BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

del Servizio Studi

The weighting process in the SHIW

by Ivan Faiella and Romina Gambacorta

Number 636 - June 2007

THE WEIGHTING PROCESS IN THE SHIW

by Ivan Faiella* and Romina Gambacorta*

Abstract

The design of a probability sample jointly determines the method used to select sampling units from the population and the estimator of the population parameter. If the sampling fraction is constant for all the units in the sample, then the unweighted sampling mean is an unbiased estimator. In the Survey of Household Income and Wealth (SHIW), units included in the sample have unequal probabilities of selection and each observation is weighted using the inverse of the proper sampling fraction (design weight) adjusted for the response mechanism (non-response weight) and for other factors such as imperfect coverage. In this paper we present the weighting scheme of the SHIW and assess its impact on bias and variance of selected estimators. Empirical evidence shows that the increasing variability caused by the use of weighted estimators is compensated by the bias reduction even when performing analysis on sample domains. A set of longitudinal weights is also proposed to account for the selection process and the attrition of the SHIW panel component. These weights, given their enhanced description of the "panel population", should be better suited to perform longitudinal analysis; nevertheless, their greater variance implies that they are not always preferable in terms of mean square error.

JEL Classification: C42.
Keywords: survey methods.

Contents

1. Introduction....................3
2. The rationale of weighting sample survey units....................4
 2.1 The role of sampling weights....................4
 2.2 The different weighting stages....................5
3. The weighting process in the Survey of Household Income and Wealth (SHIW)....................8
 3.1 A brief description of the SHIW....................8
 3.2 SHIW weights....................9
 3.3 Impact of weighting on SHIW estimates....................13
4. A set of longitudinal weights for the SHIW....................14
 4.1 A proposal to build a set of longitudinal weights in the SHIW....................15
 4.2 The validation of longitudinal weights on SHIW estimates....................19
5. Conclusions....................20
References....................21
Appendix A - Statistical tables....................24

* Bank of Italy, Economic Research Department - Sample Surveys and Statistical Methods Unit.

1. Introduction[1]

The design of a probability sample jointly determines the method used to select the sampling units from the population and the estimator employed (*sampling strategy*). If, according to the selection process, all the elements of the population have an equal chance to be included in the sample we have an *epsem* (equal probability selection method) design.

If the design is *epsem* and the sample size is fixed, given the sampling scheme, then the unweighted sampling mean is an unbiased estimator of the population mean. In practice, the sample units are often selected with unequal probabilities. This can be the consequence of a disproportional allocation of the sample to the strata, of the oversampling of some sub-classes, or it can reflect the result of response process or practical constraints (Verma, 2000). In a *design-based* perspective,[2] in order to obtain an unbiased estimator each observation should be weighted using the inverse of the proper sampling fraction (*design weight*) adjusted for the response mechanism (*non-response weight*) and for other factors (such as imperfect coverage, alignment to some features of the population, etc.).

The final weight is approximately proportional to the inverse of the selection probability of each sample unit. This weight is the key ingredient of an unbiased estimator of the total in the population, the *Horvitz-Thompson* estimator (Kish, 1965) also known as the π estimator (Särndal et al., 1992). The unbiasedness of the π estimator is not for free: there is an increase in variance due to the use of weights that is related to the squared coefficient of variation of the weights (Kish, 1992).

In the present study we describe the weighting process of the Survey of Household Income and Wealth (SHIW) and to assess its effects on the estimates of some key variables and their variability.

The study is structured as follows Section 2 is devoted to describing the rationale of weighting; in Section 3 the weighting process in the SHIW is illustrated and its effects on the estimates and their variability are appraised. Section 3 refers to the cross-sectional weights, currently disseminated with the micro-data, while Section 4 proposes a method to obtain a new set of weights (longitudinal weights) explicitly designed to conduct analyses on the panel households. Finally, in Section 5 the main conclusions are drawn.

[1] We wish to thank Giovanni D'Alessio, Leandro D'Aurizio, Stefano Iezzi, Federico Signorini, Raffaele Tartaglia and two anonymous referees for their suggestions and comments. The views expressed are those of the authors and do not necessarily reflect those of the Bank of Italy. E-mail: ivan.faiella@bancaditalia.it, romina.gambacorta@bancaditalia.it.

