Survey response and survey characteristics: Micro-level evidence from the European Community Household Panel*

Cheti Nicoletti
ISER, University of Essex

Franco Peracchi
University of Rome "Tor Vergata"

January 2005

Abstract

This paper presents micro-level evidence on the role of the socio-demographic characteristics of the population and the characteristics of the data collection process as predictors of survey response. Our evidence is based on the public use files of the European Community Household Panel (ECHP), a longitudinal household survey covering the countries of the European Union, whose attractive feature is the high level of comparability across countries and over time.

We model the response process as the outcome of two sequential events: (i) contact between the interviewer and an eligible interviewee, and (ii) cooperation of the interviewee. Our model 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 observe whether a person is a respondent only if she has been contacted.

Keywords: Panel data, survey response, bivariate probit model.

JEL classification: C33, C35, C81

* 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 the editor, two anonymous referees, Marco Bonetti, Alberto Holly, Peter Lynn, Arthur van Soest, Ernesto Villanueva and seminar participants at CORE, DIW, Ente Einaudi and Universitat Pompeu Fabra for useful comments.

1 Introduction

Identifying the determinants of survey response is of considerable practical importance. First, estimates of the probability of survey response 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 (2001) for the estimation of conditional means in the presence of missing data. Second, modelling survey response is crucial in the construction of two-step estimators of regression models with sample selection and, more generally, in the joint estimation of a regression model and 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 when the available data are subject to nonresponse. A third reason for studying the determinants of survey response 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.

In this paper we focus on response to longitudinal household surveys, and present an exploratory analysis of the determinants of the probability of survey response from one wave to another using comparable micro-level data for several European countries. Our evidence is based on the public use files of the European Community Household Panel (ECHP), a longitudinal survey of households and individuals, centrally designed and co-ordinated by the Statistical Office of the European Communities (Eurostat). Comparable international data on contact and cooperation have also been used by de Heer (1999) and de Leeuw and de Heer (2002) to relate international trends in household survey nonresponse to sampling and survey design, fieldwork, and survey organization. Their analysis, however, is based on aggregate response rates and does not take into account changes in the composition of the national populations along dimensions that are correlated with survey response.

Our paper combines two strands of the literature. The first has to do with the relative importance of different types of predictors of survey response, namely the socio-demographic characteristics of the population on the one hand, and the characteristics of the data collection process on the other hand. The role of the former has been emphasized, among others, by Hausman and Wise (1979), Ridder (1992), and Fitzgerald *et al.* (1996), whereas the role of the latter has been emphasized by Campanelli *et al.* (1997) and Lepkowski and Couper (2002).

The second strand of the literature (Campanelli *et al.* 1997, Groves & Couper 1998, Lepkowski & Couper 2002, Lynn *et al.* 2002) looks in more detail into nonresponse by separately considering location, contact and cooperation of the sample units. For example, Lynn *et al.* (2002) argue that a "major weakness of much previous research is that it either confounds ease of contact with reluctance or isolates one without considering simultaneously the effect of the other". To our knowledge, they are the first to investigate whether there is a relationship between propensity to cooperate and ease of contact. Their aggregate analysis across surveys and over time finds no evidence of correlation between contact and cooperation propensity. In this paper, we look again at this issue at the micro level, controlling for a broad set of variables characterizing the individuals, their households, and the fieldwork.

The 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. We also analyze the patterns of survey participation, distinguishing between ineligibility and nonresponse. Because of data availability, we can only study participation after the first wave, that is, our analysis of survey participation is conditional on survey participation in the first wave.

In the second part (Sections 3 and 4), we focus attention on survey response, i.e. on participation given eligibility. We use micro-level data to predict response in the next wave given response in the current wave, focusing on how the probabilities of contact failure and refusal to cooperate vary with the characteristics of the data collection process and the socio-demographic composition of the national populations. 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 paper estimates 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 can only observe cooperation for those who have been succesfully 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 2.3 analyzes the main differences in the survey design and the data collection process across countries. We then focus attention on survey nonparticipation, especially on unit nonresponse caused by contact failure or lack of cooperation. Section 2.4 analyzes participation in a single wave of the survey, while Section 2.5 analyzes the patterns of participation across waves.

2.1 Target population

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

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 persons are eligible for interview 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 the EU) is "traced" and interviewed again if she returns to the target population. Ineligible nonsample persons are not traced.

2.2 Country coverage

In Belgium and the Netherlands, the ECHP was linked from the beginning to existing national panels. 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, the Luxembourg's Social Economic Panel, and the British Household Panel Survey. Starting from the fourth (1997) wave, the ECHP data for Germany, Luxembourg and the UK have also been derived from the existing national panels. For Sweden, a comparable ECHP data set, derived from the Swedish Living Conditions Survey is available from the fourth (1997) wave.

To reduce the impact of unmeasurable differences in survey design and organization, we focus on the seven countries (Denmark, France, Greece, Ireland, Italy, Portugal, Spain) for which the ECHP ran independently of existing national surveys and data are available for all five waves included in the 2002 User Data Base (UDB). The UDB is an anonymized and user-friendly version of the ECHP data. The first release of the UDB, covering waves 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.

