# Italy 1991: 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 | 1991 |
| Collection period | May to October 1992 |
| 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 | 8,188 households |
| Individuals | 24,930 individuals (of which 13,882 income earners and 21,016 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 plus a rigorous check of the latter by a certification 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 famiglie italiane nell'anno 1991", *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year III, No. 44, July 1993 (see **http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/boll_stat/ supplemento famiglie 1991 n.44_93.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 8,188 households that made up the sample in this survey, 350 had participated since 1987 and 1,837 since 1989. The remaining 6,001 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 1991 survey was 8,188 households. The proportion of panel households was 26.7 per cent, registering a high increase with respect to the previous survey (14.5 per cent in 1989).

Sampling design

The sample was drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size. Within each stratum, the municipalities in which interviews would be conducted were selected by including all municipalities with a population of more than 40,000 and randomly selecting smaller towns. The individual households to be interviewed were then selected randomly.

In order to form the panel, the municipalities were selected from among those already sampled in the 1989 survey (panel municipalities); panel-households were selected in these municipalities among those who had signaled an availability to re-interview in the preceding survey. The non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities. Households were interviewed in 283 municipalities of which 233 were panel households and 50 non-panel households (Table 2a).

**Table 2a**
**Survey municipalities**

| Geographical area | Panel | Non-panel | Total |
|---|---|---|---|
| North | 92 | 25 | 117 |
| Centre | 49 | 10 | 59 |
| South and Islands | 92 | 15 | 107 |
| **Total** | **233** | **50** | **283** |

## C. Data collection and acquisition

<u>Data collection</u>

The interviews for the sample survey of Italian household budgets in 1991 were conducted between May and October 1992.

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 25,210 households, of which about one third (32.5 per cent) agreed to be interviewed (Table 3a). The sample was therefore composed of 8,188 households, of which 2,187 were panel households and 6,001 non-panel households. The participation rate was higher for panel households (52.9 per cent, compared with 28.5 per cent for non-panel households). This difference is due to the fact that panel households were selected among those that, in the previous survey, had signaled their availability to be again interviewed.

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

| | Panel | Non-panel | Total |
|---|---|---|---|
| Households: | | | |
| Interviewed | 52.9 | 28.4 | 32.4 |

| | | | |
|---|---|---|---|
| not interviewed | 47.1 | 71.6 | 67.6 |
| **Total** | **100.0** | **100.0** | **100.0** |
| Reasons for non-interview: | | | |
| refusals | 40.6 | 57.7 | 55.7 |
| not contacted | 55.0 | 39.1 | 40.9 |
| Other | 4.4 | 3.2 | 3.4 |
| **Total** | **100.0** | **100.0** | **100.0** |

The most common reason for non-participation was the impossibility to contact the household (55.7 per cent; Table 3a). In 40.9 per cent of cases, the household refused to be interviewed. In the remaining 22.98 per cent of the cases, it was not possible to obtain the interview for other 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 1991 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 1991.

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. 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.[1]

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.

The operations of the data collection company, and more specifically of the interviewers, was submitted to a rigorous check by a certification company. The check had the purpose of verifying that: *i)* the interviewers, in the contact phase and during the interviews, follow all the indications provided to them, and *ii)* the collection company, during the phases of codification and cleaning, adopt the right instruments to find and correct data errors and inconsistencies.

---

[1] 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.

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

Weighting: the sample estimates

The estimation procedure, which is similar to that used in the last survey, consists of three stages:

*a) Calculation of the sampling weights for households*
Each member of the household is assigned an initial weight defined as the inverse of his/her probability of inclusion in the sample. Given the sample design, the coefficient is constant at the municipality level and is equal to:

$$(1) \quad w_{hi} = \begin{cases} \dfrac{P_h}{\sum P_{hi}} \dfrac{P_{hi}}{n_{hi}} \\[2em] \dfrac{1}{m_h} \dfrac{P_h}{n_{hi}} \end{cases}$$

respectively for municipalities with more than 40,000 inhabitants and for municipalities with up to 40,000 inhabitants, where $P_h$, $\sum P_{hi}$ 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.

*b) Post-stratification of the panel households*
The socio-demographic characteristics of the panel households may differ somewhat from those of the entire sample in 1991, mainly owing to missing interviews. In order to correct for this possible source of distortion in the estimates, the panel section of the sample is post-stratified on the basis of a number of characteristics of the previous survey (geographical area, income classes, professional status of head of household) so as to modify the initial weight of this subset of households.

*c) Estimation of aggregates*

An unbiased estimator of the mean of variable x is given by the Horwitz-Thompson estimator:

$$(2) \qquad \bar{x} = \frac{\sum x_j w_j}{\sum w_j} \quad j = 1, \dots, N$$

However, if the values of variable x measured on two successive waves are correlated, an optimal estimator of the mean is given by:

$$(3) \qquad \bar{x}_t^* = a\bar{x}_t^q + (1-a)\bar{x}_t^p + (1-a)r\left(\bar{x}_{t-1} - \bar{x}_{t-1}^p\right)$$

$$(4) \qquad \text{with} \quad a = \frac{Q(1-r^2 Q)}{1-r^2 Q^2}$$

where $\bar{x}_t$ and $\bar{x}_{t-1}$ are respectively the means of variable $x$ at time t and time t-1, $\bar{x}_t^p$ and $\bar{x}_t^q$ are the means of variable $x$ at time t for the panel and non-panel parts of the sample respectively, $r$ is the correlation coefficient between $\bar{x}_t$ and $\bar{x}_{t-1}$ and $Q$ is the share of non-panel households.

The estimator (3) 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:

$$(5) \qquad \bar{x}_{t-1} = \bar{x}_{t-1}^p$$

and the last term of (3) disappears. In addition, given that the correlation coefficients for the main variables examined are between 0.6 and 0.7, giving $r$ the intermediate value $\tilde{r} = 0.65$, it is possible to approximate the estimator (3) by way of:

$$(6) \qquad \bar{x}_t^+ = a\bar{x}_t^q + (1-a)\bar{x}_t^p$$

which is obtained as the mean of the data measured at time t, weighted with coefficients equal to:

$$(7) \qquad w_{hij}^* = \begin{cases} w_{hij}\dfrac{1-a}{1-Q} \\[2mm] w_{hij}\dfrac{a}{Q} \end{cases} \quad \text{respectively for panel households and for non-panel}$$

households. This estimator differs from (2) since, being based on the positive correlation between the data gathered from the same households in successive surveys, it gives a higher relative weight to the panel segment of the sample than the share of panel

interviews actually conducted (34.5 per cent compared with 26.7 per cent), with a corresponding reduction in the weight assigned to the non-panel households. Since this reweighing could change the structure of the sample, the final sample is modified to assume the same characteristics as the population with regard to sex, age group, geographic area and size of municipality of residence.

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 | 531 | 1.42 |
| Household consumption | 396 | 1.40 |
| Household net wealth | 5,127 | 3.17 |

**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 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 1991", *Supplementi al Bollettino Statistico – Note metodologiche e informazioni statistiche*, Bank of Italy, Year III, No. 44, July 1993, the Gini coefficient of concentration is 0.325 for the distribution of household income.