NON SAMPLING ERROR CONTROL IN HOUSEHOLD SURVEYS Grazia G. Arangio-Ruiz, Central Statistical Institute of Italy

Household consumption survey

1. It would be against common sense to start a statistical survey without starting at the same time some very accurate control. Italian household consumption survey is not an exception to this. From the very beginning controls were establ<u>i</u> shed concerning each respondent household considered by itself; the aims were: - to spot errors and inconsistencies;

- to verify (for food expenses only) the balance between quantity bought and amount paid.

2. But in recent years the said controls began to prove inadequate for a num ber of reasons deriving from the growing difficulties in the collection of primary data from respondent households.

In practice, on the one hand sample Communes (1) give a less complete col laboration as they are overwhelmed by a lot of tasks, statistical and not statistical; on the other hand, households are less prepared to collaborate for many re asons:indifference, impatience, mistrust for the enumerator and his Office, fear to let an unknown person into the house (we know about interviews carried out on the landing or in the porter's lodge).

The first consequence of all that is a growth in the failures of sample units of first and second stage, which in turn brings about a reduction of sample units actually interviewed compared with the expected ones (quantitative deficien cies). The logical consequence is a growth of inflation factor and hence a rise in total error.

In the first nine months of 1982 re sponse rate was 85.3 per cent, not too b ad compared with what happens in other <u>c</u> ountries, but too low compared with that of previous years in Italy. The hearth of the matter is the territorial distribut<u>i</u> on of bad collaboration so that response rate ranges from 65 per cent in Calabria, to 96 per cent in Marche. The quality of regional data, therefore, is not the same in every region and at national level the sample of households actually interviewed is not well-balanced from a terr<u>i</u> torial point of view.

3. The second consequence is the increase of incomplete or unsatisfactory \underline{i} nterviews (qualitative deficiencies) wh<u>i</u> ch is much more serious in the presence of the sample reduction already mentioned.

This situation led to a widening of the controls in order to single out those new inadequacies.

It is worth-while pointing out the dan ger of households which agree to collaborate but for some reason, indicate only part (and a small one) of expenses met in the reference period. From the point of vi ew of survey results this behaviour is mu ch worse than that of households who do <u>n</u> ot collaborate at all.

New controls started at the end of 1979 do not consider each household separately, but compare replies given by all sample \underline{h} ouseholds of a Commune.

It is easily understood we can not obj ect to <u>one</u> household which the ten days <u>r</u> egistration period does not buy (or autoconsume) very common items such as meat, bread, cigarettes, but we can not accept that <u>all</u> (or almost all) responding hous<u>e</u> holds in a Commune do not buy nor autoco<u>n</u> sume such products or other ones of the <u>s</u> ame kind.

The aims of the controls are three in number, all of them equally important:

- to localize qualitative and quantitative deficiencies in the collection of primary data;
- to individuate incomplete replies of all (or almost all) sample households in the same Commune;
- to individuate incomplete replies of single households living in different same ple Communes.

If some irregularities are discovered, the control gives rise to two types of $i\underline{n}$ tervention:

- a feed back to the field organization <u>a</u> sking for a better collaboration in the collection of primary data;
- a correction of the results in order to avoid underestimation of expenditure.

When planning the whole correction pattern, the exigency was kept in mind of avoiding any delay in the dissemination of \underline{r} esults and at the same time bettering the ir quality with various kind of intervention, also preventive.

4. The first control - at communal level - consists of a tabulation whose pattern is given in the Annex I. The whole <u>s</u> ample of households interviewed in each <u>c</u> ommune is analysed according to the aver<u>a</u> ge monthly per capita expenditure for some food and non food items and according to the average number of affirmative replies about the ownership of sixteen durable go ds listed the questionnaire.

In addition, the table signalizes also the single maximum expenditure - in absolute value - indicated within the Commune, and the item concerned.

This very simple table emphasizes three e pathological situations which may occur in a Commune, i.e.:

- Communes showing unreliable average per capita total expenditure for the items considered;
- Communes showing reliable average per <u>c</u> apita total expenditure, but where a ho usehold reported a very high expense wh ich brought about a growth of the avera ge value;
- Communes showing very few replies as to the ownership of durable goods which ch aracterise the household level of living (i.e. Communes where very few households - or none - have a television set or a refrigerator). This is assumed as an indicator of inaccurate interview ma king.

This first control, as it was said, is not aimed at a direct correction of $d\underline{a}$ ta, but gives indications to be discussed with local authorities when pathological situations are spotted.

5. The second control we want to describe is established at the level of the \underline{t} hree big Italian geographical districts: North, Centre, South. Its aim is to single out the respondent households reporting such a low average per capita expenditure to be considered as unreliable.

This time the immediate aim is the correction of the results through the elimination of unreliable households which are replaced by other ones which passed through the control.