2.3 Survey differences across countries

This section summarizes the main differences in the design and organization of the survey across countries and waves, focusing on observable survey characteristics which may be relevant for understanding cross-country differences in survey participation. We refer to Peracchi (2002) for a more complete review of differences across countries.

Table 1 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).

The average number of visits ranges from a minimum of 1.1 in Greece to a maximum of 3 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 is lowest in Greece and Portugal (about 30 percent) and highest in Ireland (about 80 percent).

In general, personal interviews tend to last longer than household interviews. In France, however, household interviews tend to last longer than personal interviews because the household questionnaire is quite time-demanding, whereas in Greece, Italy and Portugal, both interviews tend to be quite short. The average duration of the household interview ranges between a minimum of 17 minutes in Greece and Italy and a maximum of 28 minutes in France. The average duration of the personal interview ranges instead between a minimum of 18 minutes in France and a maximum of

32 minutes in Ireland. Finally, fieldwork lasts on average more than one year in Ireland, but only three months in France and Spain.

Table 2 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 the respondent, telephone interview, and proxy interview. In general, the most common interview mode is the traditional PAPI. The main exceptions are Greece and Portugal, where the most common interview mode is instead CAPI. The percentage of self-administered and telephone interviews is low and only exceeds 1 percent in Italy and Spain. Finally, proxy interviews are rare in Denmark and Greece (2.3 and 1.6 percent respectively) but are non negligible (10 percent or more) in all other countries, with a maximum in Italy and Spain (15.8 and 16 percent respectively).

2.4 Survey participation in a single wave

We say that a person does not participate in a given wave of the panel if she is ineligible in that wave or if she is a unit nonrespondent. Unit nonresponse occurs when an eligible person (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, namely the personal residential status, which gives details on whether or not a person is within the scope (except for the first wave), and 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.

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 vice versa: it includes institutionalization, migration to a foreign country, movement of a nonsample person to a household without sample individuals, etc.

3. Absence of the person at the address.

4. 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.

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

Notice that if the interviewer neither succeeds in contacting a person nor obtains information from relatives, neighbours 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 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 provides no information on ineligibility and unit nonresponse in the first wave. An indication of the magnitude of the problem is given by Table 3, which reports household response rates in the first three waves of the ECHP as computed by Eurostat (1997).

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.

Averaging over the seven countries considered, we obtain an overall household response rate in the first wave of 76 percent, which 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 rate in Ireland (55.8 percent) mainly reflects outright refusal to respond, whereas the high response rates in Greece and Italy (90.1 and 90.7 respectively) may reflect the fact that cooperation is compulsory in these two countries. 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 Ireland and Spain.

Table 4 reports the fraction of nonparticipants by country and wave (except the first wave), that is, the ratio between the number of nonparticipants in a given wave and the number of people who participated in at least one wave of the ECHP. The fraction of nonparticipants is below 30 percent in all countries except Ireland. It shows a clear upward trend in Denmark, a clear downward trend in Ireland and Spain, but no clear trend in the other countries considered.

2.5 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 ECHP.

Let D_j be a 0–1 indicator of survey participation in wave j ($D_j = 1$ for survey participants). Since the data used in this paper contain the first five waves of the survey, 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. continued participation: $D = (1, 1, 1, 1, 1)$;
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 at most once. Monotone participation patterns are the first three cases. We say that a participation pattern is nonmonotone if D_j changes value more than once, as in the last three categories of our classification.

Table 5 compares participation patterns across countries. Continued participation is always the most frequent pattern, followed by monotone attrition and new entry. It represents 50 percent or more of the cases in all countries except Denmark and Ireland, where the percentage is somewhat lower because of the high frequency of monotone attrition. Nonmonotone participation patterns

are infrequent and, in all countries except Denmark, never represent more than 10 percent of the cases.

Table 6 shows, for the various types of pattern, the relative importance of the different causes of participation and nonparticipation. For people 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 response patterns, we report the causes of both nonparticipation before entry and of dropout after participation. For the occasional nonresponse patterns, we only report the causes of nonparticipation before re-entry, whereas for the very irregular response patterns, we only report the causes of nonparticipation before the last entry.

Quantitatively, monotone attrition is much more important than new entry. Moreover, new entry is mainly linked to eligibility (demographic events or out of scope), while monotone attrition is mainly due to contact failure or lack of cooperation. The latter 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 within the scope of the survey (58.9 percent of the cases).

3 Modelling survey response

In this section we focus on eligible people who have been successfully interviewed in a given wave of the ECHP and investigate their patterns of survey response in the next wave. We cannot study noncontact or refusal to cooperate in the first wave because the public-use files of the ECHP do not provide such information. On he other hand, as convincingly argued by Lepkowski and Couper (2002), the response process in later waves of a panel differs in important ways from the initial wave, as a result of both the self-selection of the sample units and the increasing information and organizational experience of survey agencies at each successive wave.

