A Cross-Country Comparison of Survey Participation in the ECHP

Cheti Nicoletti and Franco Peracchi

ISER Working Papers
Number 2002-32

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:

Part of this paper is based on work carried out during Cheti Nicoletti's visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex, supported by the Access to Research Infrastructure action under the EU Improving Human Potential Programme. We thank Marco Bonetti, Alberto Holly, Peter Lynn, Arthur van Soest and seminar participants at DIW and Ente Einaudi for useful comments.

Readers wishing to cite this document are asked to use the following form of words:

> **Nicoletti, Cheti. and Peracchi ,Franco. (December 2002) 'A Cross-Country Comparison of Survey Participation in the ECHP ', *Working Papers of the Institute for Social and Economic Research*, paper 2002-32. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

The European Community Household Panel (ECHP) is a very ambitious project whose main purpose is to collect comparable economic, social and demographic information at the individual and the household level throughout the European Union (EU). An attractive feature of the ECHP is its comparability across countries and over time. Indeed, a great deal of effort was devoted to harmonizing the survey characteristics. Nevertheless, the design and the organization of the survey are not completely standardized across countries. This paper exploits the variability of the survey characteristics, both across countries and over time, in order to identify the effects of various features of the data collection process on survey participation.

We focus attention on survey nonparticipations caused by ineligibility, contact failure or lack of co-operation. We analyze cross-country differences in participation rates trying to disentangle the role played by differences in the socio-demographic composition of the national populations and the differential characteristics of the data collection process.

NON-TECHNICAL SUMMARY

The European Community Household Panel (ECHP) is a longitudinal survey of households and individuals, centrally designed and co-ordinated by the Statistical Office of the European Communities (Eurostat) and covering all countries of the European Union (EU). An attractive feature of the ECHP is its comparability across countries and over time. Indeed, Eurostat devoted a great deal of effort at harmonizing the survey characteristics. Nevertheless, the design and the organization of the survey are not completely standardized across countries. The main differences have to do with the sampling frame, the sampling procedures, the following rules, the adaptation of the questionnaire and the field operations. The aim of this paper is to exploit some of these differences in order to identify the effects of survey design and organization on the probability of survey participation.

Our analysis of survey participation in the ECHP shows that the three main causes of nonparticipation are, in order, refusal to co-operate, contact failure and ineligibility. Looking at the patterns of survey participation, we find that monotone participation patterns (mainly attrition) are much more frequent than irregular response patterns. Furthermore, entry into the panel is mainly due to ineligibility, while exit is mainly due to contact failure and refusal to co-operate.

We pay special attention to contact failure and refusal to co-operate, and how the probability of these events is affected by the characteristics of the data collection process on the one hand and the socio-demographic composition of the national populations on the other hand. Two different types of analysis are conducted. The first is an aggregate analysis of participation rates by region and wave as a function of region- and time-specific characteristics of the survey. The second uses individual-level information to predict participation in the next wave given participation in the current wave. The first type of analysis focuses on the differences in the response process between the first and the following waves, the second focuses instead on individual response after the first wave.

From the second wave onward, interviewees can use past experience to decide whether the cost of survey participation is too high, while survey organizers can use past experience to improve the data collection process. One would therefore expect response in the initial wave to be different from later waves. We indeed find that co-operation in the first wave is statistically different from the following waves, but we find no evidence of this for contact. When we try to predict future survey participation based on current information, we find that several individual and household characteristics have good predictive power. In particular, the number of children, the length of residence at the current address, home ownership, household income, and the index of nonresponse to household income are good predictors of future contact, whereas age, schooling level, labor force status, living in a couple and frequent contacts with the neighbors are good predictors of future co-operation given contact. We also find that several characteristics of the data collection process in the current wave help predict survey participation in the next wave. In particular, the number of contacts, the length of the fieldwork, the length of the household interview and the duration of the survey significantly affect the probability of future contact, whereas the interview mode and the use of the same interviewer across waves significantly affect the probability of future co-operation given contact. We think that this finding has important consequences for the specification and estimation of regression models with sample selection, where identification is typically achieved through exclusion restrictions, because it provides a justification for including variables characterizing the data collection process in the model for sample participation while excluding them from the model for the outcome variable of interest. Finally, we find that the ease of contact and the propensity to co-operate are negatively correlated even after controlling for a broad set of explanatory variables. This result may be important when it comes to evaluate the impact of an increased effort to contact people on the final response rate.

A cross-country comparison of survey participation in the ECHP*

Cheti Nicoletti
ISER, University of Essex

Franco Peracchi
University of Rome "Tor Vergata"

June 2002
Revised: December 2002

Abstract

The European Community Household Panel (ECHP) is a very ambitious project whose main purpose is to collect comparable economic, social and demographic information at the individual and the household level throughout the European Union (EU). An attractive feature of the ECHP is its comparability across countries and over time. Indeed, a great deal of effort was devoted to harmonizing the survey characteristics. Nevertheless, its design and the organization are not completely standardized across countries. This paper exploits the variability of the survey characteristics, both across countries and over time, in order to identify the effects of various features of the data collection process on survey participation.

We focus attention on survey nonparticipations caused by ineligibility, contact failure and lack of cooperation. We analyze cross-country differences in participation rates trying to disentangle the role played by differences in the socio-demographic composition of the national populations and the differential characteristics of the data collection process.

* Part of this paper is based on work carried out during Cheti Nicoletti's visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex, supported by the Access to Research Infrastructure action under the EU Improving Human Potential Programme. We thank Marco Bonetti, Alberto Holly, Peter Lynn, Arthur van Soest, Ernesto Villanueva and seminar participants at DIW, Ente Einaudi and Universitat Pompeu Fabra for useful comments.

1 Introduction

The European Community Household Panel (ECHP) is a longitudinal survey of households and individuals, centrally designed and co-ordinated by the Statistical Office of the European Communities (Eurostat) and covering all countries of the European Union (EU). An attractive feature of the ECHP is its comparability across countries and over time. Indeed, Eurostat devoted a great deal of effort at harmonizing the survey characteristics. Nevertheless, its design and the organization are not completely standardized across countries. The main differences have to do with the sampling frame, the sampling procedures, the following rules, the adaptation of the questionnaire and the field operations. The aim of this paper is to exploit some of these differences in order to identify the determinants of survey participation, in particular the effects of survey design and organization.

There are several reasons why this problem is important. First, estimates of the probability of survey participation play a key role in reweighting procedures for (nearly) unbiased estimation of population means and totals (see e.g. Särndal, Swenson and Wretman, 1992). Extensions of these methods, based on the propensity score, have recently been considered by Robins and Rotnitzky (1995), Robins, Rotnitzky and Zhao (1995) and Abowd, Crépon and Kramarz (1997) for the estimation of conditional means in the presence of missing data. Second, estimates of a linear predictor (or "index") of survey participation are crucial in the construction of two-step estimators of regression models with sample selection, as well as in the joint estimation of a regression model and of a response probability model (see Heckman, 1979, and the recent review article by Vella, 1998). In both cases, the main question is how to carry out valid inference about population parameters of interest when the available data are subject to unit (or item) nonresponse. A third reason for studying the determinants of sample participation is the relevance of the issue at the survey design stage, where resources have to be allocated between the possibly conflicting goals of increasing precision of estimation and reducing nonresponse biases.

This paper is organized in two parts. In the first part (Section 2), we give a description of the ECHP, pointing out the differences in survey design and organization across countries and over time, and we define and examine survey participation. We are especially interested in analyzing the relative importance of the different types of nonparticipation in a single wave (ineligibility, contact failure, refusal to cooperate) and the patterns of survey participation across waves.

In the second part (Section 3), we study the variability of survey participation rates across countries and waves. We pay special attention to contact failure and refusal to cooperate, and

how the probability of these events is affected by the characteristics of the data collection process on the one hand and the socio-demographic composition of the national populations on the other hand. Two different types of analysis are conducted. The first is an aggregate analysis of survey participation rates by region and wave as a function of region- and time-specific characteristics of the survey. The second uses individual-level information to predict participation in the next wave given participation in the current wave. The first type of analysis focuses on the differences in the response process between the first and the following waves, the second focuses instead on individual response after the first wave.

