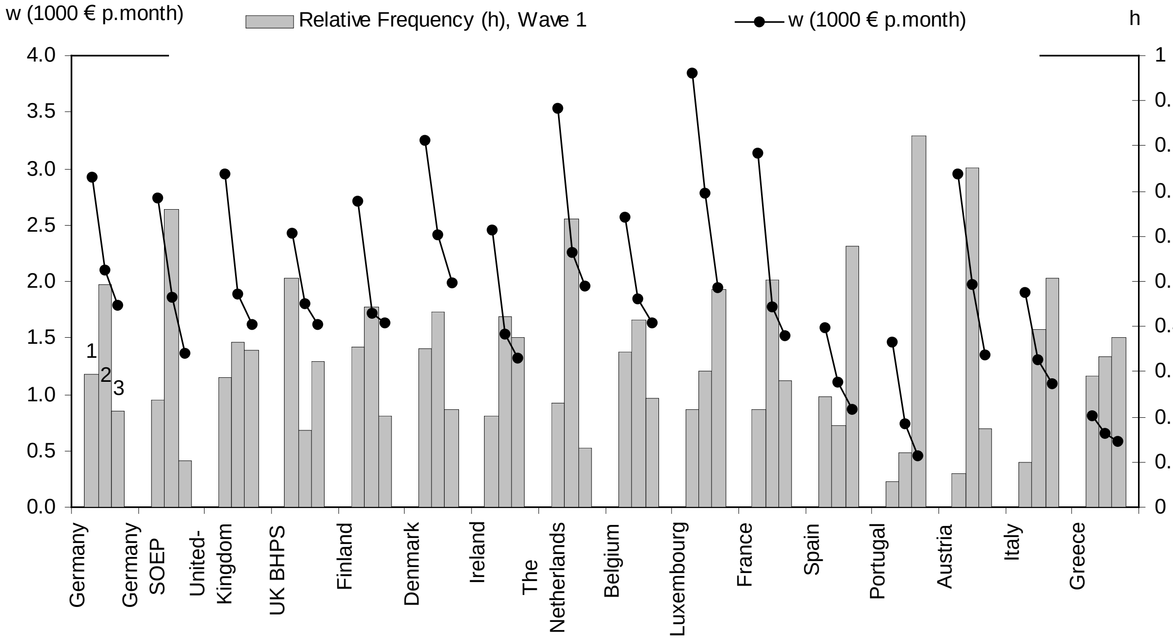
Comparing the frequency distributions of the three levels of education, we find that in the majority of countries the largest share of persons taking part in the survey obtained second stage education. The only survey containing most third level educated persons is the British Household Panel Survey. This result is quite different to the proportions we obtain when examining the ECHP-sample for the United Kingdom. Less than second stage education is the

13 We do not apply any weihting of the national samples to correct for unequal sampling probabilities. Hence, the distribution found in the samples will not necessarily mimic the distribution in the total population.

category with highest frequencies in the national subsampbles from countries of Southern Europe, Spain, Portugal, Italy and Greece.

Turning to the average wages for the three different levels of education, there is the uniformly observation that the average wage increases with the level of education obtained. The wage increase is much more pronounced comparing third level to second stage education, while the wage difference between persons having less than second stage education to persons with second stage education is rather small. The empirical result for Greece is unique in the way that only small wage differentials are observed.

Fig. 4: Earnings and level of education across countries



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).

3.4. Mincerian earnings equations

In this section we estimate cross-sectional log-earnings equations for males of the following type by country (*j*) and by wave (*t*):

$$\ln w_i = \beta_0 + \beta_1 \cdot age_i + \beta_2 \cdot age_i^2 + \beta_3 \cdot edu3_i + \beta_4 \cdot edu1_i + \beta_5 cumemp_i + \beta_6 married_i + u_i$$

Concerning education, we use second level education as the standard category and the dummy variables *edu*3 and *edu*1 to indicate third level education and less than second stage education respectively. The variable *cumemp* denotes the cumulated unemployment time in month and the dummy variable *married* indicates whether the person is married while all other marital stati (divorced, widowed, not married) are the base category.

To ease readability, we give only summarized results in Table 1. We indicate whether the effect was positive significant at the five (+) percent level (negative significant (-) respectively) by signs.

Table 1: Log-earnings equations by country and by wave, summarized findings

Country	wave	n	Age	Age2	Third level education	Basic Education	Unempl.time	Married
Germany	1	1,610	+	-	+	-	-	+
	2	1,528	+	-	+	-	-	
Germany SOEP	1	1,913	+	-	+	-	-	+
	2	1,821	+	-	+	-	-	+
	3	1,784	+	-	+	-	-	+
	4	1,574	+	-	+	-	-	+
United-Kingdom	1	1,804	+	-	+	-	-	+
	2	1,261	+	-	+	-	-	+
United-Kingdom BHPS	1	1,733	+	-	+	-	-	+
	2	1,676	+	-	+	-	-	+
	3	1,639	+	-	+	-	-	+
	4	1,620	+	-	+	-	-	+
Finland	3	1,115	+	-	+	-	-	+
Denmark	1	1,204	+	-	+	-	-	
	2	1,117	+	-	+	-	-	
	3	1,025	+	-	+	-	-	+
	4	947	+	-	+	-	-	
Ireland	1	1,300	+	-	+	-	-	+
	2	1,063	+	-	+	-	-	+
	3	869	+	-	+	-	-	+
	4	808	+	-	+	-	-	+
The Netherlands	1	1,549	+	-	+	-	-	+
	2	1,463	+	-	+	-	-	
	3	1,419	+	-	+		-	+
	4	1,371	+	-	+	-	-	+
Belgium	1	412			+	-	-	+
	2	697	+	-	+	-	-	+
	3	643	+	-	+	-	-	+
	4	597	+	-	+	-	-	+
Luxembourg	1	363	+	-	+	-	-	
	2	330	+	-	+	-	-	
France	1	1,837	+	-	+	-	-	+
	2	1,720	+	-	+	-	-	+
	3	1,546	+	-	+	-	-	+
	4	1,238	+	-	+	-	-	+
Spain	1	2,062	+	-	+	-	-	+
	2	1,863	+	-	+	-	-	+
	3	1,730	+	-	+	-	-	+
	4	1,663	+	-	+	-	-	+
Portugal	1	1,435	+	-	+	-	-	+
	2	1,393	+	-	+	-	-	+
	3	1,385	+	-	+	-	-	+
	4	1,373	+	-	+	-	-	+
Austria	2	1,198	+	-	+	-	-	
	3	1,067	+	-	+	-	-	+
	4	964	+	-	+	-	-	
Italy	1	1,843	+	-	+	-	-	+
	2	1,739	+	-	+	-	-	+
	3	1,571	+	-	+	-	-	+
	4	1,562	+	-	+	-	-	+
Greece	1	1,196	+	-	+	-	-	+
	2	1,044	+	-	+	-	-	+
	3	958	+	-	+	-	-	+
	4	919	+	-	+	-	-	+

We find an almost uniformly result concerning the sign and the level of significance for all countries and all waves. Like it was already visible in the non-parametric approach, there is a concave shape in the wage profile according to age. This is implied by the positive linear and negative quadratic age effect, which is evident in all countries. The positive effect of third level education as well as the negative effect of less than second stage education is highly significant in all countries and all waves, too, the third wave in Netherlands being the only exception.

For all countries there is a strong negative effect on cumulated unemployment time found.

While in most countries married males earn significant higher wages compared to non-married males, this effect is absent in (some waves) in the German ECHP, in the Netherlands, in Denmark and Austria. In Luxembourg there is in none of the two waves a significant effect of being married on wages present.

While table 1 gives an overview of the significance patterns of the income equation in all countries and for all waves, the following table 2 contains the parameters, as well as the approximate Wald-statistics and some model information for the first wave only.