3.1 Cross-country differences in survey response

For eligible people, participation and response are equivalent events, and the response process may be described as the outcome of two sequential events: (i) contact between the interviewer and the interviewee, and (ii) cooperation of the interviewee. We do not distinguish between location of a sample unit and contact given location because the information contained in the ECHP does

not allow a clear distinction between these two events. Lepkowski and Couper (2002) find that "noncontact once a subject has been located is typically a relatively rare status in longitudinal surveys".

Table 7 presents the overall response rate (the fraction of respondents in the current wave who also respond in the next wave) by country and wave, and its two components: the contact rate (the fraction of eligible people who have successfully been contacted in the next wave) and the cooperation rate (the fraction of contacted people who completed the personal interview in the next wave). The table shows that response rates after the first wave are high. With the exception of Ireland, however, they tend to decline over time. In some countries (France, Greece and Portugal) this appears to reflect an increasing difficulty in contacting people, whereas in other countries (Denmark, Ireland and Italy) it appears to reflect an increasing difficulty in obtaining cooperation from contacted people.

The observed cross-country differences in response rates may reflect differences in the composition of the national populations along dimensions that are correlated with the survey response decision. For example, the propensity to cooperate may be linked to personal characteristics such as gender, age and schooling attainments. Because contact may be harder for people who move frequently or live alone, labour force status and living in a couple may be important predictors of contact failure. However, because the ECHP is not completely harmonized across countries, differences in survey response may also reflect differences in the data collection process.

To investigate the role played by the characteristics of the data collection process and the socio-demographic composition of the population, we carry out a micro-level analysis that uses the information on the respondents in the current wave to predict survey response in the next wave.

3.2 The statistical model

Let Y_1 be the indicator of the event that a survey respondent is contacted in the next wave, and let Y_2 be the indicator of the event that the person cooperates in the next wave. Conditional on response in the current wave, the response process in the next wave is completely described by two elements: the probability $\pi_1 = \Pr\{Y_1 = 1\}$ of future contact, and the probability $\pi_{1|1} = \Pr\{Y_2 = 1 \,|\, Y_1 = 1\}$ of future cooperation given contact. The probability of response in the next wave is then simply the product $\pi_{11} = \pi_{1|1}\,\pi_1$.

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

one can find a vector X of covariates such that $\Pr\{Y_2 = 1 \,|\, Y_1 = 1, X\} = \Pr\{Y_2 = 1 \,|\, X\}$. In practice, the conditional independence assumption may be restrictive because it ignores correlation arising from omitted individual or survey characteristics affecting 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 nonresponse. For example, a possible source of correlation is the fact that people who do not want to be found, and pretend to be absent when an interviewer knocks at their door, are likely to be less cooperative once contacted. In this case, one may well say that unwillingness to cooperate is misclassified as noncontact.

A simple parametric model that allows for conditional correlation is the bivariate probit model

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

where Y_1^* and Y_2^* are 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_1 and X_2 are predictors of contact and cooperation respectively, and U_1 and U_2 are regression errors distributed independently of the predictors according to a bivariate Gaussian distribution with zero means, unit variances, and correlation coefficient ρ. The normalization of the variances is necessary because the model parameters are only identifiable up to scale.

The vector of model parameters is $\theta = (\theta_1, \theta_2, \rho)$, where $\theta_j = (\alpha_j, \beta_j)$ is a k_j-vector, and the parameter space is $\Theta = \Re^k \times (-1, 1)$, with $k = k_1 + k_2$. When $\rho = 0$, the model implies conditional independence between the ease of contact and the propensity to cooperate, as in the Lepkowski and Couper (2002) model.

Model (1) says that contact occurs ($Y_1 = 1$) if the ease of contact is high enough ($Y_1^* > 0$ or, equivalently, $U_1 > -\alpha_1 - \beta_1^\top X_1$). Similarly, cooperation occurs ($Y_2 = 1$) if the propensity to cooperate is high enough ($Y_2^* > 0$ or, equivalently, $U_2 > -\alpha_2 - \beta_2^\top X_2$). Both the ease of contact and the propensity to cooperate depend on observable and unobservable personal and survey characteristics. Because unobservables matter, the ease of contact and the propensity to cooperate may be correlated even if we control for the observables.

Construction of the log-likelihood is straightforward after noticing that cooperation may only be observed for those who are contacted, that those who are contacted and cooperate contribute to the likelihood the probability π_{11} of contact and cooperation, those who are contacted and do not cooperate contribute the probability π_{10} of contact and noncooperation, whereas those who are not contacted only contribute the probability $1 - \pi_1$ of unsuccessful contact. Thus, the log-likelihood

for a sample of n independent observations is

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

where the subscript i indexes the individuals in the sample. Under model (1), we have

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

where $\mu_{ij} = \alpha_j + \beta_j^\top X_{ij}$ $(j = 1, 2)$, $\sigma = \sqrt{1 - \rho^2}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the density and the distribution function of the standardized Gaussian distribution. A maximum likelihood estimate of θ maximizes (2) over the parameter space Θ.