We model the response process as the outcome of two sequential events: (i) the contact between the interviewer and an eligible interviewee, and (ii) the cooperation of the interviewee. As a result, conditional on eligibility, the response process is completely described by two elements: the probability of contact and the probability of cooperation given contact. Groves and Couper (1998) and Lepkowski and Couper (2002) assume independence between these two events after conditioning on a set of observables. This assumption is somewhat restrictive, especially if one is interested in forecasting the effect of changes in the characteristics of the data collection process on future response probabilities. In this paper we consider a more general model that allows for dependence between the ease of contact and the propensity to cooperate taking into account the censoring problem caused by the fact that we know if an individual is respondent only if she has been contacted.

2 Description of the ECHP

This section describes the ECHP and the main differences in survey organization across countries and over time. In particular, Section 2.1 describes the target population and the country coverage. Section 2.2 analyzes the main differences in the survey design and the data collection process across countries. We then focus attention on survey nonparticipation, especially ineligibility and unit nonresponse caused by contact failure or lack of cooperation. Section 2.3 analyzes participation in a single wave of the survey, while Section 2.4 analyzes the patterns of participation across waves.

2.1 Target population and country coverage

The target population of the ECHP consists of all individuals living in private households within the EU.[1] In its first (1994) wave, the ECHP covered about 60,000 households and 130,000 individuals aged 16+ in twelve countries of the EU (Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and the UK). Austria, Finland and Sweden began to participate later, respectively from the second, third and fourth wave.

The ECHP distinguishes between sample and nonsample persons. Sample persons are all individuals belonging to the national samples drawn from the target population in the first wave. Sample persons also include children, born after the first wave, that have at least one parent sample person. Nonsample persons are all other individuals.

Sample and nonsample persons may or not be eligible for interview in a given wave. Sample persons are eligible if they are aged 16 or older and belong to the target population (that is, they live in a private household within the EU). Nonsample persons are eligible if, in addition, they live in a household containing at least one sample person. We classify the different causes of ineligibility into two mutually exclusive categories: natural demographic events and all other causes, which we lump together into the single category "out of scope". A sample person who is "out of scope" (that is, homeless, institutionalized or outside of EU) is "traced" and interviewed again if she returns to the target population.[2] Ineligible nonsample persons are not traced.

In Belgium and the Netherlands, the ECHP was linked from the beginning to already existing national panels, namely the Panel Study of Belgium Households (PSBH) and the Dutch Socio-Economic Panel (ISEP). In Germany, Luxembourg and the UK, instead, the first three waves of the ECHP ran parallel to existing national panels with similar content, namely the German Social Economic Panel (GSOEP), the Luxembourg's Social Economic Panel (PSELL) and the British Household Panel Survey (BHPS). Starting from the fourth (1997) wave, the ECHP data for Germany, Luxembourg and the UK have also been derived from the existing national panels. To allow comparisons across waves, the information in the earlier waves of the GSOEP and the BHPS has been harmonized according to the ECHP standards. For Luxembourg, this operation is not yet implemented in the 2002 User Data Base (UDB) of the ECHP, and comparable data sets for the first three waves will be made available only in the next release of the data.[3]

[1] Some specific areas of the EU (some small islands in Italy and the UK, DOM-TOM in France, and Ceuta and Melilla in Spain) are excluded from the target population.

[2] The exception is when a person is institutionalized or moves outside the EU and there is no information available for tracing.

[3] The UDB is an anonymized and user-friendly version of the data. The first release of the UDB, covering waves

In conclusion, in the 2002 UDB:

- for Germany and the UK, there are two data sets for the first three waves, one from the original German and British ECHP and the other from the national panel (GSOEP and BHPS respectively), while for the fourth and fifth waves there is only the comparable data set from the national panels;
- for Luxembourg, there is only the ECHP data set for the first three waves and the comparable data set from the PSELL for the fourth and fifth waves;
- for Belgium and the Netherlands, the ECHP was linked to the existing national panels (PSBH and ISEP respectively) from the first wave;
- for Austria, the ECHP data are available only from the second wave;
- for Finland, the ECHP data are available only for the third and fourth waves;
- for Sweden a comparable ECHP data set, derived from the Swedish Living Conditions Survey (ULF) is available only from the fourth wave;
- for the remaining seven countries (Denmark, France, Greece, Ireland, Italy, Portugal, Spain), the ECHP survey run independently of existing national surveys and data are available for all five waves.

2.2 Survey differences across countries

This section summarizes the main differences in the design and organization of the survey across countries and waves, focusing on the survey characteristics which are relevant for understanding cross-country differences in survey participation.[4]

Unfortunately, the data available in the 2002 UDB to analyze survey participation are incomplete for some of the countries. In particular:

1. no information is available on ineligibility in the first wave;
2. for Germany-ECHP, it is not known in which month the interview was carried out;

1 and 2, was issued by Eurostat in December 1998, three years after completion of fieldwork for wave 2. The second release, covering the first three waves, was issued in December 1999. The third one, covering waves 1–4, was released in June 2001. The fourth one, covering waves 1–5, was released in February 2002 and is the data set used in this paper. For a description of the ECHP UDB we refer to Eurostat (1999).

[4] We refer to Peracchi (2002) for a more complete review of the ECHP.

3. for Netherlands, the duration of the household and personal interviews is not available and one cannot identify those who are ineligible;

4. for Spain, it is not known whether the interview was carried out by the same interviewer as the previous wave;

5. for Sweden, no information is available on the data collection process;

6. for the UK-BHPS, one cannot distinguish between contact failure due to absence of the person and other reasons.

Cross-country differences in the ECHP partly arise from the fact that the national panels did not all start in the same year. This is important because, in general, nonresponse is not independent of the duration of the survey. Lepkowski and Couper (2002) argue that the "nonresponse process in later waves of panel surveys differs in important ways from cross-sectional surveys or the initial wave". This reflects both the self-selection of the sample units and the increasing information and organizational experience at each successive wave.

Table 1 reports, for each country, the name of the survey used for the ECHP, the year when it began, the year of the last wave available in the ECHP, and the total number of waves available until 1998. Table 2 reports country-specific averages of several variables that characterize the data collection process: the number of visits to the household, the fraction of cases in which the interviewee was contacted by the same interviewer as the previous wave, the duration of the household and personal interviews (in minutes), and the length of the fieldwork (measured by the number of months between the first and the last household interview). For comparability reasons, we exclude from the analysis Austria, Finland and Sweden, for which participation began later than all other countries. We also exclude the ECHP data for Germany, Luxembourg and the UK, because they are only available for the first three waves. For Germany and the UK we instead consider the comparable datasets obtained from the original national panels (GSOEP and BHPS).

The average number of visits ranges from a minimum of 1.14 in Greece to a maximum of 2.98 in Denmark, and is generally lower in Southern European countries. This variability across countries may reflect both a different organization of the callbacks and differences in the ease of contact. In particular, a high average number of visits may signal contact difficulties (Lynn *et al.* 2002). The percentage of cases in which the same interviewer has been used to contact a given household[5] is

[5] The percentage is computed using people with records in two consecutive waves.

lowest in Greece and Portugal (less than 50 percent) and highest in Belgium, Germany, Netherlands and the UK, where the ECHP data are collected through the already existing national panels.

In general, personal interviews tend to last longer than household interviews. There are several exceptions, however. The GSOEP and the BHPS adopt a short household questionnaire and a relatively long personal one. In France, on the contrary, the household questionnaire is quite time-demanding, while the personal questionnaire is short. In Belgium both questionnaires are quite long, while in Greece, Italy and Spain they are both quite short. The average duration of the household interview ranges between a minimum of 11 minutes in the UK and a maximum of 28 minutes in France. The average duration of the personal questionnaire ranges instead between a minimum of 18 minutes in France and a maximum of 47 minutes in Belgium. Finally, fieldwork lasts on average more than one year in Ireland, but only three months in France and Spain.

Table 3 reports the relative importance of five different interview modes, namely pencil-and-paper face-to-face personal interview (PAPI), computer-assisted face-to-face personal interview (CAPI), self-administered by respondent, telephone interview, and proxy interview. In general, the most common interview mode is the traditional PAPI. The main exceptions are Greece, Netherlands and Portugal, where the most common interview mode is CAPI, and Belgium where it is instead self-administered questionnaire by respondents. The latter interview mode is also quite important in Germany. The percentage of telephone interviews is usually very low, and only reaches 1 percent in Italy and Spain. Proxy interviews are less than 2 percent in most countries, except France, Ireland, Italy, Portugal and Spain where they represent 10 percent or more.