[2] In this framework a randomised selection mechanism (the sampling scheme) is assumed and the sample is a realisation of this procedure (Särndal et al., 1992).

2. The rationale of weighting sample survey units

2.1 The role of sampling weights[3]

Kish (1992) identifies several reasons to use weights in analysing sample survey data. Some are a consequence of the selection procedure such as the *sampling design*, for example when stratification with disproportional allocation is used, the *non-response* process and the correction of *frame imperfections*. Others depend on the estimators used in the analysis, as in the case of post-stratification, generalised regression estimators or other methods included in the more general class of *calibrated estimators*.

On the other hand, using weights introduces a further complication in the analysis and can increases the variance of the estimators in proportion to the variability of the weights.

The analyst is then faced with a trade-off: neglecting weights can have a considerable effect on the statistical analysis because the unweighted estimators are biased under the sampling design. On the other hand, using weights will increase the sampling variance. A priori, the effect on the mean square error is unclear.

Various studies, using simulated as well as real data, point out that when using weights with large samples the bias reduction usually outweighs the increase in variance (Hansen et al., 1983). For instance, Purdon and Pickering (2001), analysing the Workplace Employee Relations Survey, assess whether the increase in bias associated with unweighted estimates is compensated for by the decrease in the standard error, by estimating the mean square error (i.e. the square of the bias plus the design based variance). With the exception of one variable, all the figures are smaller for the weighted estimates than for the unweighted ones.

To implement unbiased (i.e. weighted) estimators limiting the impact of weighting on variance it is possible to limit the influence of extreme weights or to exploit the properties of *calibrated estimators* (Särndal and Lundström, 2005).

In the following section we explore the several stages that constitute the weighting process, within the *design based* approach to sampling theory.

---

[3] An important topic, not discussed in the present study, concerns the long-standing debate on the use of weights in regression analysis (see for example Lohr, 1999 and Deaton, 1997). According to a model-based approach, if the model is correctly specified we can avoid using weights. Särndal et al. (1992) criticize a purely model-based approach, where design unbiasedness is neglected, underlining that the parameters estimated using sampling weights are more robust because they are model unbiased if the model is true and design consistent if it is not (on this point see also Little, 1989).

2.2 The different weighting stages

As pointed out by Groves et al. (2004) the final weight is the product of different stages. In what follows we focus on three stages of the weighting process. The first one reflects the selection process implied by the sampling scheme (design weight – $w^{(0)}$); the second integrates the correction for participation in the survey (non-response adjusted weight - $w^{(1)}$); the third exploits the available auxiliary information vector (at population or sample level) and incorporates it in the final weight (calibrated weight - $w^{(2)}$)

The design weights

In presence of a finite population of N elements $U=\{1,2,\ldots,N\}$ suppose that a list covering the whole population U is available.[4] From this list a probability sample, $s=\{1,2,\ldots,k,\ldots,n\}\subseteq U$, of n elements is drawn with a probability *p(s)* according to a given sampling plan. Once this *sampling design* is determined, the inclusion probability of the element k is the sum of the sample probabilities over all possible samples that include k, i.e. $\pi_k = \sum_{k \in s} p(s)$.

For the *k-th* element we define the inverse of this probability as the *design weight*: $w_k^{(0)} = \dfrac{1}{\pi_k}$.

For each element in the sample a given character y_k is measured. The total of y in the population is $Y = \sum_{k=1}^{U} y_k$ and an unbiased estimator of this total is given by $\hat{Y} = \sum_{k=1}^{n} w_k^{(0)} y_k$, also known as the *Horwitz-Thompson estimator* (HT) of the population total. The idea behind the HT estimator is that each *k-th* unit in the sample "expands" the information collected to $w_k^{(0)}$ units in the population.

The non-response adjusted weights

Once the sample is drawn, during field operations the sample elements are contacted for the interview. Some of them will participate in the survey while others will not. Some of the non-responding units are not found at home (non-contact) while others explicitly decline to participate (refusals).[5]

[4] In presence of coverage imperfections (i.e. the list used to draw the sample is not covering correctly the target population) the design weight can be corrected appropriately (for example inflating it in presence of under-coverage and deflating vice versa).

