Italy 1989: Survey Information

Summary table

Generic information	
Name of survey	Survey on Household Income and Wealth – SHIW *(Indagine sui Bilanci delle Famiglie Italiane)*
Institution responsible	Bank of Italy
Frequency	Every two years
Survey year / Wave	1989
Collection period	May to July 1990
Survey structure	Cross-sectional and panel
Coverage	Private households in the whole territory
Geographic information	20 administrative regions (more detailed info on provinces is not available for external users)
Files delivered	Two sets of files, one for the historical database and one for the annual database, each of which comprising several files at different levels (household, individual, pensions, transfers, employment activities, real estates, etc.).
Sample size	
Households	8,274 households
Individuals	25,150 individuals (of which 13,864 income earners and 21,002 aged over 15)
Sampling	
Sampling design	*Initial sample (1987):* two-stage stratified sampling, with the stratification of the PSUs (municipalities) by region and demographic size. *Subsequent samples:* households residing in panel municipalities that had signaled an availability to re-interview in the preceding survey were all included in the sample; the non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities.
Sampling frame	Municipal registry office records.
Questionnaires	Paper-based questionnaire with a modular structure: general part addressing aspects relevant to all households and a series of annexes with questions relevant to specific subsets of households.
Standard classifications	
Education	6 categories
Occupation	6 categories for employees, and 6 for self-employed, used as labour force status and not occupation
Industry	9 sectors
Income	
Reference period	Income in the preceding calendar year (which coincides with the fiscal year)
Unit of collection	Mostly at the individual level, except for property income (household level)
Period of collection	Mostly monthly income with number of months, some annual.
Gross/net	All variables are recorded net of taxes and contributions.
Data editing / processing	
Consistency checks	Standard post-survey consistency checking procedure by the data collection company.
Weighting	Survey data can be grossed up to aggregate values thanks to appropriate weights assigned to each household according to its probability to be included in the survey.
Imputation	All the elementary variables that make up the aggregates are imputed; regression models are used to estimate the values to assign to the missing answers on the basis of other available information that is correlated with the missing data.

This document draws extensively upon the methodological Annex to the "I bilanci delle familglie italiane nell'anno 1989", *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year I, No. 26, October 1991 (see **http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/boll_stat/supplemento famiglie 1989 n.26_91.pdf**).

Table of contents:

A. General characteristics

Official name of the survey/data source:
Survey on Household Income and Wealth – SHIW *(Indagine sui Bilanci delle Famiglie Italiane)*

Administrative Unit responsible for the survey:

Bank of Italy
Research Department
Divisione Rilevazioni e Metodi Statistici - R.M.S.
Address: Via Nazionale 91, 00184 ROMA.
WWW: http://www.bancaditalia.it
e-mail: studi.indagini@insedia.interbusiness.it

The Survey on Household Income and Wealth (SHIW) began in the 1960s within the Research Department of the Bank of Italy with the aim of gathering data on the incomes and savings of Italian households. Over the years, the scope of the survey has grown and now includes wealth and other aspects of households' economic and financial behaviour such as, for example, which payment methods are used.

Until 1987 the Bank of Italy's survey of Italian household budgets was conducted with time-independent samples of households. In order to facilitate analysis of changes in the phenomena being investigated, since 1989 part of the sample has comprised households that were interviewed in previous surveys (panel households). The sample used in the most recent surveys (carried out generally every other year) comprises about 8,000 households (24,000 individuals), distributed over about 300 Italian municipalities and it is representative of the whole Italian population.

The survey results are regularly published in the Bank's Supplements to the Statistical Bulletin. The data on the households is freely available, in an anonymous form, for further elaboration and research. Other than the annual datasets, a historical database has been put together comprising a slightly restricted selection of variables available in the normal annual files but consistent over all the years since 1977.

B. Population, sampling size and sampling methods

Coverage

The sample is representative of the whole Italian population.

Sample size

Table 1a shows the sample size used between 1987 and 2000, indicating the number of households interviewed in more than one survey. For example, of the 8,274 households that made up the sample in this survey, 1,206 had participated since 1987 and the remaining 7,08 were being interviewed for the first time.

