

Italy 1998: Survey Information

Summary table

| Generic information | |
|----------------------------------|---|
| Name of survey | Survey on Household Income and Wealth – SHIW (<i>Indagine sui Bilanci delle Famiglie Italiane</i>) |
| Institution responsible | Bank of Italy |
| Frequency | Every two years |
| Survey year / Wave | 1998 |
| Collection period | February to July 1999 |
| 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 | 7,147 households |
| Individuals | 20,901 individuals (of which 12,717 income earners and 17,769 aged 15 and over) |
| Sampling | |
| Sampling design | <i>Initial sample (1987)</i> : two-stage stratified sampling, with the stratification of the PSUs (municipalities) by region and demographic size. <i>Subsequent samples</i> : households residing in panel municipalities that had participated in at least 2 surveys were all included in the sample; the remaining panel households were selected randomly from among those interviewed in the previous survey only; the non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities. |
| Sampling frame | Municipal registry office records. |
| Questionnaires | Both CAPI (two thirds) and paper-based questionnaires were used; both have 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 | 8 categories (only 6 available in historical database) |
| Occupation | 6 categories for employees, and 6 for self-employed, used as labour force status |
| Industry | 10 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 | The CAPI survey method performs a number of checks, making it possible to remedy any inconsistencies in the data supplied directly in the presence of the household. Standard post-survey consistency checking procedure was used for the interviews conducted with the paper-based questionnaire. |
| 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 “Italian Household Budgets in 1998”, *Supplements to the Statistical Bulletin – Methodological Notes and Statistical Information*, Bank of Italy, Year X, No. 22, April 2000 (see (http://www.bancaditalia.it/statistiche/ibf/statistiche/ibf/pubblicazioni/boll_stat/shiw98.pdf)).

Table of contents:

- A. General Characteristics
- B. Population, sample size and sampling methods
- C. Data collection and acquisition
- D. Definition of the survey units
- E. Contents
- F. Quality of data
- G. Uses of the survey

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 7,147 households that made up the sample in this survey, 85 had participated since 1987, 459 since 1989, 1,169 since 1991, 583 since 1993, and 373 since 1995. The remaining 4,478 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 1998 survey was 7,147 households. The proportion of panel households was 37.3 per cent.

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 1995 survey (panel municipalities). Households residing in these municipalities that had participated in at least two surveys were all included in the sample; the remaining panel households to be interviewed were selected randomly from among those interviewed in the previous survey only.¹

The non-panel households were selected randomly from municipal registers in both panel and non-panel municipalities. Households were interviewed in 318 municipalities of which 304 were panel households and 14 non-panel households (Table 2a).²

Table 2a
Survey municipalities

| Geographical area | Panel | Non-panel | Total |
|-------------------|------------|-----------|------------|
| North | 126 | 4 | 130 |
| Centre | 67 | 2 | 69 |
| South and Islands | 111 | 8 | 119 |
| Total | 304 | 14 | 318 |

C. Data collection and acquisition

Data collection

The interviews for the sample survey of Italian household budgets in 1998 were conducted between February and July 1999.

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.³ The participating households may request a copy of the results of a previous survey.

¹ As in the previous survey, in order to obtain information on intergenerational aspects, all households that had been established out of the original panel households were also contacted (these were normally new households formed by the children of the original household). There were a total of 38 such households, compared with 102 in the previous survey. The decline is attributable to the greater difficulty that the firm conducting the survey had in tracing such households.

² As in the previous survey, panel households that had changed their residence were, as far as possible, interviewed at their new address even if this was in a different municipality, as long as it was in Italy. This ultimately expanded the number of municipalities in which interviews were conducted to 328.

³ Households receive no compensation for interviews. When the results of the survey are published, the participants are sent a thank-you letter with copies of newspaper articles commenting on the survey.

Interviewers contacted 16,268 households, of which 43.9 per cent agreed to be interviewed (Table 3a).⁴ The sample was therefore composed of 7,147 households, of which 2,669 were panel households and 4,478 non-panel households. The participation rate was, as is normally the case, higher for panel households (66 per cent, compared with 36.6 per cent for non-panel households).