Because the partial effect of education on earnings is of considerable interest, we visualize the findings of the regression in the following figures. Figure 5 displays the partial effect of third level education on the logarithm of the wage across countries.[14]

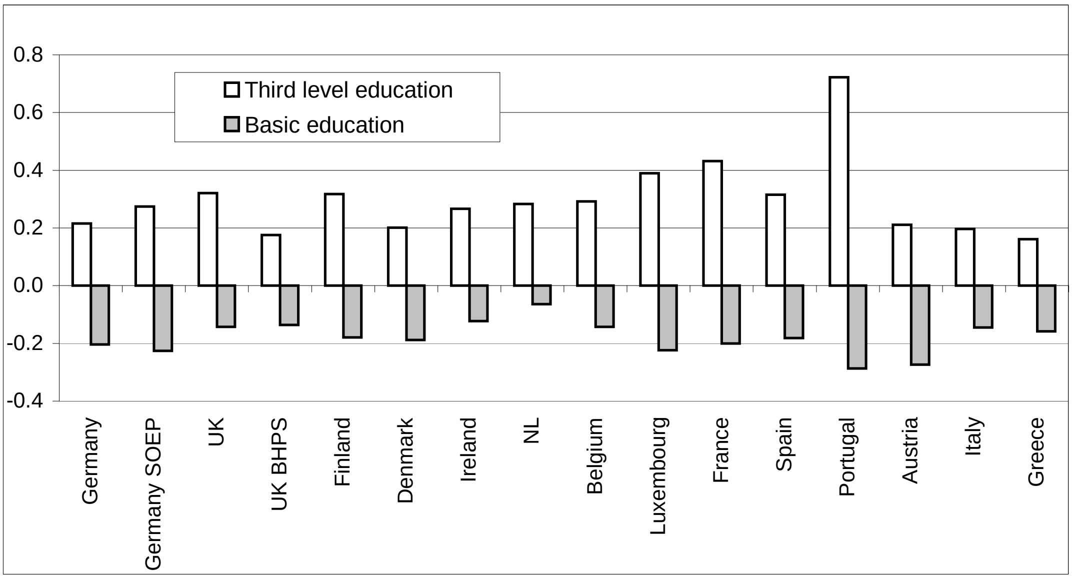
Table 2: Log-earnings equations by country, first wave results

Country	Germany	Germany SOEP	United-Kingdom	United-Kingdom BHPS	Finland	Denmark	Ireland	The Nether-lands
Wave	1	1	1	1	3	1	1	1
Intercept	4.7939	4.8867	4.6806	4.6692	5.2566	5.3606	5.0293	4.5530
	(35.38)	(35.42)	(35.99)	(40.91)	(31.14)	(39.92)	(32.88)	(27.28)
Age	0.1370	0.1314	0.1381	0.1429	0.1026	0.1191	0.1039	0.1486
	(17.94)	(16.94)	(19.23)	(22.74)	(10.92)	(16.42)	(11.46)	(16)
Age2/100	-0.1619	-0.1593	-0.1647	-0.1710	-0.1162	-0.1394	-0.1237	-0.1730
	(-16.69)	(-16.31)	(-18.89)	(-22.11)	(-9.6)	(-15.58)	(-10.84)	(-14.48)
Education - third level	0.2150	0.2747	0.3502	0.1758	0.3111	0.2075	0.2801	0.2663
	(7.72)	(10.15)	(10.92)	(5.22)	(10.99)	(7.07)	(7.46)	(9.7)
less than 2nd stage of	-0.2041	-0.2258	-0.1338	-0.1361	-0.1811	-0.1804	-0.1382	-0.1001
	(-6.7)	(-6.6)	(-4.38)	(-3.82)	(-4.85)	(-5.41)	(-4.4)	(-2.92)
Unempl.time	-0.0025	-0.0040	-0.0075	-0.0090	-0.0072	-0.0091	-0.0070	-0.0072
	(-2.25)	(-5.27)	(-6.21)	(-3.73)	(-6.29)	(-5.77)	(-7.64)	(-5.83)
Married	0.0765	0.0542	0.1062	0.0986	0.0689	0.0550	0.2911	0.0902
	(2.67)	(2.02)	(3.38)	(3.54)	(2.34)	(1.87)	(7.35)	(3.24)
r2	0.40	0.36	0.36	0.39	0.37	0.37	0.37	0.36
n	1,610	1,913	1,804	1,733	1,115	1,204	1,300	1,549

Country	Belgium	Luxem-bourg	France	Spain	Portugal	Austria	Italy	Greece
Wave	1	1	1	1	1	2	1	1
Intercept	6.6957	5.5591	5.7261	5.2078	5.0525	5.4395	5.7224	4.9514
	(22.61)	(17.8)	(37.83)	(47.99)	(47.14)	(36.93)	(57.47)	(38.24)
Age	0.0246	0.1170	0.0738	0.0764	0.0613	0.1021	0.0668	0.0628
	(1.48)	(6.38)	(8.66)	(12.48)	(10.39)	(11.44)	(11.74)	(8.62)
Age2/100	-0.0206	-0.1525	-0.0791	-0.0862	-0.0753	-0.1186	-0.0804	-0.0703
	(-0.96)	(-6.05)	(-7.11)	(-11.13)	(-10.45)	(-9.65)	(-10.93)	(-7.72)
Education - third level	0.2918	0.3825	0.4301	0.2834	0.7230	0.2146	0.1964	0.1630
	(6.76)	(6.57)	(16.61)	(10.26)	(13.13)	(4.14)	(6.89)	(6.01)
less than 2nd stage of	-0.1428	-0.2193	-0.1921	-0.2027	-0.2872	-0.2989	-0.1446	-0.1646
	(-2.85)	(-4.42)	(-7.55)	(-8.41)	(-8.94)	(-8.13)	(-8.88)	(-6.24)
Unempl.time	-0.0038	-0.0278	-0.0120	-0.0036	-0.0034	-0.0104	-0.0023	-0.0022
	(-2.38)	(-3.7)	(-9.76)	(-5.92)	(-2.69)	(-3.77)	(-8.49)	(-1.98)
Married	0.1315	0.0037	0.1264	0.1180	0.1221	0.0170	0.0944	0.1911
	(2.94)	(0.08)	(5.31)	(5)	(4.16)	(0.5)	(4.71)	(6.87)
r2	0.23	0.42	0.34	0.35	0.39	0.34	0.28	0.29
n	412	363	1,837	2,062	1,435	1,198	1,843	1,196

14 As we mentioned in section 3.2, the measured effect will include schooling as well as ability effects. For an attempt to correct for a selectivity bias and to isolate the schooling effect in the case of three choices see Garen (1984).

Fig. 5: Partial educational effects on log-wages across countries



The uniformly negative effect of less than second stage as well as the uniformly positive effect of third level education show considerable variation across countries. The premium for third level education is by far highest in Portugal. Here the wage increase exceeds 70% percent compared to second stage education. In Luxembourg and France with about 40% the earnings-differential is also rather strong. Males with less than second stage education suffer the highest losses compared to second stage education in Portugal and Austria.[15]

4. Analyzing the effect of attrition on the earnings equations

While in the preceding section we did not take into account any possible biasing effects of attrition on the income findings, we now turn to explicit analyze such possible attrition effects. This attrition analysis is carried out in several steps. First we model the response behavior of active mails by the means of a logit model. In a second step we analyze whether income equations in period 1 show significant differences, if the estimation includes persons responding in wave 2 only compared to estimations based on the full sample. In a third step we apply an inverse probability weighted estimating approach (IPW) and assess whether this method leads to a reduced bias in the estimating results compared to the un-weighted approach.

4.1. Earnings and response behavior

In this section we analyze the relation between earnings and the response behavior of active mails.[16] 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 the response rate (bold line) in wave two. To obtain the density of earnings we employ a triangular kernel with a constant bandwidth of 300 € for all countries. The