[5] A share of the sample elements is not eligible, i.e. it cannot be contacted (e.g. a bankrupted firm or a household that has moved abroad).

This response process can be viewed as a further stage of selection.[6] If the group of non-respondents is a fixed group of the population that does not participate in sample surveys, assessing the impact of non-response on the estimates is a daunting task as non-respondents can never be observed, so that this further stage of sampling is not measurable (these “hardcore non-respondents” have zero probability of being observed in sample surveys). Alternatively, it can be supposed that each element of the population *U* has a given (non-zero) propensity to participate in sample surveys. In this circumstance, the actual (unknown) response process can be evaluated through the estimated probability of responding of each household $\hat{\theta}_k$ that can be used to extend the rationale of the HT estimator, $\hat{Y}_{NR} = \sum_{k=1}^{n} w_k^{(1)} y_k$ where $w_k^{(1)} = w_k^{(0)} \hat{\theta}_k^{-1}$. In this framework, the information collected for each sampling unit is additionally expanded according to its (estimated) response propensity. The estimated response probabilities are influenced by the selected sample *s* (the “first stage”) and by the hypothetical response process *r* (i.e. $\hat{\theta}_k = E\left(\theta_{s,r}\right)$).

In practice, the response probabilities can be estimated by grouping the units into strata (response homogeneity groups – RHGs) or by applying models in order to derive a probability controlling for household characteristics. With both methods, the non-response-adjusted HT estimator is unbiased only if the grouping (or the model) appropriately accounts for non-response behaviour. Furthermore, the variance of this estimator will be higher the lower is the propensity to participate[7] and the more detailed is the method of modelling this propensity.

[6] Oh and Scheuren (1983) termed this approach a “quasi-randomisation” framework, given that the true response function is not known.

[7] In particular, cells or individuals with very low probability to participate will have a large non-response weight, with the risk of an unduly strong effect on the variability of the response-adjusted estimator

The calibrated weights

After the survey has been conducted it is possible to exploit additional information about the population - possibly not available at the design stage - coming from larger surveys, census data or other sources believed to be more reliable than survey estimates. This information covers some population characteristics and is employed in order to improve the precision of survey estimates in terms of both reducing bias and increasing efficiency. In all major sample surveys, a final weighting step involves some form of post-stratification (Groves et al., 2004).[8] In this case the HT estimator can be further improved to incorporate external information: $\hat{Y}_{PS} = \sum_{k=1}^{n} w_k^{(2)} y_k$ where $w_k^{(2)} = \delta_k w_k^{(1)}$ and δ_k is constructed in order to satisfy the following constraint: $\sum_{k \in j} \delta_k w_k^{(1)} x_k = X_j$ for each of the *j=1…J* post-strata defined from the crossing of all the auxiliary variables (known as calibration equations).[9] In other words, the weighted estimated total of some population characteristics should be equal to the one coming from the external source (X_j).[10]

In a more general attempt to include auxiliary information, all classical sampling textbooks suggest the use of the regression estimator as a tool to increase the precision of sample estimates (Kish, 1965, Chapter 12.3; Cochran, 1977, Chapter 7). More recent textbooks (in particular Särndal et al., 1992 and Särndal and Lundström, 2005) extend this concept, focussing explicitly on *model-assisted* estimation, i.e. using auxiliary information to adjust the sampling weights to obtain more efficient estimators. An important family of calibration estimators is represented by the *Generalised Regression Estimator* (GREG).[11] $\hat{Y}_{GREG} = \hat{Y}_{HT} + \left(\sum_{k=1}^{N} x_k - \sum_{k=1}^{n} x_k \right)' * B_{s,w^{(0)}}$. The idea behind the GREG is to supplement the HT estimator using a set of auxiliary information (in a multivariate context) correlated to the study variable. Practically, this implies estimating on sample data a vector of regression coefficients $B_{s,w^{(0)}}$ (i.e. conditioned on the sample *s* and estimated using for each observation the design

[8] For an extensive review of the different uses of post-stratification weights in survey estimates see Smith (1991).