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

| Households: | Panel | | Non-panel | | Total | |
|----------------|--------------|--------------|---------------|--------------|---------------|--------------|
| | Number | Percentage | Number | Percentage | Number | Percentage |
| Interviewed | 2,669 | 66.0 | 4,478 | 36.6 | 7,147 | 43.9 |
| Refusals | 1,128 | 27.9 | 5,313 | 43.5 | 6,441 | 39.6 |
| Not contacted | 245 | 6.1 | 2,435 | 19.9 | 2,680 | 16.5 |
| Total | 4,042 | 100.0 | 12,226 | 100.0 | 16,268 | 100.0 |
| Ineligible (*) | 176 | 4.4 | 1,224 | 10.0 | 1,400 | 8.6 |

(*) Households not at the address listed in the municipal register (wrong addresses, deceased, moved).

The most common reason for non-participation was the unwillingness of the household (39.6 per cent; Table 3a). In 16.5 per cent of cases, the household could not be contacted by telephone or during the three visits paid by interviewers on different days and at different times.

The questionnaire

In this survey, the standard interview method based on a paper-based questionnaire (PAPI survey - *Paper and Pencil Personal Interviewing*) was modified to include computer-assisted interviews for about two thirds of the sample (CAPI -*Computer-Assisted Personal Interviewing*).⁵ Under the latter method, households provide responses to an electronic questionnaire, which is essentially a computer program that in addition to storing data also performs a number of checks on it, making it possible to remedy inconsistencies in the responses directly with the household.⁶ This ensures higher data quality, albeit at the price of the more complex programming of the questionnaire.

The interviews conducted with a paper-based questionnaires were subsequently transferred to electronic media by the survey company using the CAPI program as the input screen.

The questionnaire, which was based on that used in the previous survey, was subjected to the usual pretesting before the start of the general survey in order to reduce difficulties in comprehension and answering. The test survey, which involved about 100 households from around Italy, was conducted by pairs of interviewers. One conducted the interview proper,

⁴ The participation rate for the previous survey net of ineligible households (deaths, wrong addresses, change of residence) was 57 per cent.

⁵ A total of 4850 interviews (67.9 per cent) were conducted using the CAPI method. In the previous survey the CAPI method was tested with about 200 non-sample households.

⁶ There are many possible causes for such inconsistencies: the interviewee may not understand the question correctly, may recall certain information erroneously or may even be reluctant to provide information considered confidential. The most common mistakes by interviewers are coding errors or entering values in a different unit from that provided for in the questionnaire.

while the other noted difficulties on a special questionnaire.⁷ This exercise provided information that was helpful in reformulating certain questions more clearly.

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. In order to reduce the burden of answering, some sections were only administered to a random subset of the sample. In particular, households had to answer only one of the two sets of questions regarding working conditions and income expectations, depending on the year of birth (odd or even) of the head of household.⁸

Interviews lasted an average of 53 minutes, compared with 64 minutes in the previous survey. However, there was considerable variability within the sample, which was positively correlated with income, wealth and number of household members (Table 4a).

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 1998 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 1998.

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);

⁷ The strategy adopted was broadly that described in L. Oksenberg, C. Cannel and G. Kalton, “New Strategies for Pretesting Survey Questions” in *Journal of Official Statistics*, vol. 7, no. 1, Statistics Sweden, Stockholm, 1991, pp. 349-365.

⁸ In addition to producing estimates based on a smaller sample, this approach makes it impossible to make joint use of the responses to the two sections. In this instance, the relationship between the two aspects involved was felt to be of little interest.

- 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

1. Non-participation

Non-response can be a problem in statistical surveys since it may produce samples in which the less-cooperative segments of the population are underrepresented, thus generating biased estimates (*selection bias*). One indication of the extent of the phenomena is provided by the number of contacts needed to obtain an interview (Table 4a). In order to conduct the 7,147 interviews, interviewers made a total of 10,712 contact attempts, including 8,358 personal visits and 2,354 telephone calls (the latter were made solely to fix an appointment).⁹ The difficulty of obtaining an interview increased with income, wealth and the educational qualification of the head of household. It was less difficult to obtain interviews in smaller municipalities, with households of small size and where the head of household was retired or female. A number of measures were taken to limit the potentially distorting effects of failure to participate. First, households that could not be interviewed were replaced by others selected randomly in the same municipality. Second, at the end of the survey the sample was post stratified on the basis of certain individual characteristics of the interviewees, making it possible to reweight the various segments of the population within the sample. Studies of the data from the 1989 survey suggest that the bias of the estimates due to non-participation is small, thanks in part to the measures taken.¹⁰

2. Response reliability

An additional aspect that can influence the quality of estimates is the reluctance of households to report their sources of income or the real or financial assets they hold.¹¹ Although participation in the survey is voluntary and the content of the survey is known to the interviewee before the start, it is possible that respondents are not entirely truthful in their responses to the more “sensitive” questions, such as those regarding income or wealth. In order to assess the extent of such phenomena, which by their very nature are difficult to investigate, interviewers were asked to express a summary evaluation of the

⁹ A total of 15,138 contact attempts were made for the 10,521 households that were not interviewed.