Table 1a
Households interviewed in the 1987-2000 surveys

Year of first interview	Year of survey						
	1987	1989	1991	1993	1995	1998	2000
1987	8,027	1,206	350	173	126	85	61
1989		7,068	1,837	877	701	459	343
1991			6,001	2,420	1,752	1,169	832
1993				4,619	1,066	583	399
1995					4,490	373	245
1998						4,478	1,993
2000							4,128
Sample size	**8,027**	**8,274**	**8,188**	**8,089**	**8,135**	**7,147**	**8,001**

The overall size of the sample for the 1989 survey was 8,274 households. The proportion of panel households was 14.5 per cent.

Sampling design

As already mentioned, in this survey the sampling design was modified to foresee the re-interview of a proportion of the households already interviewed in the previous survey (rotating panel).

The sample, including as in the previous survey about 8,000 households, was divided in two parts: the first constituted by a subset, or rather a sample, of the households interviewed in 1987 (panel part), and the second by a sample of households that are

selected for the first time in this survey (non-panel part).[1] The size of the two parts of the sample was determined following the results of a pilot survey, carried out with the purpose of estimating the portion of the households interviewed in 1987 which would be available to be re-interviewed in 1989. The results of this pilot survey highlighted the difficulties of obtaining a new interview: the portion of households available for re-interview is in fact lower than that of the households interviewed for the first time in 1987.[2] Taking thus into account the necessity of contacting a high number of households in order to interview only a fraction of them, the number of panel households was fixed at 1,200, i.e. about 15 per cent of the households interviewed in 1987.

The election of panel households was carried out following a sampling design analogous to that adopted in 1987. More specifically, considering that the 1987 sample was drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size,[3] it was first necessary to select the panel municipalities, i.e. those already selected in 1987, and subsequently the households to re-interview. All municipalities with a population of more than 40,000 were included among the panel municipalities, and the portion of households to interview in those municipalities was fixed to 15 per cent. The other municipalities, that in 1987 had been selected from regional stratum with probability proportional to the size of the municipality, were instead sub-sampled with constant probability.[4] The portion of households to re-interview in those municipalities was fixed in such a way that even here the proportion of panel households result equal to 15 per cent.

Households were interviewed in 294 municipalities of which 261 were panel households and 33 non-panel households (Table 2a).

Table 2a
Survey municipalities

[1] For several reasons, it could seem preferable to adopt a panel scheme without rotation, i.e. that foresees the re-interview of all families already interviewed in 1987. However, it should be stressed that, other than the objective problems of contacting and availability of the households, which make it necessary only a partial re-use of those, rotation is used in order to preserve over time the represenatativity of the sample selected with respect to the universe.

[2] It should be stressed that households interviewed in the previous survey were not informed of the possibility of a new interview in the future, and this could have contributed to make households less available to the second contact.

[3] Since the 1986 survey, the stratification and selection criteria for the primary sampling units used by ISTAT for the Labour Force Survey have been adopted.

[4] As a result, the probability of a municipality to be included in the 1989 sample (probability given by the product of the probabilities of being included in the 1987 sample by a constant) is still proportional to the demographic dimension (assumed to be constant between 1987 and 1989) of the sample.

Geographical area	Panel	Non-panel	Total
North	96	12	108
Centre	60	9	69
South and Islands	105	12	117
Total	**261**	**33**	**294**

C. Data collection and acquisition

Data collection

The interviews for the sample survey of Italian household budgets in 1989 were conducted between May and July 1990.

Interviews were conducted by a specialized company using professional interviewers. The interview stage was preceded by a series of meetings at which Bank of Italy officials and representatives of the company gave instructions directly to the interviewers. The households contacted for interviews, who are guaranteed complete anonymity, receive a booklet describing the purpose of the survey and giving a number of examples of the ways in which the data are used. The participating households may request a copy of the results of a previous survey.

Interviewers contacted 22,344 households, of which just above one third (37 per cent) agreed to be interviewed (Table 3a). The sample was therefore composed of 8,274 households, of which 1,208 were panel households and 7,066 non-panel households. The participation rate was lower for panel households (23.3 per cent, compared with 41.2 per cent for non-panel households).