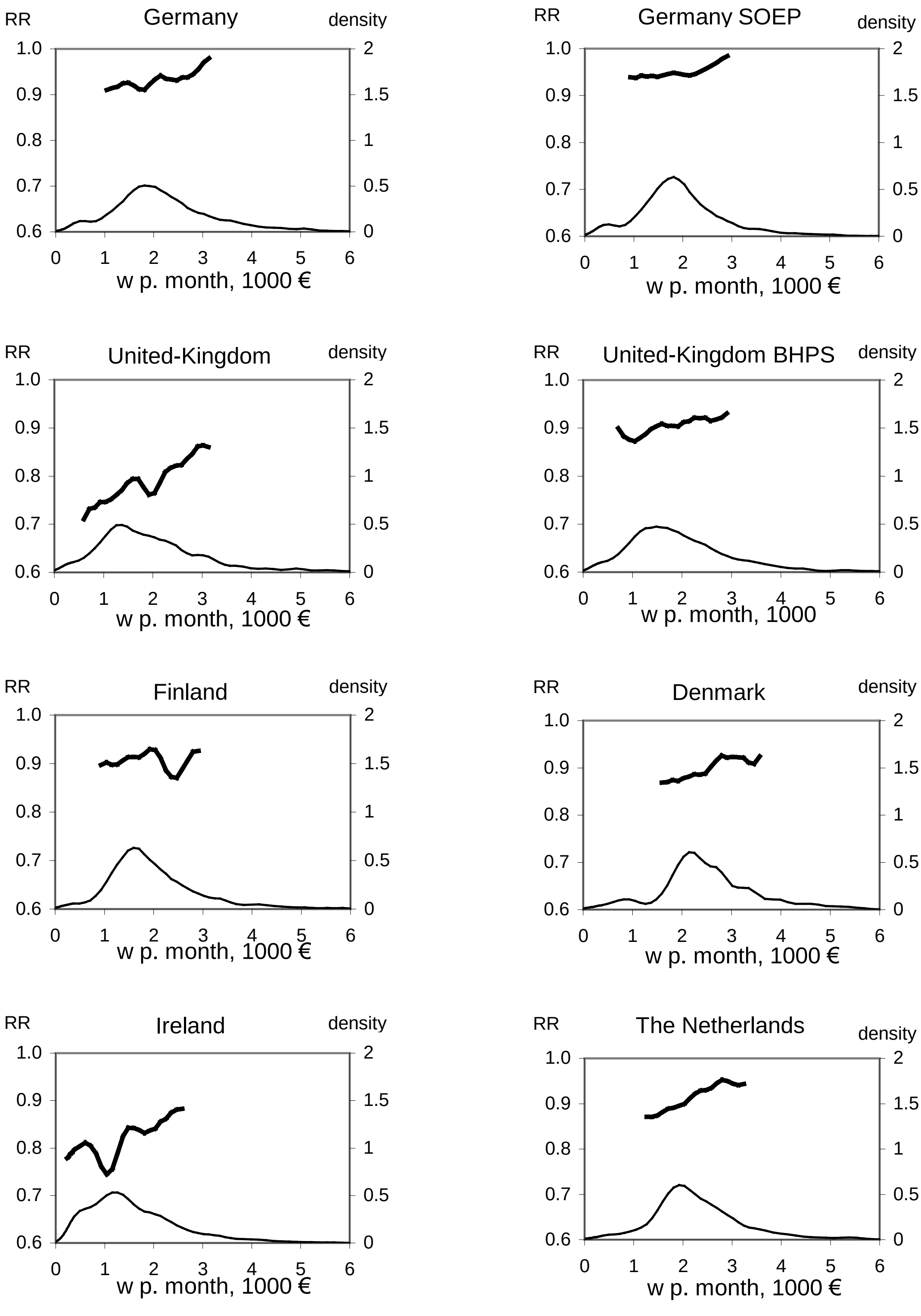
---

15 For attempts to estimate social rates of return to schooling, which includes the difficult measure of public costs, see e.g. Hines/Tweeten/Redfern (1970) an Vaillancourt (1995).

16 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-nonresponse and imputation methods in the ECHP.

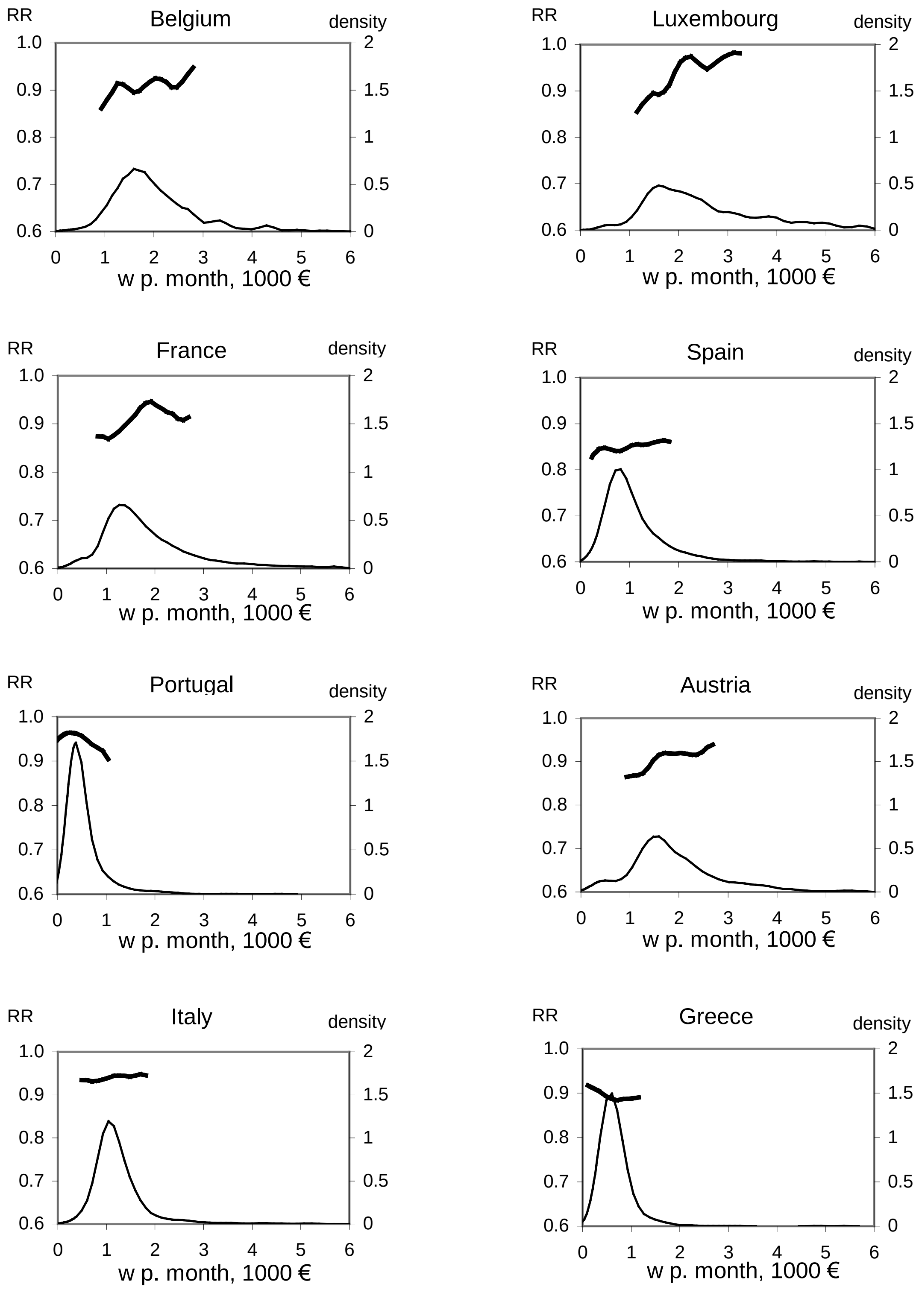
response rate is obtained through the estimation of non-parametric regressions using triangular kernels as weights and bandwidth 300 €.[17]

Fig. 6: Earnings and response rates across countries



---

17 See e.g. Härdle (1990) for an overview of non-parametric methods.



There is no clear relation visible that holds for all countries. But for most of the countries we find that the response is increasing with earnings. The only exceptions are Portugal and Greece. The curvature of the response rate can be described as fairly linear for most of the countries. Noteworthy exceptions are United Kingdom, Finland and Ireland, where in the interval of high density a sharp increase followed by an increase is visible.

4.2. Modeling the response probability

In this section we estimate response probabilities using a logit model. Hence we assume that the latent variable R_t^* is distributed according to the logistic distribution conditional on its expected value, which we model as a linear function of a set of variables $X_{t\text{-}1}$ and a set variables V_t. While X is observable only prior to the attrition period, the set of additional (field-) variables V is observable in the period of attrition. Field variables, known to by rather important for attrition behavior, are in particular the information of a move in the attrition period and the information about a chance of interviewer. Both variables were found highly significant in empirical studies (see e.g. Rendtel (1995), Behr/Bellgardt/Rendtel (2002)).

In the logistic model it is assumed that the observed response R takes the value of 1, if the logistic error term is below 0.

$$R_t^* = \gamma_1' X_{t-1} + \gamma_2' V_t + \delta_t$$

$$R_t = \begin{cases} 1 & if \quad R_t^* = \gamma_1' X_{t-1} + \gamma_2' V_t + \delta_t > 0 \\ 0 & else \end{cases}$$

We use the following variables for explaining the attrition behavior in the logit analysis:[18]

- log-earnings (ln *w*), lagged one period

- *age* as well as age raised to the power of two (age^2), lagged one period

- dummy variable *edu*1 less than second stage education (second level education we use as the base category), lagged one period

- dummy variable *edu*3 to indicate third level education, lagged one period

- variable *cumemp* denotes the cumulated unemployment time in month, lagged one period

- dummy variable *married* indicates if the person is married (all other marital stati (divorced, widowed, not married) we use as the base category), lagged one period

- move of a household in the wave under analysis[19]

- interviewer change (if available) in the wave under analysis

Note that we include in our logit analysis to explain the response behavior all explanatory variables contained in the income equation, supplemented by the additional variables lagged log-earnings and the contemporaneous dummy variables indicating whether the person has moved and whether the interviewer changed.

---

18 In our analysis we only consider variables that vary at the individual level. For an analysis including country characteristics and further information of the data collecting process, like whether the interview was by phone or personal interview, see Nicoletti/Peracchi (2002).

19 The variable indicating a change of the interviewer was not available in the ECHP-User Data Base for Denmark, Spain, Portugal and Greece. This indicated by a missing value in the table (.). For some waves the inclusion of the variables indicating a move and the change of the interviewer caused numerical problems in the estimation procedure due to almost perfect dependence of variables (attrition, moved and change of interviewer). In these cases the variables moved and, if necessery, the variable change of interviwer were dropped.