Some conditions are necessary for the above censored bivariate probit model to be identifiable. If the covariates in the contact and the cooperation equations consist of the same set of dummy variables, then the model is not identifiable. Identifiability becomes possible if there is at least one covariate (either continuous or categorical with more than two values) that enters linearly into the model, or there are exclusion restrictions, namely the covariates in X_1 and X_2 are different.

In our empirical application we impose both types of assumption. In particular, we impose that 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 justified by the argument in Section 3.3. Moreover, in both contact and cooperation equations we impose that some continuous and ordered categorical variables enter linearly. In the contact equation we will consider the following variables entering linearly: number of adults, number of children, number of years since last change of address, household income, number of interviewer visits to the household, length of the fieldwork, and duration in minutes of the household interview. In the cooperation equation we consider instead age, age square, and duration in minutes of the personal interview.

Within this censored bivariate model, testing the hypothesis of conditional independence between ease of contact and propensity to cooperate is equivalent to testing if $\rho = 0$. An alternative test based on the likelihood 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 simple probit models, one for Y_{i1} and one for Y_{i2} conditional on $Y_{i1} = 1$. Instead of probit, Lepkowski and Couper (2002) use logit models for the case of conditional independence. Nevertheless, the Monte Carlo evidence in Morimune (1979) and the theoretical results in Horowitz (1993) show that logit and probit lead to very similar results.

3.3 Choice of predictors

Recent work by Fitzgerald *et al.* (1996), Campanelli *et al.* (1997), and Lepkowski and Couper (2002), among others, offers suggestions about which variables are likely to help predict contact and cooperation. These variables include both survey features and household and personal characteristics.

In principle, one may distinguish between contact at the household and the personal level. One may therefore distinguish between two sets of variables that explain the probability of contact: household-specific variables linked to the probability of contacting a household, and person-specific variables linked to the probability of contacting a person. In practice, however, contact rates at the individual and the household level tend to coincide because either all or none of the members of a household are contacted. Because, in all countries considered, contact at the household and the personal level coincide except for a very small number of cases; we think it is not too restrictive to assume that only household-specific variables affect the probability of contact.

The probability of contacting a household is inversely related to its degree of geographical mobility and to the probability of finding someone at home. Moreover, because people may pretend to be absent when an interviewer knocks at the door, the contact probability may also be related to a household's willingness to cooperate. Since households who move around may be more difficult to locate, variables that help explain household mobility are also likely to help explain contact failure. A larger household size, the presence of children, and home ownership are all associated with lower mobility. Household size and the presence of children may also be related to the probability of finding someone at home. On the other hand, households that changed address recently may be more likely to move again. Thus, our predictors of contact include the number of adults (`adults`) and the number of children (`children`) in a household, homeownership (the dummy `nowner` for not owning home), and the number of years since the last change of address (`tmove`). The probability to find someone at home is likely to be lower for a household whose contact required a higher number of visits last year (`nvisits`). A longer fieldwork (`tfieldw`) in the current wave may be related to

a higher number of contact attempts and, possibly, a higher probability to find someone at home. Finally, failed contact for people pretending to be absent may be linked to the perceived cost of completing the interview and the household's past experience with the interview. We therefore add to the predictors the equivalized household income (`hincome`), that is, household income divided by a measure of household size, an index of item nonresponse to household income (`itemnr`), and the duration of the household interview (`hminint`) in last wave.

Once a household is successfully contacted, lack of cooperation is mainly the result of a personal decision that reflects personal characteristics, related to the perceived cost of completing the personal interview, and a person's past experience with the survey. The personal characteristics that we consider include age (which enters as a quadratic term), gender (the dummy `female`), schooling attainments represented by two dummies, one for completed college education (`college`) and one for completed secondary education (`secondary`), labor force status represented by two dummies, one for being unemployed (`unemployed`) and one for being out of the labor force (`inactive`), and indicators for not living in a couple not living with a spouse (`nocohab`) and infrequently talking to neighbors (`nosocial`).

To capture a person's past experience with the survey, we include features of the personal interview process in the current wave, namely the duration of the personal interview (`pminint`), a dummy for the presence of the same interviewer as in the last wave (`pintid`), and dummies for the interview mode. Because of the importance of the interviewer-respondent interaction (Groves & Couper 1998, Laurie *et al.* 1999, Hox & de Leeuw 2002), a person contacted by the same interviewer as in previous waves is likely to be more willing to cooperate again. Having had a face-to-face pen-and-pencil personal interview (PAPI) in the current wave may increase the psychological cost of refusal relative to less personal and direct interview modes, such as interview by telephone, self-administered, or computer assisted (CAPI). The duration of the personal interview may reflect the interest of the interviewee and her propensity to cooperate but, on the other hand, too long an interview can lead to a refusal in the following wave.

Both the model for contact and the model for cooperation also include country dummies to capture time-invariant unobserved heterogeneity across countries, year dummies to capture country-invariant time effects, and indicators for the number of times a person was previously interviewed (`int1`, `int2` and `int3` for one, two and three times). They also include a set of indicators (variables ending with `mis`) for missing predictors (years of residence at the current address, talking to neighbours, mode of interview, number of visits to the household, and the dummy for the presence

of the same interviewer as last year).