Table 3a
Households contacted and reason for non-participation (percentages)

	Panel	Non-panel	Total
Households:			
Interviewed	23.3	41.2	37.0
not interviewed	76.7	58.8	63.0
Total	***100.0***	***100.0***	***100.0***
Reasons for non-interview:			
Fear	35.0	45.7	42.7
Not contacted	25.8	28.1	27.4
Refusals	30.2	21.9	24.3
Other	9.0	4.3	5.6
Total	***100.0***	***100.0***	***100.0***

The most common reason for non-participation was the fear (42.7 per cent; Table 3a); this factor results less relevant for panel households. In 27.4 per cent of cases the reason for non-interview was caused by the impossibility to contact the household, with similar

rates for panel and non-panel households. In 24.3 per cent of cases, the household refused to be interviewed; the frequency of this reason is higher among panel households. In the remaining 5.6 per cent of the cases, it was not possible to obtain the interview for other non-specified reasons.

The questionnaire

The questionnaire used in the survey has a modular structure. It is composed of a general part addressing aspects relevant to all households and a series of annexes with questions relevant to specific subsets.

The questionnaire for panel households reported also some information given by the same family in the previous interview, in order to help the interviewer to track and remedy, in presence of the household, inconsistencies in the responses as emerging from two different surveys.

D. Definition of the survey units

Household

The basic survey unit is the household, which is defined as "a group of individual linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes". Are to be included in the household all persons that normally lived in the dwelling at 31 December 1989 who contributed at least part of their income to the household; this includes any members temporarily absent (e.g. on vacation, temporarily away for study, etc) and any non-relatives that lived stably in the household at 31 December 1989.

Head of household

The head of the household is defined at the survey stage as the person who says he/she has "the most responsibility for family finances", which satisfies the need to determine the best informed person.

E. Contents

The survey contains information about:

- demographic characteristics of the household members (including education) and the parents of the head and spouse;
- employment status of the household members aged 15 or over (incl. employment opportunities and lifetime work experience);
- questions on working times (only for workers) and on future working perspectives for workers and job searchers (rotation section);

- incomes from payroll employment, self-employment, pensions, other transfers for all adult members of the household;
- payment instruments and forms of savings of the household;
- principal residence and other property of the household;
- consumption and other family expenditures;
- forms of insurance of the household (life insurances, private pensions and annuities, health insurances, accident insurances);
- information to be provided by the interviewer.

F. Quality of data

Quality of estimates

Households that could not be interviewed were replaced by others selected randomly in the same municipality.[5] Obviously, this technique does not eliminate the risk of obtaining samples in which the less-cooperative segments of the population are underrepresented, thus generating biased estimates *(selection bias*). A recent study carried out on the data from the 1989 survey have nevertheless suggested that the bias of the estimates due to non-participation is small, thanks in part to the measures taken.[6]

Checking data and imputing missing data

The questionnaires are checked first of all to verify that no annexes are missing, no questions have been skipped and that there are no editing errors. In this phase the codification of open-answer fields is carried out (i.e. the answer "other – please specify"). Data are subsequently entered into magnetic support and automatic checks are carried out to verify the consistency of single fields or correlated fields.

Once the checks were completed, work began on imputing missing answers, which could have been due to reticence on the part of the interviewee or difficulties that respondents had in replying to the question. It is necessary to impute answers for all the elementary variables that make up the aggregate, since the absence of even one component would prevent calculation of the aggregate (for example, it is necessary to impute fringe benefits such as lunch coupons in order to calculate income from salaried employment).

Regression models are used to estimate the values to assign to the missing answers on the basis of other available information that is correlated with the missing data. In order to avoid an excessive concentration around average values, a random component is added, extracted from a normal variable, with a mean of zero and a variance equal to that of the

[5] The substitution of non interviewed households was carried out initially through reserve lists, compiled following the same criteria of the basic lists; in a limited number of cases, in order to facilitate the survey, after having unsuccessfully contacted the reserve households, the neighbouring households were interviewed instead.