Table 3: Logit results by country and by wave

Country	wave	n	lnW	Age	Age2	Third level education	Basic education	Unempl.Time	Married	HH has moved	Interviewer change
Germany	1	1,610	+						+	-	-
	2	1,528						-			
Germany SOEP	1	1,913								-	
	2	1,821								-	
	3	1,784								-	-
	4	1,574					-		+	-	-
United-Kingdom	1	1,804						-	+	.	.
	2	1,261							+	.	.
United-Kingdom BHPS	1	1,733	+								
	2	1,676		+							
	3	1,639									
	4	1,620		+							
Finland	3	1,115								.	.
Denmark	1	1,204							+	.	.
	2	1,117								.	.
	3	1,025	+							.	.
	4	0,947					-			.	.
Ireland	1	1,300							+	.	.
	2	1,063					+			.	.
	3	0,869								.	.
	4	0,808		+	-					.	.
The Netherlands	1	1,549							+	.	.
	2	1,463							+	.	.
	3	1,419		+	-				+	.	.
	4	1,371					-			.	.
Belgium	1	0,412								-	
	2	0,697									
	3	0,643		+	-			-			-
	4	0,597								-	-
Luxembourg	1	0,363					+			.	.
	2	0,330							+	.	.
France	1	1,837		+	-				+	.	.
	2	1,720					-		+	.	.
	3	1,546		+	-		-			.	.
	4	1,238		+	-					.	.
Spain	1	2,062						-	+	.	.
	2	1,863							+	.	.
	3	1,730							+	.	.
	4	1,663		-	+				+	.	.
Portugal	1	1,435					+			.	.
	2	1,393	-						+	.	.
	3	1,385	-							.	.
	4	1,373								.	.
Austria	2	1,198								.	.
	3	1,067							+	.	.
	4	0,964							+	.	.
Italy	1	1,843							+	-	-
	2	1,739									-
	3	1,571	-					+		-	-
	4	1,562								-	-
Greece	1	1,196	-						+	.	.
	2	1,044								.	.
	3	0,958								.	.
	4	0,919					+			.	.

To ease the readability and to allow for an overview of the results of the 54 logistic regressions, we reduced the information given in table 3.[20] 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 behavior. We note a

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

significant positive (+) as well as a significant negative (-) influence by means of a sign, each at the five percent level, while we did not use any sign when the parameter was insignificant.

Turning to the results, we find that the log-wage is significant in some regressions only. Rather unexpectedly we find a mixed picture for the direction of the influence. While in the northern countries the log wage exerts a positive influence, in the southern European countries we find that the response probability is decreasing with increasing wage. The altogether rather weak influence of wage on the response probability could be seen as indicating a case of missing at random (MAR). But of course it has to be taken into account, that we include the log wage lagged one period.

Age and the age raised to the power of two exerts only in some waves significant influence. If the age effects are significant, in most cases we find a positive linear and a negative squared effect for age.

The move of the household exerts a negative influence on the response behavior, almost whenever the variable could be included in the logistic regression. In Germany (both surveys), Denmark, Belgium and Italy we find a significant lower response probability for moving households.

The marital status was reflected using two different categories only. Compared to the reference category (not married) we find that married males have an almost unambiguous tendency to higher response rates.[21]

For the level of education we find only in very few significant parameters. Third level education has in no country and no wave a significant influence on the response behavior. People who gained the highest level of education show no significant tendency to respond differently in any of the 54 waves.[22] For persons with less than second stage education there is only in very few waves a significant influence present and the direction of the influence varies across countries.[23]

In several regressions cumulated unemployment time decreases the response probability, the only exception being the third wave in Italy.

The variable indicating the change of the interviewer is significant negative in most of the cases, if included in the regression. Hence, the change of the interviewer significantly increases the risk of attrition. This result confirms strongly the findings of Rendtel (1995) for the German SOEP and of Behr/Bellgardt/Rendtel (2002) for the ECHP (all persons). The same holds for persons changing their dwellings. The move significantly increases the risk of attrition.

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. In no country we find one variable to exert significant influence in opposite direction in different waves. This indicates that the pooling of the data across waves within countries would be more in accordance with the data then pooling across countries, where often opposing influences of the covariates were found.

---

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

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

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

The following table contains the parameter estimates of the first available wave for each country in detail.

Beside the parameters the table gives some model information. The overall chi-square-test of model significance clearly rejects the hypothesis of no combined explanatory power of the model except four United Kingdom (BHPS), Finland and Austria. McFadden's Likelihood-Ratio-Criterion has a rather low value about 0.04 in most countries. The R^2 (R2MZ) suggested by McKelvey and Zavoina (1975) has a slightly higher value in average (0.07). 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 somewhat in contrast to the Chi-Square test proposed by Hosmer and Lemeshow (1980), which indicates according to the high *p*-value a satisfactorily model fit.

Table 4: First wave logit results by country

Country	Germany	Germany SOEP	United-Kingdom	United-Kingdom BHPS	Finland	Denmark	Ireland	The Nether-lands
Wave	1	1	1	1	3	1	1	1
Intercept	1.6919	2.4065	1.0574	-0.4680	0.5534	1.9489	0.5125	-0.7511
	(1.07)	(1.27)	(1.42)	(-0.24)	(0.3)	(1.3)	(0.49)	(-0.55)
lnW	0.4759	0.4533	0.0559	0.9324	-0.1448	-0.0308	-0.0400	0.0489
	(2.16)	(1.91)	(0.54)	(3.24)	(-0.58)	(-0.15)	(-0.28)	(0.26)
Age	-0.0984	-0.1595	-0.0257	-0.1608	0.1295	0.0050	0.0455	0.1097
	(-1.33)	(-1.54)	(-0.74)	(-1.72)	(1.6)	(0.08)	(0.93)	(1.48)
Age2	6.836E-04	2.231E-03	3.112E-04	1.951E-03	-1.334E-03	1.600E-04	-4.094E-04	-1.172E-03
	(0.74)	(1.6)	(0.74)	(1.68)	(-1.27)	(0.2)	(-0.65)	(-1.21)
Third level education	0.2672	0.0318	0.0396	0.2873	-0.2355	0.0477	-0.0076	0.4047
	(0.88)	(0.09)	(0.26)	(0.66)	(-0.97)	(0.2)	(-0.04)	(1.55)
Basic education	-0.1990	-0.2479	-0.1578	0.2144	0.3130	-0.4539	-0.0557	-0.4120
	(-0.7)	(-0.7)	(-1.17)	(0.48)	(0.87)	(-1.92)	(-0.35)	(-1.8)
Unempl.Time	0.0053	0.0201	-0.0129	0.0333	-0.0067	-0.0150	-0.0070	0.0136
	(0.47)	(1.44)	(-2.64)	(0.71)	(-0.87)	(-1.57)	(-1.55)	(1.06)
Married	0.7140	0.3454	0.5574	0.1018	0.1083	0.5193	0.5171	0.5668
	(2.55)	(1.14)	(4.05)	(0.27)	(0.43)	(2.2)	(2.48)	(2.6)
HH has moved	-2.0357	-1.3416		-0.3069				
	(-4.27)	(-3.47)		(-0.48)				
Interviewer change	-0.9646	-0.0694		-0.0164				
	(-3.3)	(-0.18)		(-0.04)				
CHI	46.14	25.36	34.05	11.94	12.02	23.30	29.42	49.66
p	0.00	0.00	0.00	0.22	0.10	0.00	0.00	0.00
LRI	0.07	0.04	0.02	0.03	0.02	0.03	0.02	0.05
R2MZ	0.11	0.09	0.03	0.07	0.04	0.06	0.04	0.09
CHI,HL	13.29	4.39	7.16	7.44	15.60	3.19	2.85	3.38
p,HL	0.10	0.82	0.52	0.49	0.05	0.92	0.94	0.91
n	1,610	1,913	1,804	1,733	1,115	1,204	1,300	1,549

Country / Wave	Belgium 1	Luxem-bourg 1	France 1	Spain 1	Portugal 1	Austria 2	Italy 1	Greece 1
Intercept	-5.6927	-4.7680	-0.8418	0.2388	3.3276	1.5939	5.5426	4.8341
	(-1.54)	(-1.29)	(-0.58)	(0.23)	(1.89)	(1.12)	(2.69)	(2.98)
lnW	0.6223	0.6823	-0.1627	0.1214	-0.3169	0.2880	-0.3376	-0.5539
	(1.51)	(1.35)	(-0.9)	(0.82)	(-1.19)	(1.52)	(-1.16)	(-2.32)
Age	0.1941	0.0191	0.2271	0.0378	0.0521	-0.0878	-0.0169	0.0186
	(1.47)	(0.09)	(3.72)	(0.92)	(0.76)	(-1.35)	(-0.24)	(0.29)
Age2	-2.676E-03	6.526E-04	-2.864E-03	-7.268E-04	-6.132E-04	1.002E-03	9.788E-05	-3.032E-04
	(-1.58)	(0.22)	(-3.65)	(-1.43)	(-0.74)	(1.12)	(0.11)	(-0.38)
Third level education	-0.5366	0.7110	0.2468	-0.0860	-0.5832	-0.2514	-0.0124	-0.2871
	(-1.27)	(1.24)	(1.07)	(-0.46)	(-1.31)	(-0.72)	(-0.04)	(-1.35)
Basic education	-0.2409	1.0794	-0.3190	0.2288	0.6544	-0.2203	-0.2562	0.4812
	(-0.5)	(2.54)	(-1.67)	(1.38)	(2.09)	(-0.89)	(-1.26)	(1.94)
Unempl.Time	0.0260	-0.0806	-0.0024	-0.0082	0.0257	-0.0238	0.0063	0.0093
	(0.68)	(-1.46)	(-0.25)	(-2.25)	(0.99)	(-1.7)	(1.39)	(0.9)
Married	0.7034	0.0300	0.4215	0.4612	-0.1447	0.2911	0.6074	0.5845
	(1.73)	(0.07)	(2.18)	(2.87)	(-0.45)	(1.26)	(2.55)	(2.54)
HH has moved	-1.5607						-0.8693	
	(-2.14)						(-2.33)	
Interviewer change	-0.3102						-1.0562	
	(-0.7)						(-5.72)	
CHI	19.41	17.79	32.63	26.78	25.07	9.18	51.50	30.36
p	0.02	0.01	0.00	0.00	0.00	0.24	0.00	0.00
LRI	0.08	0.08	0.03	0.02	0.04	0.01	0.05	0.03
R2MZ	0.13	0.16	0.05	0.03	0.07	0.02	0.10	0.07
CHI,HL	16.53	2.29	8.62	6.74	12.85	10.06	7.92	7.40
p,HL	0.04	0.97	0.38	0.56	0.12	0.26	0.44	0.49
n	412	363	1,837	2,062	1,435	1,198	1,843	1,196