Table 8 presents, for all predictors considered, the number of nonmissing observations, the mean and the standard deviation.

Before discussing our empirical results, we would like to point out the relationships between our models and those estimated by Fitzgerald *et al.* (1996), Campanelli *et al.* (1997) and Lepkowski and Couper (2002). We use a probit specification for the response model as in Fitzgerald *et al.* (1996) but, like Campanelli *et al.* (1997) and Lepkowski and Couper (2002), we distinguish between two processes, one for contact and one for cooperation, and use data collection characteristics as well as personal and household variables as predictors.

As in Campanelli *et al.* (1997), but unlike Lepkowski and Couper (2002), we take into account the role of the interviewer. In particular, we consider the effect of different interview modes and the use of the same interviewer across waves. We would like to emphasize that the interviewer continuity dummy may be not be completely exogenous if areas where people are less cooperative (for example big cities) are also areas in which the turnover of the interviewers is higher (see Campanelli and O'Muircheartaigh 2002). Unlike Campanelli *et al.* (1997), we do not consider interviewer and/or area effects because no such information is available in the ECHP. We instead introduce time and country dummies to control for residual unobserved heterogeneity.

We do not consider variables related to the survey experience, such as cooperation with the interviewer and level of understanding of the questions, because they are not available in the ECHP. We instead consider other variables that are likely to help predict future contact and cooperation, such as the income nonresponse index, the number of visits, and the durations of the household and personal interviews in the previous wave.

Finally, unlike all other papers, we allow for correlation between the errors in the contact and the cooperation models. Further, by exploiting the high degree of comparability of the ECHP to pool the available ECHP data for our set of countries, we are able to work with a very large sample size. This allows us to identify more easily the effects of the variables considered.

4 Empirical results

Tables 9 and 10 compare the estimates obtained for three alternative specifications of the models with and without the conditional independence assumption. Table 9 presents the results for the probability of contact, Table 10 for the conditional probability of cooperation given contact. The estimates are based on the pooled data from the various countries and waves. Pooling the data

helps reducing collinearity problems due to the limited within-country variability of some of the variables, especially those related to the data collection process.

After dropping cases with equivalized annual household income below Euro 100 or with missing information on educational attainments, labor force status, cohabitation status or home ownership, the sample size consists of 323,694 observations on 100,874 individuals (13,304 observed only once, 11,103 observed two times, 17,249 observed three times, and 59,218 observed four times). At the bottom of each table we report the number of estimated regression parameters (k_1, k_2 and k), minus the maximized log likelihood ($-\hat{L}$), the Akaike information criterion (AIC), and a standard measure of goodness-of-fit (pseudo R^2). At the bottom of Table 10, we also report the likelihood ratio statistics (LR stat.) for testing the conditional independence assumption.

The first specification (Model 1) excludes current survey features from the models for the probability of future contact and cooperation. The second (Model 2) ignores instead the role of household and personal characteristics. The third specification (Model 3) is the most general and includes as predictors both survey features and household and personal characteristics. For simplicity, all specifications only include the main effects and ignore interactions between variables. We always take Italy and the first (1994) wave as the reference.

The intercepts α_1 and α_2 are directly interpretable as the inverse transforms of the probabilities of contact and cooperation for the reference case. Thus, for Models 1 and 3, $\Phi(\alpha_1)$ corresponds to the probability of contact in 1995 of an Italian interviewed in 1994 for the first time, living in a 2-adult household with one child, residing at the same address for 10 years, homeowner, with a fully reported equivalized household income of Euro 10,000, whereas $\Phi(\alpha_2)$ corresponds to the probability of cooperation in 1995 of an Italian male employee aged 50, interviewed in 1994 for the first time, with at most lower secondary education completed, living in a couple, and frequently talking to neighbours.

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. Estimated standard errors are derived from the robust or "sandwich" estimate of the asymptotic variance matrix, under the assumption that observations on the outcome variable are independent across persons but not necessarily for the same person. In particular, other things being equal:

- The number of children in the household, home ownership and the length of residence at the current address are positively related to the probability of future contact. The number of adults is also positively related, but its effect is small and not statistically significant.

- The index of item nonresponse to household income is negatively related to the probability of future contact. Equivalized household income is also negatively related, but its effect is small and not statistically significant.

- The length of the fieldwork and the duration of the household interview are positively related to the probability of future contact, whereas the number of visits to a household is negatively related. The latter result is likely to reflect the practice of reaching contact in difficult cases by increasing the number of callbacks.

For cooperation:

- Age does not appear to help predict cooperation after controlling for all the other variables.

- Women are more likely to cooperate than men, but the difference is not statistically significant.

- People with college (tertiary) education are more likely to cooperate than people with lower education, but the difference is not statistically significant.

- Being out of the labour force is positively related to the probability of future cooperation, whereas not living in a couple and infrequent interactions with the neighbours are negatively related.