¹⁰ 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.

¹¹ Moreover, it is not unreasonable to believe that certain sorts of liability might be deliberately understated by interviewees.

presumed reliability of the responses immediately following the interview, basing their judgement on the correspondence between the information provided and objective evidence available to them (zone and type of dwelling occupied by the household, standard of living implied by quality of furnishings, etc.).¹² As in the previous survey, although the reliability level was satisfactory on average, it was not homogeneous across the sample. The highest ratings were given to households with heads who were young, had a high educational qualification, were employees and resided in the North. Slightly lower ratings were given to households with heads who were elderly, had a low educational qualification, were self-employed or retired and resided in the South or the Islands. Reliability increased as the income and wealth reported in the survey increased (Table 4a).¹³ Additional elements used to assess the reliability of respondents can be obtained by comparing survey estimates with figures from the national accounts. Such comparisons must be made with caution, since at least part of the disparities found may be due to definitional differences.¹⁴ However, for the sources of income the recent adoption of the new system of national accounts (ESA95) has delayed the distribution of the necessary information by Istat, making comparisons impossible for the 1998 survey. A study of the surveys conducted up to 1995¹⁵ suggests that the survey understates income from interest and dividends and self-employment income more than income from transfers and salaried employment. By contrast, actual and imputed rents appear to be overstated.¹⁶ For real wealth, previous studies¹⁷ have indicated that the value of dwellings is understated by about 20 per cent. This appears to be due mainly to the failure to report second homes. Financial assets seem to be under-reported by a greater amount. Overall, the estimate that emerged from the 1998 survey was 22 per cent of the corresponding item in the financial accounts, although the latter also includes the assets of non-profit institutions. The underestimate is smaller for cash and bank or postal deposits, while that for shares, bonds and investment fund units is larger.¹⁸

¹² The interviewers' evaluations were expressed on a scale from 1 (completely unreliable) to 10 (completely reliable).

¹³ Obviously, the relationship between the level of reliability and "true" income is unknown.

¹⁴ The estimates derived from the survey were previously compared with those drawn from tax returns, which showed substantial correspondence for income from employment and a significant understatement of self-employment income declared in tax returns. For more on this issue, see L. Cannari, V. Ceriani and G. D'Alessio, "Il recupero degli imponibili sottratti a tassazione" in *Ricerche quantitative per la politica economica - 1995*, Banca d'Italia, Rome, 1997.

¹⁵ A. Brandolini, *The Distribution of Personal Income in Post-War Italy: Source description, Data Quality and the Time Pattern of Income Inequality*, Temi di Discussione del Servizio Studi, no. 350, Banca d'Italia, Rome, April 1999.

¹⁶ The percentage understatement varied from one survey to the next. On average, the survey estimates are about 70 per cent lower than the corresponding national accounts figure for interest income, 50 per cent lower for self-employment income and 20 per cent for income from salaried employment. Rental income is about 10 per cent higher.

¹⁷ L. Cannari and G. D'Alessio, "Housing Assets in the Bank of Italy's Survey of Household Income and Wealth", in Dagum and Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Springer Verlag, Berlin, 1990, pp. 326-334.

¹⁸ See L. Cannari and G. D'Alessio, "Non-Reporting and Under-Reporting Behavior in the Bank of Italy's Survey of Household Income and Wealth" in *Bulletin of the International Statistics Institute*, vol. LV, no. 3, Pavia, 1993, p. 395-412, and L. Cannari, G. D'Alessio, G. Raimondi and A.I. Rinaldi, *Le attività finanziarie delle famiglie italiane*, Temi di Discussione del Servizio Studi, no. 136, Banca d'Italia, July, 1990.