4.3. The IPW-approach

In this section we describe the inverse response probability weighting approach, which potentially reduces the panel attrition bias (Robins, Rotnitzky and Zhao (1995), Neukirch (2002)). We assume that the individual log-earnings can be modeled as a linear equation containing an error term and that the response behavior can be modeled by a logistic response equation:

$$Y_t = \beta' X_t + \varepsilon_t \;\; if \;\; R_t = 1$$

$$R_t^* = \gamma_1' X_{t-1} + \gamma_2' V_t + \delta_t$$

$$R_t = \begin{cases} 1 & if \;\; R_t^* = \gamma_1' X_{t-1} + \gamma_2' V_t + \delta_t > 0 \\ 0 & else \end{cases}$$

here Y is log-wage, R is the observed response variable and X contains explanatory variables common to both equations, while V contains additional variables considered as influencing the attrition behavior only.

If the log-wage equation is estimated making use of available respondents ($R = 1$), only in the case of missing at random (MAR) will unbiased estimates of the parameter vector of interest, β, be obtained.[24] This means that one has to rely on the assumption that, given the set of explanatory variables, the missingness mechanism is independent of contemporaneous Y_t:

$$P\left(R_t = 1 \mid X_{t-1}, V_t, Y_{t-1}, Y_t\right) = P\left(R_t = 1 \mid X_{t-1}, V_t, Y_{t-1}\right).$$

24 See Rendtel (1995) for a theoretical overview of panel attrition models.

This is equivalent to the desirable condition that $P(Y_t \mid X_t, R_t = 1) = P(Y_t \mid X_t)$ and hence estimates making use of respondents only will mirror the relation between *Y* and *X* for attriters as well as respondents.

If we denote the estimated response probabilities $\hat{\pi}$, and the diagonal matrix of estimated probabilities $\hat{\Pi}$, the IPW-estimator can be written as

$$\hat{\beta}_{IPW} = \left(X_t' \, \hat{\Pi}_{t+1}^{-1} X_t\right)^{-1} X_t' \, \hat{\Pi}_{t+1}^{-1} Y_t$$

with

$$prob(\hat{R}_{t+1} = 1) = \hat{\pi}_{t+1} = \frac{e^{X_t \gamma_1 + V_{t+1} \gamma_2}}{1 + e^{X_t \gamma_1 + V_{t+1} \gamma_2}}$$

The two-step procedure is rather intuitive. Because the observable sample contains respondents only, each available observation is weighted with the root of its inverse response probability. Because the observable observations will resemble the observations lost due to attrition the more the lower their response variability, the increased weight given to these observations (through dividing by the root of the low response probability) should improve the resemblance of the observable sample to the full sample.

Since the weights $\hat{\pi}$ are estimated and hence contain random variation, this has to be taken care of when doing inference on the estimated coefficients of the income equation. Hence, we do not only use the potentially misleading standard *t*-statistics obtained from the logit model, but rather carry out a non-parametric bootstrap simulation to assess the significance of the parameters.

4.4. A comparative empirical bias analysis

We perform the following procedure to judge the empirical performance of the weighted and un-weighted estimation procedures. Since we do not observe the full sample in the period of attrition, we have no possibility to judge the true bias of the weighted and un-weighted estimators under real conditions. To get nevertheless some hints about the possible attrition bias and whether a weighing approach tends to reduce such a possible bias, we split the full sample, hence attriters and respondents, in one wave into samples of attriters and respondents and respondents only, according to the response behavior in the ***following wave***. While this method gives the opportunity to assess empirically the performance of the different proposed estimators, the value of this comparison rests on the assumption, that the log-wage equation would be the same whether we use wave 1 or wave 2 variables. If the IPW-approach in this counterfactual analysis outperforms the un-weighted estimator (neglecting attrition), one could hope for an improvement in the realistic application of the IWP-approach in wave 2, hence in the period the attrition actually takes place.

Taking together we consider the following estimators in our analysis:

β_1 — the OLS-estimator, obtained using observations in wave 1 for persons responding in wave 2

$\beta_{0,1}$ — the OLS-estimator, obtained when using all observations in wave 1, whether persons respond in wave 2 or attrit.

$\beta_{1,IPW}$ the IPW-estimator, obtained using observations in wave 1 for persons responding in wave 2

Table 5: Pattern of relative bias, OLS and IPW, by country and by wave

Country	wave	n	Intercept		Age		Age2		Third level education		Basic education		Unempl.time		Married	
			OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW
Germany	1	1,610												+	- -	- -
	2	1,528									+	+	++	++		
Germany SOEP	1	1,913													- -	- -
	2	1,821									++	++			+	
	3	1,784									-	- -			-	-
	4	1,574									++	++			-	-
United-Kingdom	1	1,416			+	+		+	-	-	+	++	++	++	- -	- -
	2	1,082									++	++			++	++
United-Kingdom BHPS	1	1,733								+					-	-
	2	1,676														
	3	1,639									+	+				
	4	1,620											+			
Finland	3	1,017													+	+
Denmark	1	1,202							-	-	+	+	- -	- -	- -	- -
	2	1,112														++
	3	1,016									- -	- -	- -	- -	- -	-
	4	941										+	- -	-	++	++
Ireland	1	1,081							-	-		-	+		+	
	2	1,048											-	-		
	3	864									-	- -				
	4	800									- -	- -	- -	- -		
The Netherlands	1	1,430							+		- -	- -	-		+	
	2	1,379									- -	- -				-
	3	1,315									++	++				
	4	1,350				+		+			+	+				
Belgium	1	412			- -	- -	- -	- -			++	+			++	++
	2	697											-			
	3	643				+	-	++	-	-	++	++		- -	-	-
	4	597			+	++	+	++			++	++	- -	- -	-	-
Luxembourg	1	331	+		- -		- -						+		++	++
	2	311							- -	- -	+	+			++	++
France	1	1,777			+	++	++	++			+					
	2	1,610			-	-	-	-					+	+		
	3	1,403									+	+		-		
	4	1,125									-					
Spain	1	1,861							+	+	-	-		-	- -	-
	2	1,714			+	++	+	++					- -	- -	- -	- -
	3	1,516											-	-		
	4	1,498					-		+							
Portugal	1	1,433			-	-	-	-								+
	2	1,332							-	-			-		++	++
	3	1,333								-						
	4	1,322											- -	- -	+	+
Austria	2	1,069			+	+	+	+	-	-	-	- -	- -	- -	-	
	3	942							++	+				- -	- -	- -
	4	859			- -	- -	- -	- -	++	++			-	-	++	++
Italy	1	1,843			-	-	-	-			+					
	2	1,739														
	3	1,571								+						-
	4	1,562										-	+	+	-	-
Greece	1	1,053								+			- -	- -	+	
	2	928			+	+	+	+	- -	- -	+	+		+		-
	3	862								-	+	+		+		
	4	784							++	+			- -	-	-	-

Bias: >10.0% (++); between 5.0% and 10.0% (+); <-10.0% (--); between -5.0 and -10.0% (-).

In table 5 we give an overview of the relative biases for the two different estimation approaches. Comparing the parameters obtained on the basis of the respondents only with the parameters of the full sample (respondents and attriters), we denote a relative difference between 5 and 10% by one sign and differences greater than 10% by two signs.

The table also shows that there are noticeable differences in the way the earnings equations are affected in different countries. Especially for Germany, the United Kingdom (BHPS) and Italy we find that only a few parameters are slightly affected. On the other for Belgium and Austria the findings indicate a strong bias of the regression results due to attrition. Looking at the columns, hence comparing the results across all 54 available waves, we find no unique pattern. This result indicates that there is no identical pattern in which regression results are influenced through attrition across the different countries.