- The use of interview modes different from face-to-face PAPI (self-administered, telephone and proxy) is negatively related to the probability of future cooperation. But the impact of face-to-face CAPI does not differ significantly from face-to-face PAPI. The presence of the same interviewer as last year is positively related whereas the duration of the personal interview is negatively related, but these effects are small and not statistically significant.

Other things being equal, the probability of contact and the probability of cooperation increase with successive interviews. In both cases, the size and statistical significance of the country dummies indicate that time-invariant heterogeneity across countries is very important, whereas the profile of the year dummies suggests an initial increase of the probabilities of contact and cooperation, followed by a negative time trend.

Even after controlling for personal/household and data collection characteristics, the probability of contact (cooperation) tends to be higher for those countries where the unconditional contact (cooperation) rates are higher. Moreover, the conditional probability of contact tends to be higher in

countries where the conditional probability of cooperation tends to be lower. A negative correlation of about -.66 was also observed in Table 7, where unconditional rates by countries are reported. According to our estimates, this negative correlation is mainly due to the negative correlation between the time invariant country effects in the contact and cooperation equations. Only a minor part is instead due to the negative correlation between the errors in the contact and cooperation equations. A possible explanation for this phenomenon is that countries where contact was more difficult decided to invest more energy to increase cooperation rates.

We find that the likelihood ratio test strongly rejects the conditional independence assumption only for Model 3 but, even in this case, the estimated coefficients hardly change when conditional independence is relaxed. The estimated correlation coefficient between the two latent variables ($\hat{\rho}$) is negative but relatively small, and is statistically significant at the 1 percent level only for Model 3. The magnitude and statistical significance of the estimated correlation coefficient increases substantially when we omit from the model the indicators for the number of times a person was previously interviewed (`int1`, `int2` and `int3`).

Finally, a comparison of the maximized log-likelihoods for the three specifications shows that omitting household or personal characteristics always has a more severe impact than omitting survey features.

5 Conclusions

This paper analyzes a number of issues surrounding survey participation in household panels, using the ECHP as an illustration. This final section summarizes our main findings.

Looking at the patterns of survey participation in the ECHP, we find that monotone participation patterns (mainly attrition) are much more frequent than nonmonotone patterns. Furthermore, entry into the panel is mainly due to the eligibility condition, while exit is mainly due to contact failure and refusal to cooperate.

When we try to predict future survey response 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 and the index of nonresponse to household income are good predictors of future contact, whereas age, labour force status, living in a couple, and frequent contacts with the neighbours 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 response in the next wave. In particular, the number of visits, the length of the fieldwork, and the length of the household interview significantly affect the probability of future contact, whereas the interview mode significantly affects the probability of future cooperation given contact.

We think that these findings have important consequences for the specification and estimation of regression models with sample selection, where identification is typically achieved through exclusion restrictions. Indeed, our findings provide a justification for including variables characterizing the data collection process in the model for sample response, while excluding them from the model for the outcome variable of interest.

Finally, in line with Lynn *et al.* (2002), we find little evidence of correlation between the ease of contact and the willingness to cooperate. We conclude that correlation is not a problem after controlling for a proper set of personal, household and data collection characteristics, and for unobserved factors that are constant either within countries or within waves.

References

Abowd J., B. Crépon B., and F. Kramarz (2001), "Moment estimation with attrition", *Journal of the American Statistical Association*, 96, 1223–1231.

Campanelli P., and C. O'Muircheartaigh (2002), "The importance of experimental control testing the impact of interviewer continuity on panel survey nonresponse, *Quantity and Quantity*, 36, 129-Ů144.

Campanelli P., P. Sturgis, and S. Purdon (1997), *Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey response Rates*, SCPR, London.

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

de Heer W. (1999), "International response trends: Results of an international survey", *Journal of Official Statistics*, 15, 129–142.

de Leeuw E., and W. de Heer (2002), "Trends in household survey nonresponse: A longitudinal and international comparison", in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), *Survey Nonresponse*, Wiley, New York.

Fitzgerald J., P. Gottschalk, and R. Moffitt (1998), "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics", *Journal of Human Resources*, 33, 251–299.

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.

Horowitz J.L. (1993), " Semiparametric and nonparametric estimation of quantal response models", in C.R. Maddala and H.D. Vinod (eds.), *Handbook of Statistics*, 11, 45–72.

Hox J., and E. de Leeuw (2002), "The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison", in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), *Survey Nonresponse*, Wiley, New York.

Laurie H., R. Smith, and L. Scott (1999), "Strategies for reducing nonresponse in a longitudinal panel survey", *Journal of Official Statistics*, 15, 269–182.

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.

Lillard L.A., and C.W.A. Panis (1998), "Panel attrition from the Panel Study of Income Dynamics", *Journal of Human Resources*, 33, 439–547.

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.

Morimune K. (1979), "Comparison of normal and logistic models in the bivariate dichotomous analysisŠŠ, *Econometrica*, 47, 957Ů-976.

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

Ridder G. (1990), "Attrition in multi-wave panel data", in J. Hartog, G. Ridder, and J. Theeuwes (eds.), *Panel Data and Labor Market Studies*, North-Holland, Amsterdam.