Table 4a
Number of contacts, average length of interview and reliability of responses
(number, minutes, score on scale of 1-10)

| Characteristics* | Phone contacts | Visits | Total contact attempts | Households | Contacts per 100 households | Average length of interview | Response reliability |
|--|----------------|--------|------------------------|------------|-----------------------------|-----------------------------|----------------------|
| Gender | | | | | | | |
| Male | 1,835 | 6,329 | 8,164 | 5,411 | 150.9 | 54.7 | 7.6 |
| female | 519 | 2,029 | 2,548 | 1,736 | 146.8 | 48.2 | 7.5 |
| Age | | | | | | | |
| up to 30 years | 66 | 381 | 447 | 318 | 140.6 | 49.8 | 8 |
| 31 to 40 | 416 | 1,464 | 1,880 | 1,218 | 154.4 | 53.5 | 7.9 |
| 41 to 50 | 584 | 1,871 | 2,455 | 1,582 | 155.2 | 55.6 | 7.8 |
| 51 to 65 | 774 | 2,641 | 3,415 | 2,259 | 151.2 | 55.5 | 7.5 |
| over 65 | 514 | 2,001 | 2,515 | 1,770 | 142.1 | 48.3 | 7.3 |
| Education | | | | | | | |
| none | 77 | 603 | 680 | 522 | 130.3 | 42.5 | 7.1 |
| elementary school | 535 | 2,268 | 2,803 | 1,964 | 142.7 | 50.2 | 7.4 |
| middle school | 759 | 2,656 | 3,415 | 2,270 | 150.4 | 54.1 | 7.6 |
| high school | 719 | 2,144 | 2,863 | 1,811 | 158.1 | 56.4 | 7.9 |
| university degree | 264 | 687 | 951 | 580 | 164 | 58.6 | 7.9 |
| Branch of activity | | | | | | | |
| agriculture | 39 | 165 | 204 | 145 | 140.7 | 53.5 | 7.4 |
| industry | 397 | 1,478 | 1,875 | 1,247 | 150.4 | 54.8 | 7.7 |
| public administration | 425 | 1,276 | 1,701 | 1,098 | 154.9 | 55.8 | 8 |
| other sector | 489 | 1,545 | 2,034 | 1,277 | 159.3 | 57.4 | 7.7 |
| Not employed | 1,004 | 3,894 | 4,898 | 3,380 | 144.9 | 50.1 | 7.4 |
| Work status | | | | | | | |
| Employee | | | | | | | |
| blue-collar worker | 347 | 1,339 | 1,686 | 1,148 | 146.9 | 53.5 | 7.8 |
| office worker or school teacher | 499 | 1,434 | 1,933 | 1,217 | 158.8 | 54.8 | 8.1 |
| cadre or manager | 152 | 411 | 563 | 352 | 159.9 | 59.5 | 8.1 |
| total | 998 | 3,184 | 4,182 | 2,717 | 153.9 | 54.9 | 8 |
| Self-employed | | | | | | | |
| sole proprietor, member of arts or professions | 167 | 558 | 725 | 454 | 159.7 | 59.5 | 7.6 |
| other self-employed | 185 | 722 | 907 | 596 | 152.2 | 57.9 | 7.1 |
| total | 352 | 1,280 | 1,632 | 1,050 | 155.4 | 58.6 | 7.3 |
| Not employed | | | | | | | |
| retired | 822 | 3,160 | 3,982 | 2,763 | 144.1 | 50.1 | 7.4 |
| other | 182 | 734 | 916 | 617 | 148.5 | 49.8 | 7.3 |
| total | 1,004 | 3,894 | 4,898 | 3,380 | 144.9 | 50.1 | 7.4 |
| Household size | | | | | | | |
| 1 member | 306 | 1,308 | 1,614 | 1,141 | 141.5 | 44 | 7.6 |
| 2 members | 576 | 2,055 | 2,631 | 1,783 | 147.6 | 51.4 | 7.5 |
| 3 members | 585 | 2,002 | 2,587 | 1,684 | 153.6 | 55.1 | 7.6 |
| 4 members | 614 | 2,117 | 2,731 | 1,798 | 151.9 | 56.7 | 7.7 |
| 5 members or more | 273 | 876 | 1,149 | 741 | 155.1 | 58.5 | 7.5 |
| Number of earners | | | | | | | |
| 1 earner | 854 | 3,440 | 4,294 | 2,966 | 144.8 | 48.6 | 7.5 |