[9] Post-stratification is often employed to correct for marginal count in the sample with respect to the count in the population. In this case, the relevant estimator is known as the *raking-ratio estimator* (Kalton and Flores Cervantes, 2003).

[10] Another use of external survey data is to form weighting classes for non-response which allow for a greater homogeneity with respect to response probability than the ones it is possible to build using only survey data (Bethlehem, 2002). This aim can also be achieved through the use of post-stratification models (Gelman and Carlin, 2002).

[11] Deville and Särndal (1992) and Särndal and Lundström (2005) show that the *calibration estimators* are a version of the GREG estimator imposing some constraints on how the original design weights are modified. From this perspective, the calibrated weights are preferable to GREG weights (also known as *g-weights*) because the former can be better controlled in order to limit their deviation from the original weight.

sampling weights $w^{(0)}$) that links the matrix of the auxiliary variables $\boldsymbol{X}$ to the study variable $\boldsymbol{Y}$. The HT estimator is then corrected using the "gap" between sample estimate and information on the value of the auxiliary vector in the population (available from larger surveys or census data).

In general, the aim of calibration is to increase precision as it reduces differences between the sample and population distributions with respect to some auxiliary variables.

3. The weighting process in the Survey of Household Income and Wealth (SHIW)

3.1 A brief description of the SHIW[12]

The SHIW has been conducted by the Bank of Italy since 1965 to collect information on Italian households' economic behaviour, with a focus on the measurement of income and wealth components. The main objective of the SHIW is to obtain estimates of how income and wealth are distributed across Italian households.

The basic statistical unit is the household, defined as a group of individuals linked by ties of blood, marriage or affection, sharing the same dwelling and pooling all or part of their incomes. Institutional population is not included. The overall size of the sample is about 8,000 households. Data are collected by means of personal interviews conducted by professionally-trained interviewers and using computer assisted devices (Computer Assisted Personal Interviewing).

Data collection is entrusted to a specialised company and the interview stage is preceded by a series of meetings at which officials from the Bank of Italy and representatives of the company give 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.

The core sections of the questionnaire remain basically unchanged. Two monographic topics are added in each wave. In order to reduce the response burden, these sections are only administered to a random subset of the sample.

The sample is drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size. Within each stratum, municipalities are chosen by including all municipalities with a population of more than 40,000 (self representing units - SRUs) and randomly selecting smaller towns with probability proportional to the

[12] Further details regarding the SHIW are given in Bank of Italy (2006).

resident population (non-self-representing units - NSRUs). Within each selected municipality, the individual households are then selected randomly.

Until 1987 the survey was conducted with time-independent samples (cross-sections) of households. In order to facilitate the analysis of changes in the phenomena being investigated, since 1989 part of the sample has comprised households interviewed in previous surveys (panel households). This design is known as a *split panel survey* (Kish, 1987) and has the advantage of being flexible in providing both cross-sectional and longitudinal measures (Duncan and Kalton, 1987).

Microdata, documentation and publications (in Italian and English) are freely available at www.bancaditalia.it/statistiche/ibf.

3.2 SHIW weights

SHIW weights are constructed in three different steps. For each stratum h[13] the design weight is equal to the inverse of the selection probability of a unit in that stratum (i.e. the ratio between the number of households to be selected in the stratum and its total population).[14] This weight is further adjusted to account for imperfect coverage:[15]

(1) $$w_h^{(0)} = \frac{N_h^{'}}{n'_h}\frac{N_h}{N_h^{'}} = \frac{N_h}{n'_h}$$ **design weights** (corrected for imperfect coverage),

where N_h and N'_h are respectively the total resident population and the population included in the sampled municipalities in the h^{th} stratum and n'_h is the number of households to be selected in the h^{th} stratum.

These weights are then corrected for non-response, as some of the selected families cannot be found at home or refuse to be interviewed:[16]

---

[13] Strata are cross-determined by the geographical region and by a 3-class variable linked to the demographic size of the municipality. Strata with limited sample size are collapsed.

