

An Analysis of Attrition in the European Community Household Panel

Un'Analisi dell'Attrition nel Panel Europeo sulle Famiglie

Francesca Gallo, Sara Mastrovita, Isabella Siciliani

ISTAT – Dipartimento delle Statistiche Sociali

Direzione Centrale Indagini su condizioni e qualità della vita – Via A. Ravà, 150 –

00142 Roma; e-mail: gallo@istat.it

Riassunto: L'indagine panel europeo sulle famiglie (ECHP) è un'indagine campionaria di tipo longitudinale condotta dal 1994 con cadenza annuale nei Paesi dell'Unione europea. Le indagini panel, nonostante i vantaggi analitici rispetto a indagini trasversali, soffrono di alcune problematiche, quali l'*attrition* del campione, che risulta tanto più grave se è non casuale, cioè interessa in modo differenziato segmenti diversi della popolazione poiché può condurre a distorsioni nelle stime. Dopo la quantificazione dell'*attrition* del campione italiano nel periodo 1994-2000, si analizzano le caratteristiche degli *attritors* comparativamente a quelle di chi permane nel campione e, mediante la stima della probabilità di rimanere nel campione con modelli logit, si individuano le variabili che influenzano maggiormente l'*attrition*.

Keywords: panel survey, attrition, logit

1. Introduction

The European Community Household Panel (ECHP) is a multi-dimensional survey launched in the European Union by Statistical Office of the European Communities (Eurostat) in 1994. It involves annual interviewing of a representative sample of households and individuals aged at least 16 in each European Union Member State. It covers many topics as labour, income, education, health and demographic characteristics in order to have information on poverty and many other social indicators concerning living conditions and social exclusion. The longitudinal structure of the survey makes it possible to follow and interview the same individuals over many consecutive years to measure changes at micro level.

The initial sample in the European Union was around 60,500 households and 170,000 individuals while the Italian sample achieved 7,989 households and 24,063 individuals. Compared to cross-sectional surveys, panel surveys give more analytic advantages (Kasprzyk *et al.*, 1989; Ghellini, Trivellato, 1996; Trivellato, 1999) but, at the same time, they could suffer of some problems such as attrition. The attrition is related to the presence of units that respond to one or more of the early waves, but then leave the panel for ever.

The aim of this paper is to examine the extent and the nature of the attrition process affecting the Italian sample of ECHP from 1994 to 2000. The attrition represents one of the most important problem of panel surveys: the consequences of this phenomenon are related more to the nature than to the extent. It is therefore important to determine what kind of respondents are not participating in the survey, because the sample becomes

increasingly less representative of the population it is surveying, if particular kinds of respondents regularly leave the sample.

2. Follow up rules and evolution of the household sample

Because of the longitudinal structure of the survey, which follows the same households and individuals over time, it is necessary to define adequate follow up rules which identify the sample to follow in subsequent waves.

All residents in households in 1994 are initial sample persons. At any subsequent waves the eligible population consist of: 1) sample persons, that is all members of the initial sample, plus children born after 1994 by mothers sample persons; 2) non sample persons who reside in the same households with one or more sample persons. The survey does not follow up non sample persons who move into households not containing any sample person.

The follow up rules identify the households which are eligible for any wave. They are those interviewed at the previous wave, plus those not interviewed in the previous wave for reasons different from definitive refusal. However, households not interviewed for two consecutive times are dropped from the sample as well those out of scope (moved out the European Union, institutionalised, without sample persons). Moreover, newly formed households, resulting from the movement of sample persons, are added to the sample.

The introduction of new households and the exit of non eligible ones results in a change of the sample, as shown in table 1.

Table 1 – Italian ECHP - Sample changes in 1994-2000

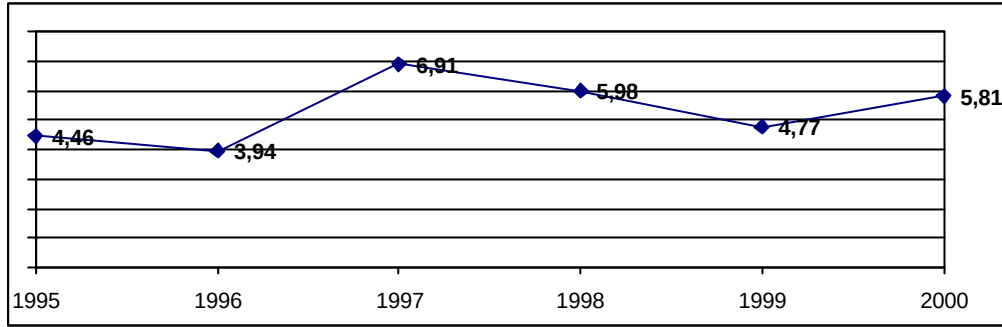
	1994	1995	1996	1997	1998	1999	2000
Household interviewed in the previous wave		7115	7128	7132	6713	6571	6370
+ Household not interviewed in the previous wave, but to include again		482	328	218	345	301	245
=Household coming from previous wave		7597	7456	7350	7058	6872	6615
+ New households	7989	305	291	312	419	264	250
- Households out of scope (without sample person, institutionalised, etc.)		60	56	69	92	90	94
- Household with a definitive refusal	35	185	208	479	415	329	390
- Households not interviewed for two consecutive waves	357*	201	133	56	98	102	86
=Households passing to the next wave	7597	7456	7350	7058	6872	6615	6295

* From 1994 to 1995 all households not interviewed for a reason different from "no contact" have been excluded.