| | | | | | | | |
|--------------------------|--------------|--------------|---------------|--------------|--------------|-------------|------------|
| 2 earners | 1,096 | 3,623 | 4,719 | 3,119 | 151.3 | 54.4 | 7.7 |
| 3 earners | 301 | 998 | 1,299 | 810 | 160.4 | 60.6 | 7.5 |
| 4 earners or more | 103 | 297 | 400 | 252 | 158.7 | 67.2 | 7.7 |
| Household income | | | | | | | |
| Up to €0.000 | 609 | 2,407 | 3,016 | 2,075 | 145.3 | 48.9 | 7.7 |
| €0.000 - €20.000 | 170 | 750 | 920 | 637 | 144.4 | 48.3 | 7.4 |
| €20.000 - €30.000 | 423 | 1,672 | 2,095 | 1,436 | 145.9 | 51.6 | 7.5 |
| €30.000 - €40.000 | 606 | 2,005 | 2,611 | 1,729 | 151 | 54.8 | 7.6 |
| More than €40.000 | 546 | 1,524 | 2,070 | 1,270 | 163 | 62.1 | 7.8 |
| Household income | | | | | | | |
| Up to €0.000 | 217 | 1,221 | 1,438 | 1,046 | 137.5 | 43.4 | 7.2 |
| €0.000 - €20.000 | 612 | 2,641 | 3,253 | 2,285 | 142.4 | 48.8 | 7.5 |
| €20.000 - €30.000 | 609 | 2,039 | 2,648 | 1,762 | 150.3 | 55.2 | 7.7 |
| €30.000 - €40.000 | 412 | 1,232 | 1,644 | 1,028 | 159.9 | 58.6 | 7.8 |
| More than €40.000 | 504 | 1,225 | 1,729 | 1,026 | 168.5 | 63.9 | 8 |
| Town size | | | | | | | |
| Up to 20,000 inhabitants | 444 | 2,228 | 2,672 | 1,908 | 140 | 50.7 | 7.5 |
| from 20,000 to 40,000 | 386 | 1,814 | 2,200 | 1,534 | 143.4 | 52.3 | 7.5 |
| from 40,000 to 500,000 | 1,121 | 3,329 | 4,450 | 2,864 | 155.4 | 54.3 | 7.6 |
| More than 500,000 | 403 | 987 | 1,390 | 841 | 165.3 | 56.4 | 7.9 |
| Geographical area | | | | | | | |
| North | 1,164 | 3,493 | 4,657 | 2,996 | 155.4 | 54.8 | 7.7 |
| Centre | 436 | 1,851 | 2,287 | 1,524 | 150.1 | 56.7 | 7.6 |
| South and Islands | 754 | 3,014 | 3,768 | 2,627 | 143.4 | 49.2 | 7.5 |
| Total | 2,354 | 8,358 | 10,712 | 7,147 | 149.9 | 53.2 | 7.6 |

(*) Referred to the head of household

Checking data and imputing missing data

The CAPI survey method sharply reduced the need for post-survey consistency checks of data quality. However, the standard checking procedure was used for the interviews conducted with the paper-based questionnaire (about one third), for which the CAPI program was used as an input screen in order to exploit its ability to flag inconsistencies. In these cases, problems were solved through telephone contacts with the households involved.

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.¹⁹ 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).

The amount of imputed data is generally small, on the order of a few dozen cases for most variables. For more complex questions that require the respondent to estimate amounts, such as fringe benefits for salaried workers, depreciation for the self-employed, the value of dwellings or business equity, imputed rents, other property and furnishings, between 5 and 10 per cent of the data must be imputed.

¹⁹ Nevertheless, while not answering was possible for some questions, the failure to indicate sources of income or the most significant components of wealth resulted in the invalidation of the interview.