2.3 Survey participation in a single wave

We say that an individual does not participate in a given wave of the panel if she is ineligible in that wave or is a unit nonrespondent. Unit nonresponse occurs when an eligible individual (that is, aged 16+ and living in a private household within the EU) fails to return the personal questionnaire. There are two broad reasons for unit nonresponse: one is contact failure, due to absence of the person or other reasons, the other is lack of cooperation. In the ECHP, unit nonrespondents are followed up in the next wave, except when nonresponse is due to incapacity or refusal to return a questionnaire that is considered as "final". If contact failure or lack of cooperation of all household members persists for two consecutive waves, then the entire household is dropped from the survey.

To identify the various causes of nonparticipation we use the age of the person and two variables in the longitudinal link file of the UDB: (i) the personal residential status, which gives details on

whether a person is in scope or out of scope (except for the first wave), and (ii) the personal interview result, which specifies whether a person has an interview completed or not completed for some reasons, or she has not been contacted.[6]

We classify the causes of nonparticipation as follows:

1. Natural demographic events: death or 16th birthday.

2. Movement from in to out of scope of the survey, or viceversa. It includes institutionalization, migration to a foreign country,[7] movement of a nonsample person to a household without sample individuals, etc.

3. Lack of cooperation (refusal to respond): it includes definite or temporary refusal to participate, individual unable to respond because of physical or language problems, and failure to return a self-completed questionnaire.

4. Absence of the person at the address.

5. Other types of contact failure: it includes the case of incomplete number of callbacks or interview not attempted for some reason, person omitted by error, inability to contact the person because address non residential or non existent, inability to locate the address, or other reasons.

Notice that if the interviewer neither succeeds in contacting a person nor obtains information from relatives, neighbors or other sources, then the person is considered eligible by the ECHP. Thus, the category contact failure may also include people that are no longer eligible.

The distinction between nonparticipation due to ineligibility (categories 1 and 2) and unit nonresponse (categories 3, 4 and 5) is very important for inference. Changes in eligibility essentially

[6] To avoid misunderstanding, we give details on how we handle the missing or not applicable cases for these two variables. For any specific wave, we exclude cases with residential status "not applicable". These cases correspond to nonsample people who were dropped from the sample in a previous wave because out of scope, people who died in a previous wave, people who only become eligible in a future wave, people belonging to households not interviewed for two consecutive waves, and all the other people who are not forwarded from the previous wave because of the rules for follow-up. We instead include people who are in scope but with interview result "not applicable". These are persons born before 1994 but younger than 16, hence ineligible for the personal interview. We report as a separate nonparticipation cause those who are in scope but with interview result missing. Finally we exclude the very few cases in which the residential status is missing.

[7] In principle, people moving to another country within the EU remain in the scope of the survey. However, because the follow up is difficult and is successful only in a few cases, we decided to classify these movements together with movements to a non-EU country.

reproduce the dynamics of the target population, while changes in the response status may create a problem of self-selection of the responding sample.

Unfortunately, the UDB does not provide adequate information on ineligibility and unit nonresponse in the first wave. An indication of the magnitude of the problem is given by Table 4, which shows household response rates in the first three waves of the ECHP as computed by Eurostat (1997).[8] The overall household response rate in the first wave is 72 percent, and is comparable to that observed in the initial wave of other panel surveys (Peracchi, 2002). More striking, however, is the large variation across countries. The low response rates in Germany and Luxembourg largely reflect outright refusal to participate. On the other hand, the high response rates in Greece and Italy may reflect the fact that survey participation is compulsory in these two countries, whereas in Belgium and the Netherlands they may reflect the fact that the first wave of the ECHP corresponds in these two countries to later waves of existing national panels. As is typical with household panels, response rates in later waves of the ECHP tend to be higher than the initial ones, often notably as in the case of Germany and Luxembourg.

Table 5 reports the fraction of nonparticipants by country and wave (except the first wave), that is, the ratio between the number of nonparticipants and the number of people included into the sample (participants plus nonparticipants). The fraction of nonparticipants is below 30 percent in all countries except Ireland and the UK. The high fraction of nonparticipants in Ireland is mainly due to ineligibility, while in the UK it is mainly due contact failure and lack of cooperation (see below). We observe that in the countries where the ECHP is derived from already existing national panels (Belgium, Germany, Netherlands and the UK), the fraction of nonparticipants tends to be higher, mainly because of the higher incidence of failed contact, lack of cooperation, and missing interview result.

2.4 Patterns of survey participation

The analysis in the previous section refers to survey participation in a single wave. We now consider the patterns of survey participation of people who participate in at least one wave of the UDB.[9]

Let D_j be a 0–1 indicator of survey participation in wave j ($D_j = 1$ for survey participants).

[8] Household response rates are defined as the ratio of the number of interviewed households to the target number for interview. For the first wave, the latter is just the number of households selected into the sample, excluding the cases which turned out to be nonexistent or otherwise ineligible. For the second and third waves, it is the number of households forwarded from the previous wave, minus those no longer existing, plus the newly formed ones.

[9] The UDB also contains a small fraction of persons who never participate. This category consists mainly of: (i) sample persons aged less than 16 in all five waves, (ii) sample persons with a definite refusal to cooperate, and (iii) sample persons who were unit nonrespondents in the first wave and then became impossible to contact.

Since 5 waves of the survey are currently available, a participation pattern is described by the 5-dimensional vector $D = (D_1, D_2, D_3, D_4, D_5)$. Thirty-one ($31 = 2^5 - 1$) participation patterns are possible, which we classify into six categories:

1. always responding: $D = (1, 1, 1, 1, 1)$;
2. monotone attrition: $D = (1, 0, 0, 0, 0)$, $D = (1, 1, 0, 0, 0)$, $D = (1, 1, 1, 0, 0)$ or $D = (1, 1, 1, 1, 0)$;
3. new entry: $D = (0, 1, 1, 1, 1)$, $D = (0, 0, 1, 1, 1)$, $D = (0, 0, 0, 1, 1)$ or $D = (0, 0, 0, 0, 1)$;
4. occasional nonresponse: $D = (1, 0, 1, 1, 1)$, $D = (1, 0, 1, 1, 1)$, $D = (1, 0, 0, 1, 1)$, $D = (1, 1, 0, 0, 1)$ or $D = (1, 1, 1, 0, 1)$;
5. occasional response: $D = (0, 1, 0, 0, 0)$, $D = (0, 1, 1, 0, 0)$, $D = (0, 0, 1, 1, 0)$, $D = (0, 0, 1, 0, 0)$ or $D = (0, 0, 0, 1, 0)$;
6. very irregular response: all other participation patterns.

We say that a pattern is monotone if D_j changes value only once. Monotone participation patterns are the first three cases.[10] We say that a participation pattern is irregular if D_j changes value more than once, as in the last three categories of our classification, and is very irregular if it changes value more than twice.

Table 6 compares participation patterns across countries. We focus on the eleven countries for which we have comparable data covering all five waves. Thus, we exclude Austria, Finland, Luxembourg and Sweden, whereas for Germany and the UK we only consider the comparable dataset obtained from the GSOEP and the BHPS. The pattern (1,1,1,1,1) is the most frequent. It represents 55 percent or more in all countries, except Denmark, Ireland and Spain where the percentage is somewhat lower because of the higher frequency of monotone attrition. For Belgium, Netherlands, Germany and the UK, the relatively high frequency of the pattern (1,1,1,1,1) may simply reflect the fact that the data are derived from already existing national panels.

Table 7 shows the relative importance of the different causes of nonparticipation for each type of pattern. For individuals who enter in wave t, we consider why they did not participate in wave $t - 1$, whereas for people who exit in wave t we look at the cause of attrition in that wave. For the occasional participation patterns, we report the causes of both attrition and entry. For

[10] This classification is based on the available information. Whether an individual is a stable participant or not can only be determined after the last wave of the survey.

the occasional nonparticipation patterns, we report the causes of re-entry. For the very irregular participation patterns, we only report the causes of last entry.