[6] See L. Cannari and G. D'Alessio, *Mancate interviste e distorsione degli stimatori*, Temi di Discussione del Servizio Studi, no. 172, Banca d'Italia, Rome, June 1992. With reference to the 1989 survey, the authors estimate that household income was understated by 5 per cent owing to non-participation.

residuals in the regression model. This preserves the mean and variance of the data actually measured.

Weighting: the sample estimates

In a two-stage sampling design, with stratification of primary sampling units, an estimator of the mean value of variable x s given by:

$$(1)\quad \bar{x} = \frac{\sum_h \sum_i \sum_j w_{hij} x_{hij}}{\sum_h \sum_i \sum_j w_{hij}}$$

where x_{hij} is the value of variable x for the j^{th} household sampled in the i^{th} municipality of the h^{th} stratum and w_{hij} is the weight to give to the same household taking the sampling probabilities into account. More specifically, the sampling design adopted makes unit sampling weighs in the two strata constant at the municipality level:

$$(2)\quad w_{hij} = w_{hi} = \begin{cases} \dfrac{P_h}{\sum_i P_{hi}} \dfrac{P_{hi}}{n_{hi}} \\ \dfrac{1}{m_h} \dfrac{P_h}{n_{hi}} \end{cases}$$

respectively for municipalities with more than 40,000 inhabitants and for municipalities with up to 40,000 inhabitants, where P_h, $\sum_i P_{hi}$ and m_h are respectively the resident population, that of the municipalities in the survey and the number of sample municipalities in the h^{th} stratum, and P_{hi} and n_{hi} are respectively the population and the number of respondents in the i^{th} municipality of the h^{th} stratum.[7]

The calculation of weights (2) is carried out separately for the panel and non-panel part of the sample. The estimator (1) thus becomes:

$$(3)\qquad \bar{x} = Q\left(\bar{x}^q\right) + P\left(\bar{x}^p\right)$$

where P denotes the portion of panel households, Q=1-P the portion of non-panel households, and $\bar{x}^p$ and $\bar{x}^q$ the means of variable x for the respective subsets, calculated taking the weighs (2) into account.

When analysing the data, it was recognised though that, with respect to 1987, the socio-demographic characteristics of the panel households differ somewhat from those of the

[7] Expressions (2) are calculated in a second stage, taking the number of interviews effectively carried out into account. The procedure consists thus of a post-stratification that accounts for the response rate at the municipality level.

entire sample, and panel household have thus biased sample means.[8] Taking into account the positive correlation that links variables in successive surveys, it is best likely to assume that the estimator (3) is biased in the same direction. As a result, a post-stratification was carried out in such a way that the panel part of the sample be representative of the entire sample selected in 1987. This, other than removing the afore-mentioned distortions, allows reducing the variability of the estimates as far as the panel part is concerned.

As is known, in presence of correlation of values taken by the same variable *x* in two successive surveys, the estimator (3) is not optimal in terms of efficiency. The optimal estimator of the mean of *x* at time *t* is given by:

$$(4) \qquad \bar{x}_t^* = a(\bar{x}_t^q) + (1-a)\left(\bar{x}_t^p + r\left(\bar{x}_{t-1} - \bar{x}_{t-1}^p\right)\right)$$

$$(5) \qquad \text{with} \quad a = \frac{Q\left(1-r^2 q\right)}{1-r^2 Q^2}$$

and where *r* is the correlation coefficient between x_t and x_{t-1}.

The estimator (4) is not a simple weighted average of the values measured at time t, since, in addition to the correlation coefficient, it refers to the values of x from the previous survey for the panel and the total sample. However, following the post-stratification described above, the main variables approximately satisfy:

$$(6) \qquad \bar{x}_{t-1} = \bar{x}_{t-1}^p$$

and the estimator (4) becomes:

$$(7) \qquad \bar{x}_t^+ = a\bar{x}_t^q + (1-a)\bar{x}_t^p$$

which is different from that defined in (3) only because of the relative weigh assigned to the panel part and to the non-panel part. This is the same as weighing the sample units with the following weighs:

$$(8) \qquad w_{hij}^* = \begin{cases} w_{hij}\dfrac{a}{Q} \\ w_{hij}\dfrac{(1-a)}{P} \end{cases}$$ respectively for non-panel households and panel households.