To assess the significance of the bias of the four different estimators, OLS and IPW in the two different attrition models, we carry out Hausman-tests to test whether the difference between each of the four parameters and the parameter using the information of attriters as well as respondents ($\beta_{0,1}$) is significant.

The estimate of the bias is

$$\hat{b}(\beta) = \hat{\beta}_1 - \hat{\beta}_{0,1} .$$

The hypothesis $b(\beta) = 0$ ist tested against the alternative $b(\beta) \neq 0$ making use of the asymptotic result that the covariance matrix of the difference (Σ_{diff}) between a consistent estimator under the null-hypothesis ($\hat{\beta}_1$) and an efficient estimator ($\hat{\beta}_{0,1}$) is given by the difference of the covariance matrix of the consistent estimator (Σ_{con}) and the covariance matrix of the efficient estimator (Σ_{eff}):

$$\Sigma_{diff} = \Sigma_{con} - \Sigma_{eff} .$$

The Hausman test statistic is then calculated as:

$$t_{Haus} = \left(\hat{\beta}_1 - \hat{\beta}_{0,1}\right)' \Sigma_{diff}^{-1} \left(\hat{\beta}_1 - \hat{\beta}_{0,1}\right) \sim \chi_k^2$$

Taking into account the additional variability of the estimated response probabilities in the IPW-approach, we use two different test procedures. Tables 6 and 7 contain Hausman-tests, which are based on the fact that the two estimates are estimated with two different data sets, whereof one is a subset of the other. Hence the parameters obtained from the data set including attriters as well as respondents should be efficient, while the subset of respondents is using the information of respondents is consistent only.

Since empirically the standard deviation of the less efficient estimated parameters are in some cases smaller than the standard errors of the estimation using the full data set, the Hausman-test can not be carried out due to negative differences in the square root of the denominator. These cases are indicated by a dot (.) in the table. The signs, which are displayed in the case of significance at the 5%-level, indicate the direction of the bias.

Comparing the results of the two different estimates, there is no superiority of one estimation procedure visible. Throughout all available 54 waves, the number of significant parameter-differences is rather small and the direction of the bias changes often from wave to wave. Nevertheless, some regressions, e.g. Austria wave 4, seem to be strongly affected by attrition.

Table 6: Hausman-tests of OLS- and IPW-bias, by country and by wave

Country	wave	n	Age OLS	Age IPW	Age2 OLS	Age2 IPW	Third level education OLS	Third level education IPW	Basic education OLS	Basic education IPW	Unempl.time OLS	Unempl.time IPW	Married OLS	Married IPW
Germany	1	1,610		+									-	-
	2	1,528							-	-		-		
Germany SOEP	1	1,913					+							
	2	1,821							-	-	-			
	3	1,784					+							
	4	1,574												
United-Kingdom	1	1,804		+										-
	2	1,261							-	-				
United-Kingdom BHPS	1	1,733	.		.	.	.	+	.		.	-	.	-
	2	1,676												
	3	1,639					-	-	-		.			+
	4	1,620												
Finland	3	1,115										.		
Denmark	1	1,204	.	.	.	.	.	.	.	.	+	.	.	.
	2	1,117												
	3	1,025	.	.	.	.	.	.	.	.	+	.	.	.
	4	947									+			
Ireland	1	1,300												
	2	1,063												
	3	869												
	4	808									+			
The Netherlands	1	1,549		+			+		+	+	+			
	2	1,463												
	3	1,419												
	4	1,371		+		-								
Belgium	1	412		-		+								
	2	697												
	3	643		+		.	-	-	-	-				
	4	597		+		-					+	.		
Luxembourg	1	363	-	.	+	.							+	+
	2	330					-	-						
France	1	1,837	+	+	-	-								
	2	1,720	-	-	+	+								
	3	1,546												
	4	1,238												
Spain	1	2,062					+			+				
	2	1,863	+	+	-	-					+	+		
	3	1,730												
	4	1,663												
Portugal	1	1,435	-	-	+	+		+			.			
	2	1,393					-	-			.		+	+
	3	1,385					-	-						
	4	1,373									+			
Austria	2	1,198	+	+	-	-			+	+		.		
	3	1,067												
	4	964	-	-	+	+	+	+				+		
Italy	1	1,843	-	-	+	+								
	2	1,739												
	3	1,571												
	4	1,562												
Greece	1	1,196												
	2	1,044					-	-						
	3	958												
	4	919												

First, because the Hausman-tests fails in several cases and second, because the IPW-estimators are based on estimated response probabilities, what is not explicitly taken into account when doing inferences, we apply alternatively bootstrap sampling to assess the significance of the parameter differences between the estimators based on the full sample and based on the respondents respectively.

Table 7: Bootstrap tests of OLS- and IPW-bias, by country and by wave

Country	wave	n	Age		Age2		Third level education		Basic education		Unempl.time		Married	
			OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW	OLS	IPW
Germany	1	1,610												
	2	1,528							-	-				
Germany SOEP	1	1,913												
	2	1,821							-	-				
	3	1,784												
	4	1,574												
United-Kingdom	1	1,416												
	2	1,082							-	-				
United-Kingdom BHPS	1	1,733									-			
	2	1,676					+							
	3	1,639											+	
	4	1,620												
Finland	3	1,017												
Denmark	1	1,202												
	2	1,112												
	3	1,016												
	4	941												
Ireland	1	1,081												
	2	1,048									+			
	3	864												
	4	800									+	+		
The Netherlands	1	1,430												
	2	1,379												
	3	1,315									+			
	4	1,350		+		-								
Belgium	1	412												
	2	697									+			
	3	643						-	-	-				
	4	597												
Luxembourg	1	331	-		+								+	
	2	311					-	-						
France	1	1,777		+	-	-								
	2	1,610	-	-	+	+								
	3	1,403												
	4	1,125												
Spain	1	1,861												
	2	1,714	+	+	-	-						+		
	3	1,516												
	4	1,498												
Portugal	1	1,433												
	2	1,332									+		+	
	3	1,333						-						
	4	1,322												
Austria	2	1,069	+	+		-			+	+				
	3	942												
	4	859	-	-	+	+								
Italy	1	1,843	-	-	+	+								
	2	1,739												
	3	1,571												
	4	1,562												
Greece	1	1,053												
	2	928	+	+	-	-	-	-						
	3	862							-					
	4	784												

For each wave we draw 1000 non-parametric bootstrap replications with replacement. The bootstrap method is applied to simulate the distribution of the bias $\hat{b}(\beta) = \hat{\beta}_1 - \hat{\beta}_{0,1}$. For each realization we got $\hat{\beta}_1^*$ and $\hat{\beta}_{0,1}^*$, the bootstrap versions of $\hat{\beta}_1$ and $\hat{\beta}_{0,1}$. To assess the significance of the estimated bias, we used the 2.5% and 97.5% quantiles of the simulated bootstrap distributions.

In general the number of significant parameter differences found decreased compared to the Hausman-tests. While in no case opposite findings occurred (significant positive bias in the un-weighted estimator turning out negative significant in the weighted estimator and vice versa), in several cases the significance changed. Altogether one has to state that the estimation of the income equations are only slightly affected by attrition. Only about a tenth of all estimated parameters turned out to be affected significantly. Taking into account that for each parameter we test at the five-percent level, the overall finding should be interpreted as strongly supportive for the view of only very mild attrition bias in the ECHP.