Quantitatively, monotone attrition is much more important than new entry. Moreover, new entry is mainly linked to eligibility (turning 16 years old and movements from out to in scope), while monotone attrition is mainly due to contact failure or lack of cooperation. Lack of cooperation is especially important among the very irregular response patterns. For the occasional nonresponse patterns, the main cause of nonparticipation is instead contact failure (41.5 percent of the cases excluding absence). For the occasional response patterns, exit is mainly due to contact failure (59.3 percent of the cases excluding absence), while entry is mainly due to people moving from out to in scope of the survey (58.9 percent of the cases).

3 Modeling survey participation by country and wave

Because the ECHP is not completely harmonized across countries, the observed differences in survey participation may reflect, at least partly, differences in the data collection process. They may also reflect differences in the composition of the national populations along dimensions that are correlated with the survey participation decision. For example, different population age structures may help explain the observed differences in the importance of ineligibility, whereas the propensity to cooperate may be linked to personal characteristics such as gender, age, schooling attainments and income. Because contact may be harder for people who move frequently or live alone, labor force status and living in a couple may be important predictors of contact failure.

To investigate the role played by the characteristics of the data collection process and the socio-demographic composition of the population, we carry out two different types of analysis. The first is an aggregate analysis of the relationship between survey participation rates and average characteristics of a region or a wave (Section 3.2). The second is a micro-level analysis that uses the information on the respondents in the current wave to predict survey participation in the next wave (Section 3.3).

3.1 Predictors of survey participation

In what follows, we relate survey participation to five sets of variables:

1. variables describing the data collection process;

2. variables representing the age structure of the population;

3. variables describing other socio-demographic characteristics of the population;

4. country dummies to capture time-invariant unobserved heterogeneity across countries;

5. year dummies to capture country-invariant time effects.

The set of variables describing the data collection process consists of the number of visits to the household (`nvisits`), the duration of the household and the personal interviews (`hminint` and `pminint`), an indicator of whether the interview was carried out by the same interviewer as in the previous wave (`pintid`), indicators for the interview mode, namely computer-assisted face-to-face personal interview (`CAPI`), self-administered by the respondent (`self`), telephone interview (`tel`), and proxy interview (`proxy`), the length of the fieldwork (`tfieldw`), and the panel duration (`duration`), that is, the number of years since the beginning of the (pre-existing or ECHP) national panel. These variables are expected to help predict contact failure and lack of cooperation, but not ineligibility.

The population age structure is expected to be an important predictor of ineligibility. It is represented by a set of age groups in the aggregate analysis of Section 3.2, and by a quadratic polynomial in age in the micro-level analysis of Section 3.3.

Turning to the variables describing the other socio-demographic characteristics of the populations, the number of children in the household (`children`) is likely to have the most direct impact on eligibility. The variables which are more likely to affect the propensity to participate include gender (represented by a dummy for females), indicators for the level of schooling (`college` and `secondary`), labor force status (`unemployed` and `inactive`), not living in a couple (`nocohab`) and whether the person infrequently talks to the neighbors (`nosocial`), the level of equivalized household income (`hincome`) and an index of item nonresponse to household income (`itemnr`). Because variables which are highly correlated with geographical mobility are likely to help predict non-participation, we also consider the number of adults (`adults`) and the number of children in the household, the number of years of residence at the current address (`tmove`), and an indicator for not owning home (`nowner`).

In the aggregate analysis of Section 3.2, we look at the relationship between the average values of these variables by region and wave and aggregate participation rates. In the micro-level analysis of Section 3.3, these variables are instead used to predict future survey participation of those responding in the current wave.

The role of some of these variables may be different at the aggregate and the individual level. For example, a higher percentage of telephone interviews may indicate a greater effort in contacting people and may be linked positively to the contact and cooperation rates. On the other hand, the fact that a person has been contacted by telephone in the last wave may indicate reluctance to cooperate in the current wave. The number of children in a household is another example. The average number of children by region and wave is an indicator of the fraction of people who are ineligible, while the number of children in a household may be an indicator of ease of contact, as households with children may have a higher probability of someone being always at home.

3.2 An aggregate analysis of survey participation

In this section, we consider three types of survey nonparticipation: ineligibility, contact failure and refusal to respond (lack of cooperation). For each of them, we sequentially define a participation rate by only considering people at risk of that specific type of nonparticipation. We define the eligibility rate as the ratio between the number of eligible people and the number of people selected into the sample.[11] We define the contact rate as the ratio between the number of eligible people who were contacted and the number of eligible people. Finally, we define the cooperation rate as the ratio between the number of contacted people who cooperated and the number of contacted people.[12]

Our basic statistical model is a grouped probit model relating survey participation rates by region and wave to the characteristics of the data collection process and to various socio-demographic indicators. Except for the country and year dummies, all predictors are averages by region and wave using all the available data.[13] This is justifiable as long as the characteristics of the data collection process and the socio-demographic characteristics of the population are relatively similar between the respondents and the nonrespondents. We focus on the countries for which we have comparable data covering at least four waves. After dropping the countries where the information on the region of residence is unavailable (Denmark and Netherlands), we are left with ten countries (Austria, Belgium, France, Germany-GSOEP, Greece, Ireland, Italy, Portugal, Spain and UK-BHPS).

[11] Calculations exclude the first wave because the UDB does not include any information on those who are ineligible in this wave.

[12] For a small fraction of people in each wave (about .5 percent), the interview is incomplete for reasons that are unknown. Because these people are eligible, we consider them as contacted.

[13] For Spain we do not know whether the same interviewer is used across waves, and for the first three waves of the BHPS and all waves of the GSOEP we do not know the frequency with which people talk to neighbors. In all these cases we imputed the missing variables using the EU average.

Table 8 presents the estimated parameters of various grouped probit models of survey participation. The first column of the table corresponds to the eligibility rate after the first wave. The next three columns correspond to the contact rate. The second and third columns provide a comparison of the results obtained respectively excluding and including the first wave. In both these cases, we leave out Belgium, Germany-GSOEP and UK-BHPS (for which no information on the first wave is available in the UDB of the ECHP, since these panels started before 1994). We also leave out the year dummies and the information on the frequency of use of the same interviewer. The estimates in the fourth column are based on all ten countries and include all covariates. We proceed in a similar way for the cooperation rate given contact (the last three columns of the table).

Since all variables (except survey duration and the country and year dummies) are expressed as deviations from the Italian average, the intercept of each model is directly interpretable as the inverse probit transform of the probability of survey participation in the 1998 wave for Italy, the country with the largest sample size.

Unlike gender, the composition of the population by age appears to have a statistically significant effect on survey participation. In particular, the fraction of people below age 16 has a strong negative impact on eligibility, whereas the fraction of young (aged 16–25) and old (aged 65+) people has a negative impact on contact.

Some characteristics of the data collection process and some socio-economic characteristics of the population appear to help explain the variability in response rates, but the sign and the magnitude of the coefficients are often difficult to interpret.

The size and variability of the country effects indicate a considerable amount of unobserved heterogeneity. In other words, there are unmeasured features associated with each country (other than the characteristics of the data collection process and the socio-demographic characteristics considered in our regressions) that are important in explaining the observed differences in survey participation rates. This is especially true for the cooperation rates. The year effects, on the other hand, are not particularly significant and show no clear pattern.

A comparison of the residual sums of squares (RSS) at the bottom of the second and the third columns provides a simple test of the hypothesis of stability of the contact rates between the first and the other waves, as the second columns excludes the first wave which is instead included in the third. A similar test of stability of the cooperation rates may be based on a comparison of the residual sum of squares at the bottom of the fifth and sixth columns.[14] The p-values of these tests

[14] Because the ineligible people are not included in the first wave of the UDB, we cannot test for stability of the eligibility rate.

are .684 for the contact rate and .002 for the cooperation rate. These findings suggest that lack of cooperation in the first wave of a panel is statistically different from the following waves, whereas lack of contact tend to be more similar between the first and the following waves.