[8] To this different composition, contribute, other than self-selection phenomena linked to the different probabilities of various household typologies in taking part to the survey, also definitory factors, such as, for example, the fact that panel households cannot include households that were created or finished in the period between the two surveys, as well as factors connected with mobility of household residency.

The coefficient *a*, used for the re-weighing of w_{hij}, is a function of the autocorrelation coefficient *r*. In theory, it could be considered that the latter coefficient is variable with respect to the considered aggregate; however, this would make the data management very complex, as in that case even the weighs assigned to each household would be variable with the considered aggregate. In order to avoid such inconvenient, *r* was assumed to be constant at 0.70, which represents an intermediate level between the values of *r* calculated for different aggregates.[9]

As a result of the re-weighing (8) and the value of *r* fixed as above, the panel section assumes a relative weigh equal to 0.25. This weigh is higher in the *P* portion (equal to 0.15), since in the described methodology it was considered that the panel also incorporates part of the information on x_{t-1}, information that can contribute to improve the precision of the estimate of the latter value (since it is correlated with the information on x_t. The weigh relative to the non-panel section, on the other hand, is reduced from 0.85 to 0.75. in this situation the variance of the estimator (7) results about 10 per cent lower than that of the estimator (3).

Standard errors

The standard errors of the means of the main variables, calculated taking the sampling design into account, are shown in Table 4a.

Table 4a
Standard errors in the estimation of the means for the main variables
(thousands of lire, percentages)

Variables	Standard error	
	Absolute value	% of estimate
Household income	549	1.57
Household consumption	334	1.30
Household net wealth	3,726	3.00

The variability of the estimators used is much lower for consumption and income than for net wealth. This seems to be determined essentially by the variability of the same phenomena, since the total effect of the sampling design results, in each of the three cases, estimable between 1.9 and 2.0.

G. Uses of the survey

Publications

The results are regularly published in the Supplements to the Statistical Bulletin of the Bank of Italy. A whole bibliography of the research carried out using data from the

[9] For example, the correlation between household income in 1987 and 1989 is equal to 0.72. The same correlation is equal to 0.64 for household consumption and to 0.62 for household net wealth.

SHIW is available (in Italian) from the Bank of Italy web-site (http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/altre/biblio.pdf). A list of the Economic Research Papers of the Bank of Italy concerning the SHIW is reported here:

E. Battistin, R. Miniaci and G. Weber (2003), What do we learn from recall consumption data?, Bank of Italy, *Temi di Discussione*, N. 466.

Giovanni D'Alessio and Ivan Faiella (2002), Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth, Bank of Italy, *Temi di Discussione*, N. 462.

Silvia Magri (2002), Italian households' debt: determinants of demand and supply, Bank of Italy, *Temi di Discussione*, N. 454.

Guido de Blasio and Sabrina Di Addario (2002), Labor market pooling: evidence from Italian industrial districts, Bank of Italy, *Temi di Discussione*, N. 453.

A. Brandolini, P. Cipollone and P. Sestito (2001), Earnings dispersion, low pay and household poverty in Italy, 1977-1998, Bank of Italy, *Temi di Discussione*, N. 427.

Andrea Brandolini and Piero Cipollone (2001), Multifactor Productivity and Labour Quality in Italy, 1981-2000, Bank of Italy, *Temi di Discussione*, N. 422.

Piero Cipollone (2001), Is the Italian Labour Market Segmented?, Bank of Italy, *Temi di Discussione*, N. 400.

G. D'Alessio e L. F. Signorini (2000), Disuguaglianza dei redditi individuali e ruolo della famiglia in Italia, Bank of Italy, *Temi di Discussione*, N. 390.

Andrea Brandolini (1999), The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality, Bank of Italy, *Temi di Discussione*, N. 350.

Poverty and income distribution

According to the publication "I bilanci delle familglie italiane nell'anno 1989", *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year I, No. 26, October 1991, the Gini coefficient of concentration is 0.334 for the distribution of household income.