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

Robins J., A. Rotnitzky, and L. Zhao (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: Mean of selected survey features by country.

	Number of visits	% same interviewer	Length hh interview	Length pers. interview	Length of fieldwork
Denmark	3.0	52.1	20	26	6
France	2.4	62.2	28	18	3
Greece	1.1	32.5	17	22	9
Ireland	2.8	80.4	20	32	13
Italy	1.7	54.9	17	19	6
Portugal	1.7	29.7	18	20	4
Spain	2.0		18	23	3

Table 2: Interview modes by country.

	PAPI	CAPI	Self-adm.	Phone	Proxy	Missing	
Denmark	95.8	.0	.9	.1	2.3	1.0	100.0
France	53.8	.0	.0	.0	9.6	40.2	100.0
Greece	20.9	76.7	.4	.4	1.6	.0	100.0
Ireland	88.2	.0	.0	.0	11.8	.0	100.0
Italy	82.1	.2	.0	1.7	15.8	.2	100.0
Portugal	30.4	58.4	.5	.1	10.6	.0	100.0
Spain	80.0	.0	2.7	1.3	16.0	.0	100.0

Table 3: Household response rates (percent) in the first three waves of the ECHP. Source: Eurostat (1997).

	Wave 1	Wave 2	Wave 3
Denmark	62.4	82.8	76.7
France	79.5	89.6	
Greece	90.1	88.5	87.4
Ireland	55.8	81.8	81.7
Italy	90.7	90.9	90.7
Portugal	88.9	90.4	96.8
Spain	67.0	86.9	84.3

Table 4: Fraction of nonparticipants (percent) by country and wave.

	Wave 2	Wave 3	Wave 4	Wave 5
Denmark	24.3	25.2	26.3	28.8
France	24.2	24.1	24.2	25.8
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
Portugal	20.7	20.9	20.0	20.0
Spain	22.3	21.7	19.7	18.9

Table 5: Participation patterns by country.

	Continued particip.	Monotone attrition	New entry	Occasional nonresp.	Occasional response	Very irregular	Total
Denmark	46.8	31.9	8.1	5.1	4.9	3.2	100.0
France	58.1	26.6	8.1	2.6	3.0	1.7	100.0
Greece	55.5	27.6	10.6	1.7	2.8	1.8	100.0
Ireland	44.7	40.0	9.1	1.1	3.8	1.3	100.0
Italy	62.4	19.5	11.0	3.3	2.2	1.7	100.0
Portugal	62.4	16.0	14.6	3.0	2.6	1.5	100.0
Spain	50.4	29.6	10.9	3.9	2.9	2.3	100.0

Table 6: Causes of participation and nonparticipation by type of participation pattern.

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

Table 7: Response rates, contact rates and cooperation rates by country and year.

	1994	1995	1996	1997
	Response rate			
Denmark	.876	.850	.851	.837
France	.892	.925	.893	.895
Greece	.909	.917	.911	.879
Ireland	.815	.833	.871	.891
Italy	.944	.953	.904	.913
Portugal	.954	.939	.945	.929
Spain	.861	.892	.874	.876
	Contact rate			
Denmark	.979	.960	.968	.975
France	.970	.926	.894	.900
Greece	.911	.918	.912	.879
Ireland	.829	.993	.990	.992
Italy	.968	.977	.968	.978
Portugal	.974	.944	.949	.938
Spain	.907	.926	.889	.910
	Cooperation rate			
Denmark	.895	.886	.879	.858
France	.920	.999	.999	.995
Greece	.998	.999	.999	.999
Ireland	.983	.839	.880	.899
Italy	.976	.975	.934	.934
Portugal	.980	.995	.995	.990
Spain	.949	.963	.984	.963

Table 8: Number of nonmissing observations (Obs.), mean and standard deviation (Std. Dev.) of the predictors in the models.

Variable	Obs.	Mean	Std. Dev.
adults	323716	2.8	1.3
children	323716	.6	1.0
tmove	317413	11.8	5.9
tmovemis	323716	.019	.138
nowner	323716	.191	.393
hincome	323716	10.298	8.375
itemnr	323716	.077	.220
nvisits	322628	2.0	1.3
visitmis	323716	.003	.058
tfieldw	323716	.5	.4
hminint	323716	19.5	16.6
age	323716	45.0	18.4
female	323716	.517	.500
college	323716	.120	.325
secondary	323716	.247	.431
unemployed	323716	.072	.258
inactive	323716	.459	.498
nocohab	323716	.360	.480
nosocial	321300	.199	.399
socmis	323716	.007	.086
CAPI	323716	.193	.394
self	323716	.007	.084
tel	323716	.006	.075
proxy	323716	.109	.311
modemis	323716	.037	.188
pintid	188264	.527	.499
pintmis	323716	.418	.493
pminint	323716	21.7	11.7
int1	323716	.270	.444
int2	323716	.236	.425
int3	323716	.182	.386
Denmark	323716	.063	.243
France	323716	.153	.360
Greece	323716	.143	.350
Ireland	323716	.098	.297
Portugal	323716	.141	.348
Spain	323716	.193	.395
year95	323716	.257	.437
year96	323716	.248	.432
year97	323716	.232	.422