3. Extent of attrition at individual level

In the subsequent part of the paper the unit of analysis is the individual and the subset of analysis is made of the individuals interviewed in the first wave. This subset has been divided into three groups: 1) non-attritors, individuals that have taken part into the survey each year since 1994; 2) attritors, individuals who leave the sample at a certain wave and no more return; 3) returnees made of people who do not answer some years and then start answering again. The first group amounts to 61% of those interviewed in 1994; the attritors are below 6% each year except 1997, where they are around 7%. The returnees are in general below 2%. The pattern of attritors is not regular across time, as shown in Figure 1.

Figure 1 – Italian ECHP - Attritors in 1995-2000 – (for 100 respondents in 1994)



4. The nature of attrition

In order to determine if the attrition process is random or not, some characteristics of individuals (non-attritors, attritors, returnees), measured in 1994, are taken into account in the analysis. The results based on chi-square test show that sex does not affect attrition, while age, marital status, level of education, main activity status, health conditions, social relationships and area in which the household is located affect the exit (temporary or definitive) from the sample.

Focusing the analysis on the group of attritors and non-attritors, as the extent of returnees is modest, six logit models have been estimated, one for each year from 1995 to 2000.

These models estimate the probability to be always interviewed until wave t , under the condition that the individual has been always interviewed until wave $t-1$ and not in wave t , taking into account some variables observed on sample units in wave $t-1$.

Let Y_t be a dichotomous variable assuming value 1 if the individual is interviewed in wave t and 0 otherwise and X_{t-1} the set of q variables measured in wave $t-1$, the model estimated is reported in [1].

$$P(Y_t=1|Y_{t-1}=1, \dots, Y_{t-i}=1, X_{t-1}) = \frac{\exp(b_0 + b_1 X_1 + \dots + b_q X_q)}{1 + \exp(b_0 + b_1 X_1 + \dots + b_q X_q)} \quad (t=1995, \dots, 2000) \quad [1]$$

($i=1, \dots, 6$)
where ($t-i$) ≥ 1994

The results of the model for $t=2000$ are listed in table 2. Characteristics affecting the attrition are: area in which the household is located, age of individuals, health conditions and frequency of conversation with neighbours.

More precisely, people living in the North and Centre of Italy are less likely to be interviewed compared to those living in the South of Italy; moreover, in the North-East, the probability to skip the interview is higher than the one in the North-West. People aged 25-64 are more likely to answer the interview compared to those aged 65 and more: among the age groups, people belonging to 55-64 are characterised by the highest probability of answering. As regards health conditions, individuals who declared fair, good/very good health are more likely to be interviewed compared to those in bad conditions. As far as the frequency of conversation with neighbours is concerned,

people who seem to be more friendly, that is talk to neighbours *on most days*, are more likely to keep on staying in the panel.

The attrition process is not evenly spread across all subgroups of the population in the same way. In practice, the results listed in the table demonstrate that the pattern of attrition affects particular subgroups more than others and the attrition is not random.

Table 2 – Estimate of the logit model on the probability of answering in each wave vs. exit from the sample in 2000.

	Probability of answering in each wave until 2000 variables of 1999			
<i>Individual and Household Characteristics</i>	Parameter estimate	Chi-square	p-value	Odds ratio
Intercept	1,8091	39,9	<0,0001	
Individual Characteristics				
Age		14,7	0,0053	
16-24	0,2111	2,3	0,1292	1,2350
25-39	0,2664	4,2	0,0403	1,3052
40-54	0,3287	7,3	0,0070	1,3891
55-64	0,4440	12,8	0,0003	1,5589
65 and more	0,0000	.	.	1,0000
Frequency of conversation with neighbours		21,1	<0,0001	
On most days	0,4898	19,8	<0,0001	1,6319
Sometimes	0,3061	7,6	0,0059	1,3581
Never	0,0000	.	.	1,0000
Health		23,9	<0,0001	
Good/Very good	0,4397	16,0	<0,0001	1,5522
Fair	0,5192	23,2	<0,0001	1,6806
Bad/Very bad	0,0000	.	.	1,0000
Household Characteristics				
Area in which the household live		136,2	<0,0001	
North-West	-0,7260	52,4	<0,0001	0,4839
North-East	-1,0756	132,4	<0,0001	0,3411
Centre	-0,5471	29,9	<0,0001	0,5786
South	0,0000	.	.	1,0000
Largest source of income		2,7	0,6093	
Wages and salaries	0,0988	0,2	0,6865	1,1039
Income from self-employment	-0,0393	0,0	0,8762	0,9615
Pensions	-0,0040	0,0	0,9869	0,9960
Unemployment, redundancy benefit, other benefits	0,1777	0,2	0,6773	1,1944
Income from investment, property, other sources	0,0000	.	.	1,0000
Equivalised household income		5,6	0,1300	
Up to I quartile	-0,1336	1,8	0,1780	0,8749
From I to II quartile	0,1054	1,0	0,3143	1,1111
From II to III quartile	0,0032	0,0	0,9737	1,0032
From III to IV quartile	0,0000	.	.	1,0000
<i>Deviance(b)</i>	422(18)			
<i>p-value</i>	<0,0001			
<i>Sample</i>	11895			

(b) Deviance = $-2\log(\text{likelihood of the model with intercept only} / \text{likelihood of the actual model})$. Degrees of freedom are in parentheses.

References

Ghellini G., Trivellato U., 1996, *Indagini panel sul comportamento socio-economico di individui e famiglie: una selezionata rassegna di problemi ed esperienze*, in Quintano C. (a cura di), *Scritti di statistica economica 2*, Rocco Curto Editore, Napoli, pp. 271-314.

Kasprzyk D., Duncan G.J., Kalton G., Singh M.P. (Eds.), 1989, *Panel Surveys*, John Wiley and Sons, United States.

Trivellato U., 1999, *Issues in the Design and Analysis of Panel Studies: A Cursory Review*, *Quality & Quantity*, n. 33, pp. 339-352.