We would like to emphasize the limitations of this aggregate analysis. First, it cannot account for individual heterogeneity within the same region and wave. Second, because our predictors are averages by region and wave over responding individuals, we are implicitly assuming that the average characteristics of the respondents are unbiased estimators of the corresponding population averages. Both limitations can be overcome by using micro-level data on responding individuals and specifying a model for their future participation based on the information currently available. The results of this type of micro-level analysis are shown in the following section, where we again focus attention on nonparticipation due to contact failure and refusal to respond.

3.3 A micro-level analysis of survey participation

The purpose of this section is to investigate cross-country differences in survey participation using the information on eligible individuals who are survey participants in wave t to predict their participation in wave $t+1$. Since we confine attention to eligible people, participation and response are equivalent events.

The response process may be described as the outcome of two sequential events: (i) the contact between the interviewer and the interviewee, and (ii) the cooperation of the interviewee. If Y_1 is the indicator of the event that a currently responding person is contacted in the next wave and Y_2 is the indicator of the event that the person cooperates then, conditional on response in the current wave, the response process in the next wave is completely described by two elements: the probability of future contact, $\pi_1 = \Pr\{Y_1 = 1\}$, and the probability of future cooperation given contact, $\pi_{1|1} = \Pr\{Y_2 = 1 \,|\, Y_1 = 1\}$. Table 9 presents estimates of these probabilities by country and year, along with estimates of the probability of response in the next wave, $\pi_{11} = \Pr\{Y_1 = 1, Y_2 = 1\}$. Response probabilities in the next wave (conditional on response in the current wave) are high, but tend to decline over time, especially from the 1995 wave. In some countries (France, Greece and Portugal) this is mainly due to an increasing difficulty in contacting people, whereas in other countries (Ireland and Italy) it is mainly due to an increasing difficulty in obtaining cooperation from contacted people.

Groves and Couper (1998) and Lepkwoski and Couper (2002) assume independence between contact and cooperation after conditioning on a set X of observable covariates, that is, they assume

that $\Pr\{Y_2 = 1 \,|\, Y_1 = 1, X\} = \Pr\{Y_2 = 1 \,|\, X\}$. This conditional independence assumption is restrictive because it ignores correlation arising from omitted individual or survey characteristics that may affect both the probability of contacting people and the probability to cooperate. It also ignores the correlation induced by incorrect classification of the different causes of survey nonparticipation.

The alternative considered in this paper is the bivariate probit model

$$\begin{aligned} Y_j^* &= X_j^\top \beta_j + U_j, \\ Y_j &= 1\{Y_j^* > 0\}, \qquad j = 1, 2, \end{aligned} \tag{1}$$

where Y_1^* and Y_2^* are two latent continuous random variables, representing respectively the ease of contact and the propensity to cooperate, $1\{A\}$ is the indicator function of the event A, X_j is a k_j-vector of covariates, β_j is a k_j-vector of unknown parameters, and the regression errors U_1 and U_2 are distributed independently of the covariates according to a bivariate Gaussian distribution with zero means, unit variances and correlation coefficient ρ.[15] The vector of model parameters is $\theta = (\beta_1, \beta_2, \rho)$. The parameter space is $\Theta = \Re^k \times (-1, 1)$, with $k = k_1 + k_2$. This model nests the conditional independence model as a special case corresponding to $\rho = 0$ but allows for unrestricted patterns of dependence between the ease of contact and the propensity to cooperate.

Under the assumption that the data $\{(X_{i1}, X_{i2}, Y_{i1}, Y_{i2}), i = 1, \ldots, n\}$ are a random sample from the joint distribution of (X_1, X_2, Y_1, Y_2) and taking into account the fact that cooperation (or lack thereof) is only observable for those who have been contacted ($Y_{i1} = 1$), we obtain the sample log-likelihood

$$L(\theta) = \sum_{i=1}^{n} \left[Y_{i1} Y_{i2} \, \ln \pi_{i11}(\theta) + Y_{i1}(1 - Y_{i2}) \, \ln \pi_{i10}(\theta) + (1 - Y_{i1}) \, \ln(1 - \pi_{i1}(\theta)) \right], \tag{2}$$

with

$$\begin{aligned} \pi_{i11}(\theta) &= \int_{-\mu_{i1}}^{\infty} \Phi\left(\frac{\mu_{i2} + \rho u}{\sigma} \right) \phi(u) \, du, \\ \pi_{i10}(\theta) &= \int_{-\mu_{i1}}^{\infty} \left[1 - \Phi\left(\frac{\mu_{i2} + \rho u}{\sigma} \right) \right] \phi(u) \, du, \end{aligned}$$

and $\pi_{i1}(\theta) = \Phi(\mu_{i1})$, where $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the density and cumulative distribution function of the standardized Gaussian distribution and $\mu_{ij} = X_{ij}^\top \beta_j$. A maximum likelihood estimate of θ maximizes (2) over the parameter space Θ.

Within this model, testing the hypothesis of independence between ease of contact and propensity to cooperate is equivalent to testing whether $\rho = 0$. An alternative test based on the likelihood

[15] The normalization of the variances is necessary because the model parameters are only identifiable up to scale.

ratio principle is easily obtained by comparing the maximized value of the log-likelihood (2) with the maximized value of the log-likelihood for the model with conditional independence. The latter is just the sum of the log-likelihoods for two binary probit models, one for Y_{i1} and one for Y_{i2} conditional on $Y_{i1} = 1$.

Tables 10 and 11 compare the estimates obtained for three alternative specifications of the models with and without the conditional independence assumption. Table 10 presents the results for the probability of contact, whereas Table 11 presents the results for the conditional probability of cooperation given contact. As before, the estimation sample includes the ten countries for which we have comparable data covering at least four waves. The sample size consists of $n = 441,548$ individuals, of whom $415,750/441,548 = 94.1\%$ have been successfully contacted in the next wave.

The first specification (Model 1) excludes current survey features from the models for the probability of future contact and cooperation. The predictors of contact include the number of adults and children in the household, the number of years of residence at the current address, home ownership, equivalized household income, and the index of item nonresponse to household income, whereas the predictors of cooperation given contact include age, age squared, a female dummy, the level of schooling (represented by two dummies, one for completed college education and one for the secondary level of secondary education), labor force status (represented by two dummies, one for being unemployed and one for being out of the labor force), and indicators for not living in a couple and infrequently talking to neighbors.

The second specification (Model 2) ignores instead the role of household and personal characteristics. In this case, the predictors of future contact include a set of characteristics of the household interview process (the number of visits to the household, the length of the household interview, the length of the fieldwork, and the panel duration), whereas the predictors of future cooperation given contact include the length of the personal interview and a set of dummies for the personal interview mode (CAPI, self-administered, telephone and proxy) and a different interviewer from last year.

Finally, the most general specification (Model 3) includes as predictors both survey features and household and personal characteristics. All specifications also include country and year dummies in the two equations, plus a set of indicators (variables ending with `mis`) to capture the fact that, for some key predictors (education, labor force status, home ownership, cohabitational status, mode of interview, length of the interview, etc.), the fraction of missing values is nonnegligible. We always take Italy and the fourth (1997) wave as the reference.

A common aspect of all three specifications is that the socio-economic characteristics of a household and the features of the household interview process only enter the model for the probability of future contact, whereas the personal characteristics and the features of the personal interview process only enter the model for the probability of future cooperation given contact. These exclusion restrictions are mainly imposed to avoid the log-likelihood from being ill behaved. An admittedly loose justification is the assumption that failure to contact a person is largely the result of failure to contact the household to which the person belongs, whereas refusal to cooperate by a contacted person is largely a personal decision.

Most of the covariates have coefficients with the expected sign. As a result of the large sample size, most of them are also statistically significant at the 1 percent level.[16] In particular:

- the age profile of the probability of future cooperation is concave, cooperation being lower for younger and older people;
- women are more likely to cooperate than men, but the difference does not apper to be statistically significant;
- people with college (tertiary) education are more likely to cooperate than people with lower education;
- the presence of children in the household, home ownership and the level of equivalized household income are positively related to the probability of future contact, whereas the index of item nonresponse to household income is negatively related;
- being out of the labor force is positively related to the probability of future cooperation, whereas not living in a couple and infrequent interactions with the neighbors are negatively related;
- the number of visits to the household is negatively related to the probability of future contact, whereas the length of the fieldwork and the panel duration are positively related;
- the presence of the same interviewer as last year is positively related to the probability of future cooperation, whereas the use of interview modes different from PAPI (CAPI, self-administered, telephone and proxy) is negatively related;

[16] Standard errors are always obtained using the "sandwich form" of the asymptotic variance matrix.