The most important point of this kind \underline{o} f control is how to decide which are <u>poor</u> households and which are <u>unreliable</u> ones.

As a first step, it was decided to classify all respondent households according to their expenditure pattern and three parameters were chosen among all the possible ones:

- Size of household (six possibilities);

- Socio-economic condition of the head of household (eight possibilities);
- Number of members of household having a subsistence mean of their own (seven <u>po</u> ssibilities).

The complete list of variables is given in Annex II. The result of all possible combinations are about three hundred groups of househo lds to be considered homogeneous as far <u>a</u> s expenditure is concerned. For each group, total average montly per capita expen diture is calculated and the "reliability limit" is fixed equal to the 30 per cent of such expenditure. So, if a group of ho useholds reports an average pro capita ex penditure of 600,000 lire, all households of the same group reporting an expenditure less than 180,000 lire are considered as unreliable, and eliminated.

For the entire procedure see Annex III. 6. The efficacy of the controls we have described and of the interventions deriving from them, is demonstrated by the decline in the number of unreliable hous<u>e</u> holds.

The first time, in the third quarter \underline{o} f 1979, households eliminated were 8.1 per cent of the total. In the second quarter of 1982 we arrived at 1.4 per cent; and we hope not to find any more unreliable \underline{h} ousehold by the end of 1983.

We feel this reduction is very important, of course, as the entire control and correction procedure is just a device to make up for a deficiency which must be radically eliminated.

7. The control at household level is <u>a</u> n effective one if only some households give unsatisfactory replies whilst all o<u>t</u> her ones give a good collaboration. It is theresfore necessary to control also "go-od" households in order point may be too low in some groups.

For that reason, other controls are $i\underline{n}$ troduced at any time when it is felt they are necessary. We list the principal ones:

- Expenditure on school fees, books, copy books, etc. of households having a student member. This control so far gave <u>s</u> atisfactory results;
- Expenditure on petrol of households having at least one car. At national level, 0.5 per cent of households owning <u>a</u> t least one car did not report expenditure on petrol. The proportion was est<u>e</u> emed acceptable as a car may not be used during the reference period for many reasons: driver ill or absent, car under repair, etc.;
- Expenditure relating to sport practising (sport wear and equipment, attendance at sporting centres, etc.) of households having at least one member who de claredhe/she prectises sport. The control so far gave a satisfactory result.
- Expenditure on power according to the <u>n</u>

umber of big electric appliances owned by the household.

The control showed a not very tight correlation, but still satisfactory between the two variables;

- Expenditure on subscription to RAI-TV (State Radio and Television Broadcasting Company) of households owning a television set (the subscription is compulsory in Italy). The control was started in 1983. We do not yet know the results. Labour force survey

8. Controls established from the very beginning for labour force survey are similar to those of the survey on household consumption. Each record passes through a program detecting inconsistencies and errors, which are eliminated by a rather so phisticated correction plan.

The new control started in 1980 deals with some parameters relating:

- to each Commune which is examinated before the expansion of data;
- to the whole stratum represented by the Commune under examination.

The parameters are listed in the Annex IV.

9. The anomalous situations detected by the control so far are as follows:

- Rate of activity and/or employment and/ or unemployment too low, particularly <u>f</u> or women.
- Too many old people in the total popula tion.

- Average size of household too small (2). The aim of the control is not that of correcting data or modifying them in any way, but to discuss distorsions (if any) with persons responsible for field organi sation, and to ask them (if it is necessa ry) to better the standard of their work.

Sometimes the anomaly is only an apparent one, because the Commune concerned <u>h</u> as a particular situation. For instance, there are in the country some small Communes situated into the mountains, which we re abandoned by the young people and have actually low activity rate and small households.

We count also, of course, on the psycological effect: certainly persons work be tter if they know their work is well controlled.

10. Another control, established only in 1982, was the comparison of average size of household resulting from the <u>d</u> emographic census of October 1981 and from quarterly labour force survey carried out in the same month. The comparison was made for each of the 95 provinces and the 95 Communes chief of province.

Only in three cases the differences we re disturbing, i.e. in the provinces of Bergamo, Rovigo and Palermo.

For the first two - Bergamo and Rovigo - the average size resulting from labour force survey is bigger than that resulting from demographic census, both for the whole province and for the Commune chief of province.

For Palermo we find the opposite: the average size resulting from labour force survey is smaller, both for the whole province and the Commune chief of province.

The results of labour force enquiry of October 1982, one year after our control, give us a different picture, as follows:

- For Bergamo and Rovigo:differences in hou sehold average size almost desappeared at provincial level and became smaller at communal level;
- For Palermo: differences in household <u>a</u> verage size remained practically the sa me at provincial level and became a lit tle smaller at communal level (see Annex V).