Fig. 7: Medians and quantiles of the parameter differences in 54 cross sections

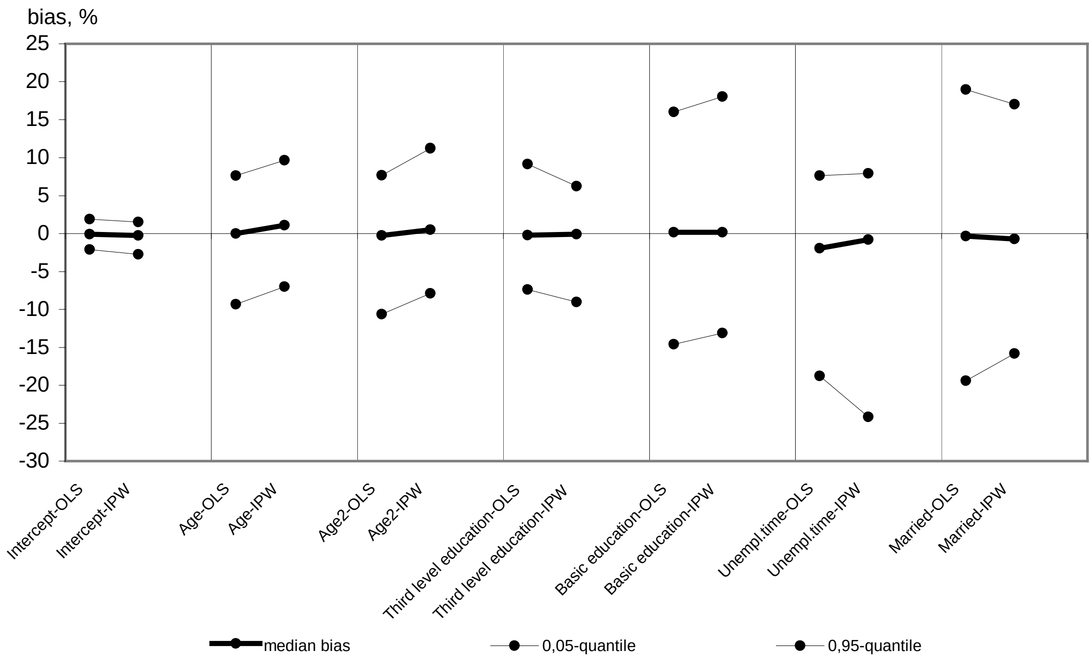
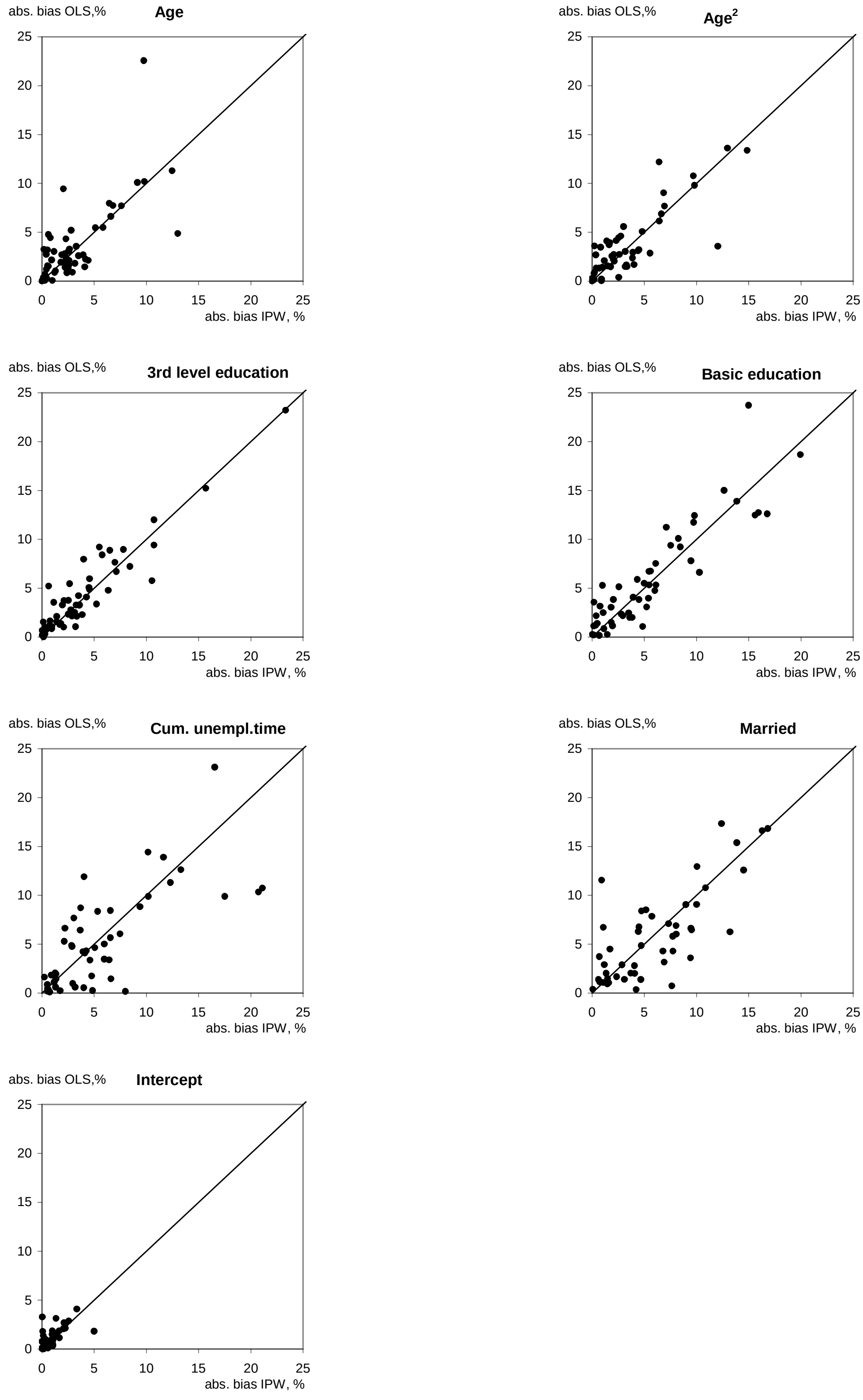


Figure 7 contains descriptive evidence on the relative parameter differences in all 54 waves between the two estimates based on the respondents (OLS, IPW) and the estimates based on the full sample in the 1000 bootstrap replications. Of course considering all 54 waves together might hide some specific results for some waves, but the overall evidence is quite impressive. For no parameter the midst 90% of the bootstrap replications show relative differences outside the interval ranging from –25 to +25 percent. This adds further evidence to the conclusion so far, that the problem of attrition bias is not severe in the earnings estimates from the ECHP.

The following figure contains scatter plots of pairs of absolute values of the relative OLS/IPW-biases of the 54 waves for each parameter. The closer the points are to the origin, the smaller is the bias. Points above the diagonal indicate the IPW-estimator being superior to the OLS estimator, while the opposite holds for the points below the diagonal.

The overall finding is that the points are rather close to the diagonal indicating only small differences between the two different estimation approaches.

Fig. 8: Comparison of OLS- and IPW-biases across all waves and countries[25]



25 To improve comparability, we exclouded a small number of outlyers in the figures.

5. Comparing the estimation results in parallel surveys

For two years in Germany and the United Kingdom occurred the specific situation of two similar surveys running parallel. In Germany the ongoing Socio Economic Panel, which started in 1984, run parallel in the years 1994-1996 with the newly started European Community Household Panel. After two years the German ECHP was stopped and the SOEP adjusted to mimic the ECHP as much as possible. A similar situation occurred in the United Kingdom where the British Household Panel Survey, which started in 1991, overlapped with the newly started ECHP-survey in the years 1994-1996. This unique situation renders possible the analysis of interesting questions concerning the comparability of different surveys running parallel and the effect of substituting older panels by newly started panels.

In this section we compare the estimation results of an identical earnings regression function making use of the two different data bases for Germany as well as for the United Kingdom. In the following table the detailed results of the earnings equations are given.

Table 8: Earnings equations based on two overlapping surveys

Country	Germany				United-Kingdom			
Wave	1		2		1		2	
Database	ECHP	SOEP	ECHP	SOEP	ECHP	BHPS	ECHP	BHPS
Intercept	4.7939	4.8867	4.8891	4.9728	4.6806	4.6692	4.7778	4.7974
	(35.38)	(35.42)	(32.44)	(33.68)	(35.99)	(40.91)	(30.93)	(40.85)
Age	0.1370	0.1314	0.1321	0.1267	0.1381	0.1429	0.1360	0.1370
	(17.94)	(16.94)	(15.89)	(15.68)	(19.23)	(22.74)	(16.32)	(21.54)
Age^2	-1.619E-03	-1.593E-03	-1.538E-03	-1.520E-03	-1.647E-03	-1.710E-03	-1.627E-03	-1.624E-03
	(-16.69)	(-16.31)	(-14.77)	(-15.13)	(-18.89)	(-22.11)	(-16.34)	(-21.06)
Education - third level	0.2150	0.2747	0.2330	0.2902	0.3502	0.1758	0.3345	0.1816
	(7.72)	(10.15)	(8.27)	(10.89)	(10.92)	(5.22)	(9.6)	(5.48)
Basic Education	-0.2041	-0.2258	-0.1353	-0.1406	-0.1338	-0.1361	-0.1080	-0.1289
	(-6.7)	(-6.6)	(-4.34)	(-4.03)	(-4.38)	(-3.82)	(-3.18)	(-3.68)
Unempl.time	-0.0025	-0.0040	-0.0035	-0.0043	-0.0075	-0.0090	-0.0132	-0.0074
	(-2.25)	(-5.27)	(-3.42)	(-5.98)	(-6.21)	(-3.73)	(-6.78)	(-3.47)
Married	0.0765	0.0542	0.0474	0.0693	0.1062	0.0986	0.1072	0.0835
	(2.67)	(2.02)	(1.6)	(2.62)	(3.38)	(3.54)	(3.12)	(3.06)
r2	0.40	0.36	0.35	0.33	0.36	0.39	0.38	0.37
n	1,610	1,913	1,528	1,821	1,804	1,733	1,261	1,676

All parameters, except the dummy variable indicating the person being married in the second wave of the ECHP data set for Germany, are highly significant. To ease the comparability of the parameter estimates, the following figure displays the relative parameter differences for the two waves in Germany and the United Kingdom. If the parameter difference is significant, this is indicated within the diagram (*).