Table 9: 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	1.719 **	1.640 **	1.709 **	1.717 **	1.637 **	1.706 **
adults	.006		.011 *	.005		.010
children	.068 **		.068 **	.069 **		.070 **
tmove	.012 **		.011 **	.012 **		.012 **
tmovemis	.086		.072	.085		.071
nowner	-.208 **		-.203 **	-.209 **		-.204 **
hincome	-.001		-.000	-.001		-.000
itemnr	-.070 **		-.064 *	-.073 **		-.069 **
nvisits		-.064 **	-.061 **		-.065 **	-.062 **
visitmis		-.774 **	-.735 **		-.781 **	-.746 **
tfieldw		.061 **	.057 **		.058 **	.048 *
hminint		.002 **	.002 **		.002 **	.002 **
int1	.363 **	.366 **	.363 **	.363 **	.366 **	.363 **
int2	.590 **	.597 **	.590 **	.589 **	.596 **	.589 **
int3	1.011 **	1.020 **	1.011 **	1.010 **	1.019 **	1.009 **
Denmark	.030	.031	.109 **	.028	.031	.108 **
France	-.476 **	-.499 **	-.439 **	-.478 **	-.500 **	-.443 **
Greece	-.665 **	-.728 **	-.729 **	-.666 **	-.728 **	-.728 **
Ireland	-.389 **	-.289 **	-.356 **	-.389 **	-.285 **	-.348 **
Portugal	-.323 **	-.292 **	-.308 **	-.324 **	-.293 **	-.312 **
Spain	-.645 **	-.583 **	-.617 **	-.645 **	-.583 **	-.618 **
year95	.067 **	.076 **	.072 **	.070 **	.078 **	.076 **
year96	-.092 **	-.085 **	-.099 **	-.086 **	-.081 **	-.090 **
year97	-.116 **	-.104 **	-.122 **	-.108 **	-.099 **	-.110 **
k_1	19	16	23	19	16	23
$-\hat{L}$	68897.2	69307.8	68643.5			
AIC	137834.4	138649.6	137335.0			
pseudo R^2	9.2	8.7	9.6			

Table 10: 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	1.379 **	1.180 **	1.207 **	1.398 **	1.193 **	1.232 **
age	-.000		-.000	-.000		-.000
age^2	.000 *		.000	.000 *		.000
female	-.004		-.007	-.003		-.007
college	.031		.032	.032		.033
secondary	-.016		-.014	-.016		-.014
unemployed	.025		.021	.026		.021
inactive	.054 **		.050 **	.054 **		.049 **
nocohab	-.103 **		-.099 **	-.103 **		-.098 **
nosocial	-.139 **		-.137 **	-.137 **		-.134 **
socmis	-1.423 **		-1.386 **	-1.408 **		-1.358 **
CAPI		-.079	-.084		-.079	-.083
self		-.202 *	-.174 *		-.200 *	-.172 *
tel		-.290 **	-.280 **		-.289 **	-.276 **
proxy		-.100 **	-.074 **		-.100 **	-.075 **
modemis		.656 **	.667 **		.656 **	.663 **
pintid		.022	.021		.022	.021
pintmis		.403 *	.401 *		.405 *	.406 *
pminint		-.001	-.001		-.001	-.001
int1	.326 **	.329 **	.330 **	.311 **	.320 **	.308 **
int2	.563 **	.570 **	.570 **	.542 **	.557 **	.540 **
int3	.958 **	.970 **	.970 **	.931 **	.954 **	.932 **
Denmark	-.468 **	-.493 **	-.487 **	-.466 **	-.492 **	-.484 **
France	.417 **	.247 **	.290 **	.433 **	.258 **	.312 **
Greece	1.349 **	1.415 **	1.398 **	1.368 **	1.428 **	1.420 **
Ireland	-.421 **	-.401 **	-.412 **	-.410 **	-.395 **	-.397 **
Portugal	.676 **	.719 **	.715 **	.682 **	.724 **	.721 **
Spain	.164 **	.040	.017	.185 **	.054	.046
year95	.055 **	.216 **	.220 **	.053 **	.214 **	.217 **
year96	-.003	.162 **	.166 **	.001	.164 **	.172 **
year97	-.081 **	.028	.031	-.077 **	.031	.039
k_2	22	20	30	22	20	30
$-\hat{L}$	43300.0	43206.8	43070.4			
AIC	86642.0	86455.6	86202.8			
pseudo R^2	14.0	14.2	14.5			
k	41	36	53	41	36	53
$\hat{\rho}$	0	0	0	-.202 *	-.116	-.320 **
$-\hat{L}$	112197.2	112513.5	111714.1	112195.1	112513.0	111707.6
AIC	224480.4	225103.9	223537.9	224478.2	225072.0	223527.2
LR stat.				3.67 *	.84	9.30 **