- missing information on the length of residence at the current address, the interview mode and the presence of the same interviewer as last year are positively related to the probabilities of future contact and cooperation, whereas missing information on all other covariates is negatively related;
- time-invariant heterogeneity across countries is very important for both the probability of contact and cooperation;
- other things being equal, the probability of contact tends to decline over time, whereas the probability of cooperation does not show any clear time trend.

We find that the likelihood ratio test always strongly rejects the conditional independence assumption, although the estimated coefficients on the model covariates do not change much if conditional independence is relaxed. The estimated correlation coefficient between the two latent variables ($\hat{\rho}$) is always negative and statistically significant at the 1 percent level. Since the baseline estimates of μ_1 and μ_2 are positive, whereas the estimates of ρ are negative, the results in the Appendix imply that the predicted positive effect on the response probability $\Pr\{Y_1 = 1, Y_2 = 1\}$ of an increase in μ_1 (for example because of an increase in the number of callbacks, or more generally, an improvement in the contact process) is larger for the model without conditional independence than for that with conditional independence.

Finally, a comparison of the maximized log-likelihoods for the three specifications shows that omitting survey features has a less severe impact than omitting household characteristics in the model for the probability of future contact, but it has a more severe impact than omitting personal characteristics in the model for the probability of future cooperation given contact.

4 Concluding remarks

This paper analyzes a number of issues surrounding survey participation in household panels, with special reference to the ECHP. This final section summarizes our main findings.

The three main causes of survey nonparticipation in the ECHP are, in increasing order of importance, refusal to cooperate, contact failure, and ineligibility. Looking at the patterns of survey participation, we find that monotone participation patterns (mainly attrition) are much more frequent than irregular response patterns. Furthermore, entry into the panel is mainly due to ineligibility, while exit is mainly due to contact failure and refusal to cooperate.

From the second wave onward, interviewees can use past experience to decide whether the cost of survey participation is too high, while survey organizers can use past experience to improve the data collection process. One would therefore expect response in the initial wave to be different from later waves. We indeed find that cooperation in the first wave is statistically different from the following waves, but we find no evidence of this for contact.

When we try to predict future survey participation based on current information, we find that several individual and household characteristics have good predictive power. In particular, the number of children, the length of residence at the current address, home ownership, household income, and the index of nonresponse to household income are good predictors of future contact, whereas age, schooling level, labor force status, living in a couple and frequent contacts with the neighbors are good predictors of future cooperation given contact.

We also find that several characteristics of the data collection process in the current wave help predict survey participation in the next wave. In particular, the number of contacts, the length of the fieldwork, the length of the household interview and the duration of the survey significantly affect the probability of future contact, whereas the interview mode and the use of the same interviewer across waves significantly affect the probability of future cooperation given contact. We think that this finding has important consequences for the specification and estimation of regression models with sample selection, where identification is typically achieved through exclusion restrictions, because it provides a justification for including variables characterizing the data collection process in the model for sample participation while excluding them from the model for the outcome variable of interest.

Finally, we find that the ease of contact and the propensity to cooperate are negatively correlated even after controlling for a broad set of explanatory variables. This result may be important when it comes to evaluate the impact of an increased effort to contact people on the final response rate.

References

Abowd J., Crépon B., Kramarz F. (1997), "Moment estimation with attrition", NBER Technical Working Paper No. 214.

Eurostat (1997), "Response rates for the first three waves of the ECHP", PAN 92/97, Eurostat, Luxembourg.

Eurostat (1999), *ECHP UDB Manual. Waves 1,2 and 3*, Eurostat, Luxembourg.

Groves R.M., and M.P. Couper (1998), *Nonresponse in Household Interview Surveys*, Wiley, New York.

Heckman J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153–161.

Lepkowski J.M., and M.P. Couper (2002), "Nonresponse in the second wave of longitudinal household surveys", in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), *Survey Nonresponse*, Wiley, New York.

Lynn P., P. Clarke, J. Martin, and P. Sturgis (2002), "The effects of extended interviewer efforts on nonresponse bias", in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), *Survey Nonresponse*, Wiley, New York.

Peracchi F. (2002), "The European Community Household Panel: A review", *Empirical Economics*, 27, 63–90.

Robins J., Rotnitzky A. (1995), "Semiparametric efficiency in multivariate regression models with missing data", *Journal of the American Statistical Association*, 90, 122–129.

Robins J., Rotnitzky A., Zhao L. (1995), "Analysis of semiparametric regression models for repeated outcomes in presence of missing data", *Journal of the American Statistical Association*, 90, 106-121.

Särndal C.-E., B. Swensson, and J. Wretman (1992), *Model Assisted Survey Sampling*, Springer, New York.

Vella F. (1998), "Estimating models with sample selection bias: a survey", *Journal of Human Resources*, 33, 127–169.

Table 1: Duration of the ECHP by country.

	Survey source	First wave	Last wave	No. of waves
Austria	ECHP	1995	1998	4
Belgium	PSBH	1992	1998	7
Denmark	ECHP	1994	1998	5
Finland	ECHP	1996	1997	2
France	ECHP	1994	1998	5
Germany	ECHP	1994	1996	3
Germany-SOEP	GSOEP	1984	1998	15
Greece	ECHP	1994	1998	5
Ireland	ECHP	1994	1998	5
Italy	ECHP	1994	1998	5
Luxembourg	ECHP	1994	1996	3
Netherlands	ISEP	1984	1998	15
Portugal	ECHP	1994	1998	5
Spain	ECHP	1994	1998	5
Sweden	ULF	1984	1998	15
UK	ECHP	1994	1996	3
UK-BHPS	BHPS	1991	1998	8

Table 2: Mean of selected survey features by country.

	Number of visits	Same interviewer	Length of hh questionnaire	Length pers. questionnaire	Length of fieldwork
Belgium	2.0	0.860	22	47	8
Denmark	3.0	0.521	20	26	6
France	2.4	0.622	28	18	3
Germany-GSOEP	2.4	0.866	19	35	9
Greece	1.1	0.325	17	22	9
Ireland	2.8	0.804	20	32	13
Italy	1.7	0.549	17	19	6
Netherlands	2.0	0.878			6
Portugal	1.7	0.297	18	20	4
Spain	2.0		18	23	3
UK-BHPS	2.6	0.868	11	36	6

Table 3: Interview modes by country (as a percentage of the total).

	PAPI	CAPI	Self-adm.	Phone	Proxy	Missing
Belgium	48.4	0.0	50.9	0.0	0.4	.1
Denmark	95.8	0.0	0.9	0.1	2.3	1.0
France	53.8	0.0	0.0	0.0	9.6	40.2
Germany-GSOEP	54.9	0.0	39.8	0.1	0.0	4.0
Greece	20.9	76.7	0.4	0.4	1.6	.0
Ireland	88.2	0.0	0.0	0.0	11.8	.0
Italy	82.1	0.2	0.0	1.7	15.8	.2
Netherlands	0.0	100.0	0.0	0.0	0.0	.0
Portugal	30.4	58.4	0.5	0.1	10.6	.0
Spain	80.0	0.0	2.7	1.3	16.0	.0
UK-BHPS	99.9	0.0	0.0	0.0	0.0	.0

Table 4: Household response rates in the first three waves of the ECHP. Source: Eurostat (1997).

	Wave 1	Wave 2	Wave 3
Belgium	.844	.873	
Denmark	.624	.828	.767
France	.795	.896	
Germany	.477	.920	.955
Greece	.901	.885	.874
Ireland	.558	.818	.817
Italy	.907	.909	.907
Luxembourg	.407	.940	.900
Netherlands	.875	.889	.916
Portugal	.889	.904	.968
Spain	.670	.869	.843
UK	.716	.843	.856

Table 5: Fraction of nonparticipants (percent) by country and wave.