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:

$$(1) w_{hi} = \begin{cases} \frac{P_h P_{hi}}{\tilde{P}_h n_{hi}} \\ \frac{1}{m_h} \frac{P_h}{n_{hi}} \end{cases}$$

respectively for municipalities with more than 40,000 inhabitants and for municipalities with up to 40,000 inhabitants, where P_h , \tilde{P}_h 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.²⁰

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 1995, 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

²⁰ The probability of a household being extracted in a selected municipality is approximately equal to n_{hi} / P_{hi} . For municipalities with more than 40,000 inhabitants, which are all included in the theoretical sample, we need to bear in mind that for organizational purposes it is not always possible to conduct interviews in all the municipalities in the stratum. The first term of equation (1) therefore allows us to take account of this circumstance. Municipalities with fewer than 40,000 inhabitants are selected with a probability proportional to their size (PPS). The selection probability of the i^{th} municipality in the h^{th} stratum is therefore equal to $m_h * P_{hi} / P_h$. The probability of a household being included in the sample can therefore be written as $m_h * n_{hi} / P_h$.

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

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

However, if the values of variable x measured on two successive waves are correlated, an optimal estimator of the mean is given by:²¹

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

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

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, \mathbf{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:

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

and the last term of (3) disappears. In addition, given that the correlation coefficients for the main variables examined are between 0.4 and 0.6, giving \mathbf{r} the intermediate value $\tilde{\mathbf{r}} = 0.5$, it is possible to approximate the estimator (3) by way of:

$$(6) \quad \bar{x}_t^+ = \tilde{\mathbf{a}}\bar{x}_t^q + (q-\tilde{\mathbf{a}})\bar{x}_t^p \quad \text{with } \tilde{\mathbf{a}} = \frac{Q(1-\tilde{\mathbf{r}}^2Q)}{1-\tilde{\mathbf{r}}^2Q^2}$$

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

²¹ See L. Fabbri, "L'indagine campionaria", La Nuova Italia Scientifica, Rome, 1989.

$$(7) \quad w_{hij}^* = \begin{cases} w_{hij} \frac{1-\tilde{a}}{1-Q} \\ w_{hij} \frac{\tilde{a}}{Q} \end{cases} \quad \text{respectively for panel households and for non-panel}$$

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 (41.4 per cent compared with 37.3 per cent), with a corresponding reduction in the weight assigned to the non-panel households. Since this reweighting 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 cannot easily be determined with the usual analytical methods. The presence of stages b) and c) as described in the previous section render useless - except with a large margin of inaccuracy - the equations for calculating standard errors of the means in a two-stage sampling with stratification of the first-stage units. For this reason, the standard errors were calculated using simulation methods that take account of the original design of the sample and subsequent adjustments. In particular, 200 bootstrap samples of equivalent size to the actual sample were replicated (drawing the units with replacement in both stages). The mean values of the main variables were obtained by performing the full estimation process. The variability of the estimators was approximated analysing the distribution of simulated mean values.

The standard errors of the means of the main variables are shown in Table 8a. The table reveals the limited variability of the means for the demographic variables, which is mainly attributable to the post-stratification carried out in stage c. As regards the main economic variables, it can be noted that the standard errors in the means for consumption and income are significantly smaller than the standard error for net wealth. The standard errors in estimates at the level of geographical area are naturally larger than those for the sample as a whole.

Table 8a
Standard errors in the estimation of the means for the main variables
(units, euros, percentages)

| Variable | North | | Centre | | South and Islands | | Total sample | |
|------------------------|--------|---------------|--------|---------------|-------------------|---------------|--------------|---------------|
| | Value | % of estimate | Value | % of estimate | Value | % of estimate | Value | % of estimate |
| Mean number of members | 0.034 | 1.3 | 0.065 | 2.3 | 0.038 | 1.2 | 0.020 | 0.7 |
| Mean age | 0.40 | 0.7 | 0.79 | 1.5 | 0.46 | 0.8 | 0.17 | 0.3 |
| Household income | 1,494 | 2.7 | 2,254 | 4.2 | 837 | 2.4 | 746 | 1.5 |
| Household consumption | 775 | 2.1 | 1,736 | 4.5 | 742 | 2.8 | 502 | 1.5 |
| Net wealth | 11,219 | 4.0 | 29,478 | 9.7 | 10,194 | 5.7 | 8,346 | 3.3 |

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 "Italian Household Budgets in 1998", *Supplements to the Statistical Bulletin – Methodological Notes and Statistical Information*, Bank of Italy,

Year X, No. 22, April 2000, the number of individuals living in low-income households (those with equivalent incomes of less than half the median income, whereby the OECD scale of equivalence was used) is equal to 14.2% of the total. The Gini coefficient of concentration is 0.376 for the distribution of household income and 0.343 for the distribution of equivalent income.