For Germany we find that the estimates of the intercept as well as the age effect and the basic education are very close. The estimates based on the SOEP for the effect of third level education are about 25% higher compared to the estimates obtained from the ECHP.

The strongest differences are found for the negative effect on cumulated unemployment time. Here the parameter based on the SOEP exceeds the parameter for the ECHP data set by 60%. For the effect of being married we find the unexpected result that the differences vary between the two waves under consideration. When testing the differences between parameters for significance, no parameter differences turn out to be significant at the 5%-level.

Fig. 9: Relative differences between the ECHP (base) and the SOEP parameters

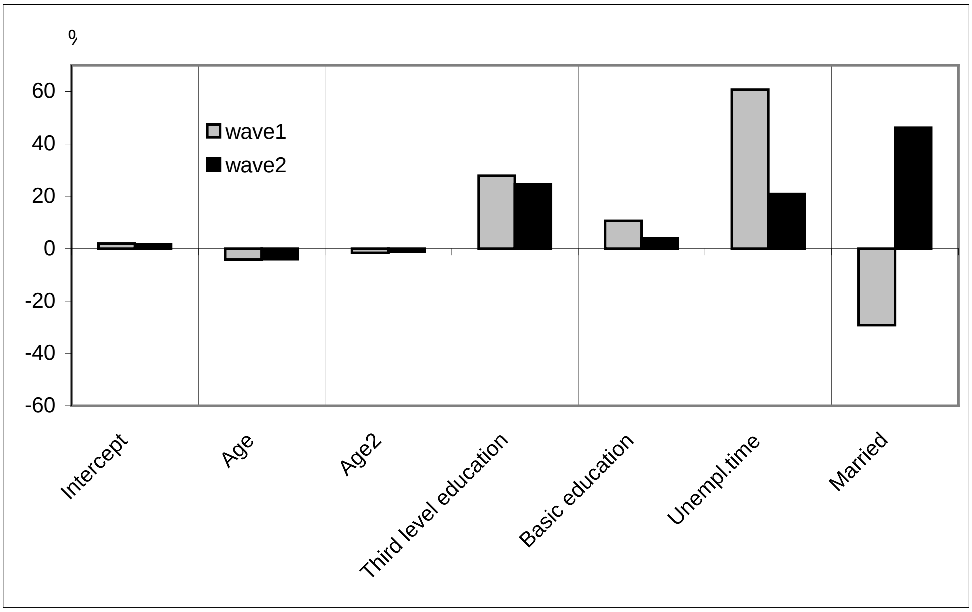
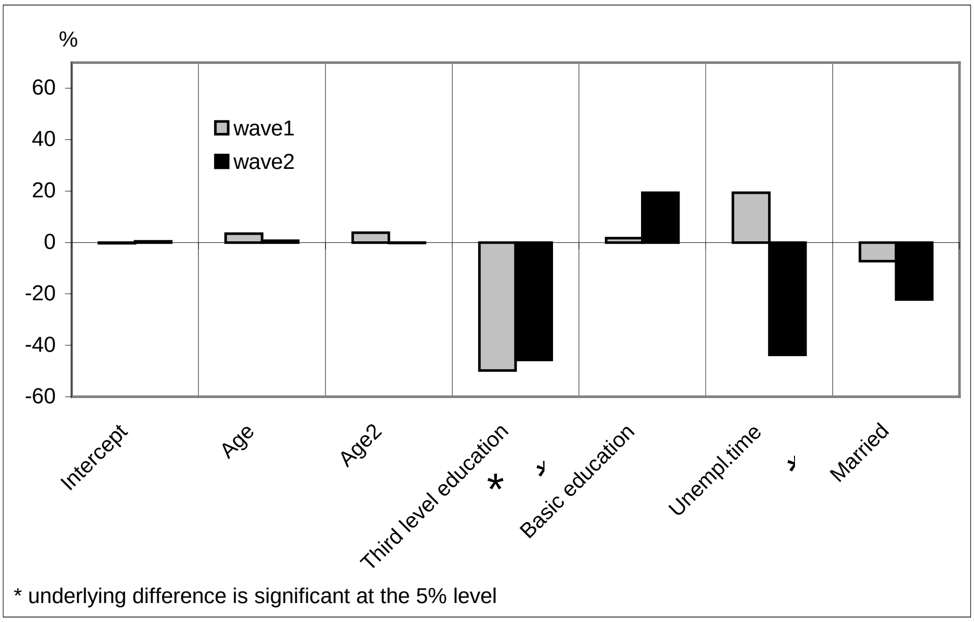


Fig. 10: Relative differences between the ECHP (base) and the BHPS parameters



Turning to the findings for the United Kingdom the empirical results resemble. While there are only minor differences in the estimated parameters found for the intercept, age and basic education, the strongest relative parameter differences show up for the effect of third level education and cumulated unemployment time. Contrary to the findings for the two German surveys running parallel, there are three significant parameter differences found between the two parallel surveys in the United Kingdom. The effects of third level education in both waves and the effects of cumulated unemployment time are statistically significant different in the two surveys.

6. Conclusion

The aim of this empirical paper is to explore the effect of attrition on earnings equations for male employees across Europe. The income equation we use is of Mincerian type based on the human capital approach of income determination. The data base underlying our research is the European Community Household Panel (ECHP) giving unique opportunity for cross country comparisons in the European Union on the level of individuals and households. Since the ECHP is plagued by panel attrition, which is documented in detail by Behr, Bellgardt and Rendtel (2002), we analyzed whether possible biases caused by non-ignorable panel attrition have to be a subject of greater concern when using the ECHP data base.

One possible solution suggested in the literature is the use of an inverse probability weighted estimator considered as correcting for attrition (Robins et al. 1995, Neukirch 2002). Concerning the question of correcting for possible attrition biases through inverse probability weighting, we conclude that the additional variance of the estimated response-probabilities outweighs possible gains from this correcting procedure and that the inverse probability weighting does not lead to an overall improvement compared to the un-weighted estimator.

Our findings indicate that the effects of attrition on income equations in general are very mild. Only few parameters were found to be estimated with significant bias when analysing 54 waves. Hence we conclude that the problem of attrition is no matter of great concern when estimating income equations of the Mincerian type based on the ECHP data.

References

Behr, Andreas/Egon Bellgardt/Ulrich Rendtel, Extent and determinants of panel attrition in the European Community Household Panel, Working Paper, URL: www.destatis.de/chintex.

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.

Garen, John (1984), The Returns to Schooling: A Selectivity Bias Approach With a Continuous Choice Variable, Econometrica, Vol. 52, No. 5, pp. 1199-1218.

Groot, Wim/Hessel Oosterbeek (1994), Earnings Effects of Different Components of Schooling; Human Capital Versus Screening, The Review of Economics and Statistics, Vol. 76, pp. 317-321.

Härdle, Wolfgang (1990), Applied Nonparametric Regression, Cambridge University Press.

Hines, Fred/Luther Tweeten/Martin Redfern (1970), Social and Private Rates of Return to Investment in Schooling, by Race-Sex Groups and Regions, The Journal of Human Resources, Vol. 5, No. 3, pp. 318-340.

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

Kroch, Eugene A./Kriss Sjoblom (1994), Schooling as Human Capital or a Signal, The Journal of Human Resources, Vol. 29, pp. 156-180.

Lillard, Lee A./Panis, Constantijn W.A. (1998), Panel Attrition from the Panel Study of Income Dynamics, The Journal of Human Resources, Vol. 33, pp. 437-457.

Mincer, Jacob (1958), Investment in Human Capital and Personal Income Distribution, The Journal of Political Economy, Vol. 66, No. 4, pp. 281-302.

Mincer, Jacob (1970), The Distribution of Labour Incomes: A Survey, Journal of Economic Literature, Vol. 8, No. 1. pp. 1-26.

Mincer, Jacob (1997), The Production of Human Capital and the Life Cycle of Earnings: Variations on a Theme, Journal of Labour Economics, Vol. 15, No. 1, pp. S26-S47.

Neukirch, Thomas (2002): Nonignorable attrition and selectivity biases in the Finnish subsample of the ECHP, Working Paper, URL: www.destatis.de/chintex.

Neumark, David/Paul Taubman (1995), Why Do Wage Profiles Slope Upward? Tests of the General Human Capital Model, Journal of Labor Economics, Vol. 13, No. 4, pp. 736-761.

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.

Robins, J./A. Rotnitzky, L. Zhao (1995), Analysis of Semi-parametric Regression Models for Repeated Outcomes in the Presence of Missing Data, Journal of the American Statistical Association, Vol. 90, pp. 106-121.

Vaillancourt, Francois (1995), The Private and Total Returns to Education in Canada, 1985, The Canadian Journal of Economics, Vol. 28, No. 3, pp. 532-554.

Weiss, Andrew (1995), Human Capital vs. Signalling Explanations of Wages, Journal of Economic Perspectives, Vol. 9, No. 4, pp. 133-154.