	Wave 2	Wave 3	Wave 4	Wave 5
Belgium	27.3	27.1	28.0	28.3
Denmark	24.3	25.2	26.3	28.8
France	24.2	24.1	24.2	25.8
Germany-GSOEP	25.3	24.9	24.5	24.4
Greece	21.6	20.7	20.1	20.6
Ireland	34.6	32.6	32.7	31.5
Italy	18.7	18.1	17.8	18.0
Netherlands	29.6	28.9	29.0	29.9
Portugal	20.7	20.9	20.0	20.0
Spain	22.3	21.7	19.7	18.9
UK-BHPS	31.2	30.5	30.2	30.3

Table 6: Participation patterns by country.

	Response	Attrition	Entry	Occas nr	Occas r	Very irreg	Total
Belgium	57.1	26.5	9.1	2.5	3.1	1.7	7726
Denmark	46.8	31.9	8.1	5.1	4.9	3.2	6885
France	58.1	26.6	8.1	2.6	3.0	1.7	16234
Germany-SOEP	63.8	16.8	12.7	2.2	2.7	1.8	14638
Greece	55.5	27.6	10.6	1.7	2.8	1.8	14660
Ireland	44.7	40.0	9.1	1.1	3.8	1.3	11458
Italy	62.4	19.5	11.0	3.3	2.2	1.7	20669
Netherlands	56.1	20.1	15.4	3.2	2.9	2.2	11721
Portugal	62.4	16.0	14.6	3.0	2.6	1.5	14194
Spain	50.4	29.6	10.9	3.9	2.9	2.3	20993
UK-BHPS	61.8	14.8	13.8	1.8	6.5	1.3	11406

Table 7: Causes of nonparticipation by type of participation pattern.

	Demogr. event	Out of scope	Contact failure	Lack of coop.	Absence	Total
Attrition	9.7	4.5	50.9	30.3	4.6	100.0
New entry	42.6	45.5	5.1	4.5	2.3	100.0
Occasional nonresponse	.0	7.7	41.5	32.6	18.1	100.0
Occasional response (dropout)	3.7	8.5	59.3	22.7	5.8	100.0
Occasional response (entry)	22.2	58.9	7.0	7.8	4.2	100.0
Very irregular response	.5	8.6	35.5	40.5	15.0	100.0

Table 8: Estimated grouped probit models for the probability of participation (** denotes an observed significance level below 1%, * denotes an observed significance level between 1 and 5%).

	Eligibility	Contact			Cooperation		
		1st excl.	All waves	All waves	1st excl.	All waves	All waves
constant	.933**	2.814**	2.995**	2.716**	2.986**	2.950**	3.177**
aged 0–15	-2.330**	-7.486	-10.901**	.055	5.923	1.726	.783
aged 16–25	-.203	-9.060*	-13.567**	-2.461*	-4.583	-7.066	-.655
aged 26–45	.069	-2.739	-4.989	-1.483	-2.930	-3.027	-1.579
aged 56–65	.013	-5.106	-7.076**	-1.685	-6.590	-3.837	-.859
aged 66–75	.031	-5.882	-8.046*	2.685*	1.208	-3.961	.202
aged 75+	.126	-2.816	-8.423**	-.400	-1.931	-4.832	1.254
female	.007	-.058	-1.080	-2.370**	-3.048	-3.246	-.508
college	-.010	-1.648	-1.173	.952**	-2.585	-.724	.376
secondary	-.027	-1.626*	-1.214	.182	-2.045*	-2.565**	-.082
unemployed	-.287**	-.032	-.681	.183	.628	.530	.937
inactive	-.027	1.124	.358	-.137	1.544	2.434*	-.582
nocohab	.075	1.737	4.206**	.400	5.259**	3.104*	.375
nosocial	.049*	-.713	-.887	-.668*	-.365	-.309	-.809*
adults	.075**	-.057	-.110	.029	-.184	-.298	-.148
children	-.233**	1.043	1.320**	.193	-1.188	-.630	-.217
tmove	-.008**	-.032	-.012	.026	.024	.038	-.021
nowner	.019	-.041	-.219	-.414*	.941	.176	-.241
itemnr	.045	-1.755	-2.192**	-1.084*	-3.182**	-2.364*	-3.517**
hincome	-.003**	.065	.008	-.003	.050	.031	-.019
CAPI	.027**	-.141	.404**	-.318*	-.306	-.174	-.655**
self	-.014	.525	1.359	-.321	-.313	.278	-.299
tel	.263**	.683	2.425	-.446	2.932	1.571	.082
proxy	-.041*	-.273	-.322	.029	.432	.521	.479
pintid	-.007*			.095*			-.045
pminint	.001	.005	.004	-.004	.019	.024*	-.003
nvisits	-.006	-.051	-.127	.014	.101	.099	.046
tfieldw	.005	.048	.042	.047	.044	.001	.083*
duration	.004	.007	-.006	-.040*	-.015	-.029	-.032
hminint	-.001	.004	.019**	.023**	.023	-.013	.014*
Austria	.084**	-.856	-.111	-.659**	-1.280*	-.692	-.453*
Belgium	.046*			-.386			-.577*
France	.058**	-.787	.019	-.299	-.904	-.537	-.291
GSOEP	.000			.460			-.346
Greece	-.018	-.777*	-.749**	-.258	.312	.114	.840**
Ireland	.099**	-1.707**	-1.135**	-1.100**	-2.001**	-1.704**	-1.133**
Portugal	.012	-.863*	-.574	-.154	-.875	-.716	-.054
Spain	.039**	-.985**	-.350	-.457**	-1.613**	-1.540**	-.824**
BHPS	.011			-.931**			-1.037**
year 1995	-.006			-.008			-.082
year 1996	-.010*			.081*			-.104**
year 1997	-.003			.082**			-.011
n	350	147	188	351	129	163	337
k	41	35	35	42	35	35	42
RSS	.060	3.351	4.421	8.324	2.530	4.171	6.907
R^2	.988	.522	.626	.834	.862	.809	.806

Table 9: Response rates in the next wave by country and year conditional on response in the current wave.

	Response				Contact				Cooperation given contact			
	1994	1995	1996	1997	1994	1995	1996	1997	1994	1995	1996	1997
Austria		.895	.899	.892		.900	.901	.896		.994	.998	.996
Belgium	.908	.905	.895	.888	.983	.975	.974	.974	.924	.928	.918	.912
France	.891	.924	.883	.894	.969	.925	.884	.899	.919	.999	.999	.995
Germany-GSOEP	.942	.940	.936	.922	.994	.991	.991	.986	.948	.949	.944	.935
Greece	.907	.918	.915	.873	.909	.918	.915	.874	.998	.999	.999	.999
Ireland	.812	.835	.873	.880	.827	.994	.991	.992	.982	.840	.881	.887
Italy	.945	.952	.901	.908	.969	.977	.968	.977	.975	.975	.931	.930
Portugal	.956	.936	.944	.927	.975	.941	.949	.936	.980	.995	.995	.990
Spain	.858	.895	.875	.880	.905	.928	.890	.911	.948	.964	.983	.967
UK-BHPS	.915	.942	.938	.937	.924	.945	.943	.953	.990	.996	.995	.983

Table 10: Parameter estimates of models for the probability of contact with and without the conditional independence assumption (** denotes an observed significance level below 1%, * denotes an observed significance level between 1 and 5%).