The data resulting from labour force <u>s</u> urvey relating to Palermo (whole province and Commune chief of province) are particularly puzzling because of their discrepancy compared with

what we know of the demographic and social background of Italy, of the South of \underline{I} taly and, more particularly of Sicily. We cannot accept, actually, that in Palermo average size of household is smaller than in the whole of Italy.

The first tests suggest the possibili ty of a distortion happened in the phase of selection of sample households from <u>po</u> pulation registers, but the problem is still under discussion with local authorities concerned.

Final remarks

11. To introduce continuosly new cont<u>r</u> ols in time is not only a way to ensure a good quality standard of data deriving f<u>r</u> om the two surveys we are speaking of, but it is also a way to keep them continuosly up to date and to oblige them to put into evidence new situations - anomalous or n-ot anomalous - which take place in the "<u>w</u> orld of household".

I can instance the case of a deficiency which occurred some years ago in the <u>i</u> ndication of expenditure on heating. We <u>f</u> ound it was caused by a modification in <u>t</u> he termes of payment asked by the suppliers of fuel as a consequence of energy crisis and of the rises in costs. This cha nge consisted in instalments less numerous, higher and paid in advance.

The registration system was no more ef fective to catch them; it was necessary, therefore, to give new instructions to ho useholds and interviewers.

A change in household way of life, pro mptly reflected by the survey.

Notes

- (1) See Appendix
- (2) We want to point out that labour f orce survey has always understima ted average size of household comparing with demographic census; on ly as for the 1981 census the datu m was the same: 3.00.

Appendix

The administrative territorial structu re of Italy is as follows:

- 20 Regions
- 95 Provinces
- 8,000 Communes

Communes are the smallest administrati ve districts and form the basis of sample design of household consumption and labour force surveys.

The sample of household consumption s urvey includes every year about 700 Commu nes and 38,000 households resident in the said Communes.

The sample of labour force survey includes every quarter 1,921 Communes and 123,000 households resident in the said C ommunes.

For both surveys field work, including selection of households from population r egisters, interview, first control of fil led up questionnaires, is the responsability of Communal authorities.

Annex I

ITEMS CONSIDERED FOR EXPENDITURE CONTROL AT COMMUNE LEVEL.

- A. Average monthly per capita expenditure - Food
- Non food
- Total
 - Meat
 - Oil and fats
 - Fruit and vegetables
 - Rent
 - Heating
- B. Maximum expenditure reported within th e Commune
- C. Average number of affirmative replies as to the ownership of durable goods.

Annex II

VARIABLES CONSIDERED FOR THE STRATIFICATI

ON OF HOUSEHOLDS

- A. Size of household
- 1 member
- 2 members
- 3 members
- 4 members
- 5 members
- 6 or more members
- B. Socio-economic condition of the head o f household
- Entrepreneur and professional
- Self-employed in agriculture
- Self-employed in non agricultural sectors
- Executive and employee
- Agricultural worker
- Worker in non agricultural sectors
- Person not in labour force
- Other
- C. Members of household having a subsiste nce mean of their own.
- Head of household only
- One member in addition to head of household
- Two members in addition to head of household
- Three members in addition to head of h ousehold
- Four members in addition to head of ho usehold
- Five members in addition to head of ho usehold
- Six or more members in addition to head of household

<u>Annex III</u>

SPOTTING AND ELIMINATION OF HOUSEHOLDS EN TERING EXPENDITURE CONSIDERED AS NOT RELI ABLE

- First phase
- District
- Household size
- Condition of head of household
- No. of members whith a subsistence mean of their own
- Average monthly per capita expenditure of household
- Number of households
- Number of households entering a monthly per capita expenditure below the re liability limit
- Second phase

Within each district - Ventilation of h ouseholds entering monthly per capita expenditure below the reliability limi t according to: month of participation to the survey; Commune and province of residence; average expenditure separatedly for food and non food items.

Third phase

The households, after the ventilation

are replaced by other households of the same category, reporting a reliable average expenditure, chosen at random.

<u>Annex IV</u>

CONTROL OF COMMUNES AT SAMPLE LEVEL

- Activity rate, by sexEmployement rate, by sex
- Unemployement rate, by sex
- Youth rate: percentage of persons 0 -
- 13 old on total population
- Old age rate: percentage of persons 65 years old and more on total population
- No. of sample households and members thereof
- No. of stratum households and members thereof
- Average size of sample households

<u>Annex V</u>

AVERAGE SIZE OF HOUSEHOLDS IN SOME PROVI-NCES

			Commune chi ef of prov.	Province
Bergamo	0			
Census,Oct.		1981	2.74	3.03
LF,	October	1981	3.92	3.22
LF,	October	1982	3.46	3.07
Rovigo				
Census,Oct.		1981	2.85	3.02
LF,	October	1981	3.58	3.36
LF,	October	1982	3.43	3.05
Palermo				
Census,Oct.		1981	3.35	3.26
LF,	Octobre	1981	2.66	2.84
LF.	Octobre	1982	2.85	2.83