	With conditional independence			Without conditional independence		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
constant	2.000 **	1.894 **	1.995 **	2.001 **	1.894 **	1.998 **
adults	-.022 **		-.019 **	-.025 **		-.020 **
children	.076 **		.077 **	.079 **		.078 **
tmove	.018 **		.017 **	.018 **		.017 **
tmovemis	.175 **		.174 **	.172 **		.172 **
nowner	-.234 **		-.229 **	-.236 **		-.230 **
ownmis	-1.197 **		-1.172 **	-1.203 **		-1.164 **
hincome	.001 **		.002 **	.001 **		.002 **
itemnr	-.112 **		-.107 **	-.123 **		-.113 **
itemmis	-1.262 **		-1.221 **	-1.382 **		-1.284 **
nvisits		-.045 **	-.040 **		-.046 **	-.041 **
visitmis		-.566 **	-.510 **		-.584 **	-.527 **
tfieldw		.047 **	.041 **		.040 **	.033 **
duration		.046 **	.056 **		.046 **	.056 **
hminint		.002 **	.002 **		.002 **	.002 **
hminmis		-.045	-.024		-.051	-.031
Austria	-.683 **	-.572 **	-.597 **	-.681 **	-.569 **	-.595 **
Belgium	.070 **	-.023	-.041 *	.069 **	-.022	-.041 *
France	-.473 **	-.487 **	-.456 **	-.475 **	-.487 **	-.457 **
GSOEP	.532 **			.531 **		
Greece	-.636 **	-.662 **	-.683 **	-.631 **	-.660 **	-.680 **
Ireland	-.434 **	-.323 **	-.421 **	-.441 **	-.315 **	-.415 **
Portugal	-.313 **	-.270 **	-.309 **	-.311 **	-.271 **	-.310 **
Spain	-.640 **	-.570 **	-.625 **	-.642 **	-.570 **	-.626 **
BHPS	-.298 **	-.439 **	-.421 **	-.296 **	-.439 **	-.422 **
year94	.022 *	.140 **	.193 **	.017	.138 **	.191 **
year95	.127 **	.198 **	.240 **	.124 **	.197 **	.238 **
year96	.003	.043 **	.059 **	.004	.045 **	.060 **
k_1	22	18	27	22	18	27
$-\hat{L}$	92328.4	93678.5	92125.4			
R^2	6.1	4.7	6.3			

Table 11: Parameter estimates of models for the conditional probability of cooperation given contact with and without the conditional independence assumption (** denotes an observed significance level below 1%, * denotes an observed significance level between 1 and 5%).

	With conditional independence			Without conditional independence		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
constant	1.666**	1.574**	1.607**	1.684**	1.589**	1.621**
age	.001*		.000	.001*		.000
age^2	-.000**		-.000**	-.000**		-.000**
female	.012		.009	.013		.010
college	.076**		.086**	.075**		.086**
secondary	.006		.011	.005		.011
educmis	-.035*		-.034*	-.034*		-.032*
unemployed	.022		.018	.023		.018
inactive	.079**		.077**	.078**		.075**
lfstmis	-.337**		-.302**	-.332**		-.299**
nocohab	-.080**		-.081**	-.079**		-.080**
cohmis	-.222		-.179	-.219		-.177
nosocial	-.147**		-.141**	-.143**		-.137**
socmis	-.272**		-.240**	-.264**		-.233**
CAPI		-.118**	-.126**		-.118**	-.124**
self		-.059**	-.057**		-.059**	-.058**
tel		-.370**	-.355**		-.366**	-.340**
proxy		-.142**	-.096**		-.141**	-.096**
modemis		.316**	.318**		.316**	.318**
pintid		.117**	.121**		.116**	.119**
pintmis		1.078**	1.118**		1.067**	1.100**
pminint		-.000	-.000		-.000	-.000
pminmis		-.327**	-.331**		-.325**	-.328**
Austria	.954**	.948**	.952**	.951**	.972**	.974**
Belgium	-.261**	-.263**	-.257**	-.262**	-.262**	-.257**
France	.350**	.222**	.274**	.366**	.243**	.296**
GSOEP	.152**	-.054**	.154**	.133**	-.063**	.139**
Greece	1.337**	1.429**	1.421**	1.321**	1.446**	1.435**
Ireland	-.418**	-.438**	-.447**	-.395**	-.419**	-.425**
Portugal	.644**	.725**	.730**	.649**	.731**	.736**
Spain	.118**	.081**	.055**	.143**	.110**	.086**
BHPS	.936**	.661**	.870**	.933**	.671**	.873**
year94	.022*	-.016	-.025	.024*	-.015	-.026
year95	.115**	.123**	.122**	.108**	.118**	.115**
year96	.076**	.094**	.091**	.074**	.094**	.091**
k_2	26	22	35	26	22	35
$-\hat{L}$	59206.9	59103.7	58805.2			
R^2	8.5	8.7	9.1			
k	48	40	62	48	40	62
$\hat{\rho}$	0	0	0	-.951**	-.477**	-.905**
$-\hat{L}$	151535.3	152782.2	150930.6	151520.7	152775.7	150908.4
LR stat.				29.3**	13.0**	44.5**

Appendix

Let $\pi_{ij} = \Pr\{Y_1 = i, Y_2 = j\}$, $i, j = 0, 1$. Under the bivariate probit model (1)

$$\pi_{11} = \int_{-\mu_1}^{\infty} \Phi\left(\frac{\mu_2 + \rho u_1}{\sigma}\right) \phi(u_1)\, du_1$$

and

$$\pi_{10} = \int_{-\mu_1}^{\infty} \left[1 - \Phi\left(\frac{\mu_2 + \rho u_1}{\sigma}\right)\right] \phi(u_1)\, du_1,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote respectively denote the density and distribution function of the standardized Gaussian distribution. Clearly

$$\pi_1 = \Pr\{Y_1 = 1\} = \pi_{10} + \pi_{11} = \int_{-\mu_1}^{\infty} \phi(u_1)\, du_1 = \Phi(\mu_1).$$

Further

$$\pi_{1|1} = \Pr\{Y_2 = 1 \,|\, Y_1 = 1\} = \frac{\pi_{11}}{\pi_1} = \frac{1}{\Phi(\mu_1)} \int_{-\mu_1}^{\infty} \Phi\left(\frac{\mu_2 + \rho u_1}{\sigma}\right) \phi(u_1)\, du_1.$$

If $\rho = 0$, then

$$\pi_{11} = \Pr\{Y_1 = 1\} \Pr\{Y_2 = 1\} = \Phi(\mu_1)\, \Phi(\mu_2),$$

$$\pi_{10} = \Pr\{Y_1 = 1\} \Pr\{Y_2 = 0\} = \Phi(\mu_1)\, [1 - \Phi(\mu_2)],$$

and

$$\pi_{1|1} = \Pr\{Y_2 = 1\} = \Phi(\mu_2).$$

Differentiating with respect to μ_1 gives

$$\frac{\partial \pi_1}{\partial \mu_1} = \phi(\mu_1) > 0,$$

$$\frac{\partial \pi_{11}}{\partial \mu_1} = \Phi\left(\frac{\mu_2 - \rho \mu_1}{\sigma}\right) \phi(\mu_1) > 0, \tag{3}$$

and

$$\frac{\partial \pi_{10}}{\partial \mu_1} = \left[1 - \Phi\left(\frac{\mu_2 - \rho \mu_1}{\sigma}\right)\right] \phi(\mu_1) > 0.$$

Further

$$\frac{\partial \pi_{1|1}}{\partial \mu_1} = \frac{\phi(\mu_1)}{\Phi(\mu_1)} \left[\Phi\left(\frac{\mu_2 - \rho \mu_1}{\sigma}\right) - \pi_{1|1}\right],$$

whose sign and magnitude depends on the difference $\Phi((\mu_2 - \rho\mu_1)/\sigma) - \pi_{1|1}$. If $\rho = 0$, then

$$\frac{\partial \pi_1}{\partial \mu_1} = \phi(\mu_1),$$

$$\frac{\partial \pi_{11}}{\partial \mu_1} = \Phi(\mu_2)\,\phi(\mu_1) > 0, \tag{4}$$

$$\frac{\partial \pi_{10}}{\partial \mu_1} = [1 - \Phi(\mu_2)]\,\phi(\mu_1) > 0,$$

and

$$\frac{\partial \pi_{1|1}}{\partial \mu_1} = \frac{\phi(\mu_1)}{\Phi(\mu_1)}\,[\Phi(\mu_2) - \Phi(\mu_2)] = 0.$$

A comparison of (3) and (4) shows that the difference in $\partial \pi_{11}/\partial \mu_1$ between the two cases of dependence ($\rho < 0$) and independence ($\rho = 0$) is equal to

$$\left[\Phi\left(\frac{\mu_2 - \rho\mu_1}{\sigma}\right) - \Phi(\mu_2)\right]\,\phi(\mu_1),$$

and can be positive or negative depending on the values of μ_1, μ_2 and ρ. It is positive whenever

$$\frac{\mu_2 - \rho\mu_1}{\sigma} > \mu_2,$$

that is, whenever

$$\frac{\mu_2}{\mu_1} > \frac{\rho}{1-\sigma} = \frac{\rho}{1 - \sqrt{1-\rho^2}}$$

When $-1 < \rho < 0$, this inequality is satisfied if μ_1 and μ_2 are both positive.