[14] Given the cluster design, in each stratum the probability of selecting a household is the product of the probability of selecting the municipality and the probability of selecting the resident household. All the municipalities with more than 40,000 inhabitants are included in the sample so that the first term is 1 (S*elf Representing Units - SRU*). The remaining municipalities are selected according to a PPS scheme. In order to have a clearer description of the different weighting stages, in the text we neglect the difference between the SRU and the PPS selected municipalities. The detailed formula can be found in Bank of Italy (2006).

[15] E.g. under-coverage generated because not all municipalities in the stratum are included in the sampling frame, which can be also seen as a non-response at PSU level.

[16] This weight is based on the assumption that each stratum constitutes a RHG. This assumption is principally due to the lack of further information about the households not interviewed.

(2) $w_h^{(1)} = w_h^{(0)} \frac{n'_h}{n_h} \frac{n_h}{m'_h} = \frac{N_h}{m'_h}$ **non-response weights** (corrected for eligibility[17]),

where n_h and m_h are respectively the number of eligible and respondent households in the h^{th} stratum ($n'_h \geq n_h \geq m_h$).[18]

Finally, weights are calibrated to account for additional information coming from the panel units and from external surveys.

Panel households are firstly post-stratified in order to adjust for the attrition in the panel.[19] According to 12 frequency cells, determined by the cross-classification of 4 income classes and 3 work status of the head of household, weights of panel families are aligned to the distribution of the whole sample measured in the previous wave:

(3) $w_c^{(2')} = w_c^{(1)} \alpha_c$ **attrition-adjusted weights** (only panel units),

where α_c is the post-stratification factor for each of the c panel post-stratification classes.

Secondly, the weights are corrected in order to gain from the positive correlation between the data gathered from the same households in successive surveys. The panel segment is then re-weighted so that its share of the total sample allows the gain in precision to be maximised: the optimum share of the panel depends on an estimate of the auto-correlation between the main survey variables (income and wealth):

(4a) $w_p^{(2'')} = w_p^{(2')} \beta_p$ **autocorrelation corrected weights** (p=panel indicator),

where β_p is the autocorrelation correction coefficient for panel and non-panel units.

[17] Cases not eligible for in-person household surveys include: out-of-sample housing units; not-a-housing unit; vacant housing units; housing units with no eligible respondent. See the standard definitions on the website of the American Association for Public Opinion Research (AAPOR), www.aapor.org.

[18] Table A1 in the Appendix contains the main factors for the construction of design and non-response weights for each stratum.

[19] For *attrition* here we refer to the sample-reducing process caused by both non-response and ineligibility (death, moving, etc). See Fitzgerald et al. (1998).

In detail, if there is a positive correlation between the variables observed on the panel in two consecutive surveys, this can be used to obtain more efficient estimators. When the values of variable y gathered in consecutive surveys are correlated, an optimal estimator of the mean is given by:[20]

(4b) $$\bar{y}_t^* = \frac{Q(1-Q\rho^2)}{1-Q^2\rho^2}\bar{y}_t^q + \frac{P}{1-Q^2\rho^2}\left[\bar{y}_t^p + \rho(\bar{y}_{t-1} - \bar{y}_{t-1}^p)\right];$$

setting $w_p^{(2")} = \frac{Q\,(1-\rho^2 Q)}{1-\rho^2 Q^2}$ we have

(4c) $$\bar{y}_t^* = w_p^{(2")}\bar{y}_t^q + \left(1-w_p^{(2")}\right)\bar{y}_t^p + \left(1-w_p^{(2")}\right)\rho\;(\bar{y}_{t-1} - \bar{y}_{t-1}^p)\;,$$

where $\bar{y}_t$ and $\bar{y}_{t-1}$ are respectively the means of variable y at time t and time t-1, $\bar{y}_t^p$ and $\bar{y}_t^q$ are the means of variable y at time t for the panel and non-panel parts of the sample respectively, and ρ is the correlation coefficient between $\bar{y}_t$ and $\bar{y}_{t-1}$, and Q is the share of non-panel households. The estimator (4c) can be regarded as a *composite estimator* equal to the weighted average of two adjusted estimators: the first uses the information on y_t available for the sample of non-panel households; the second is based both on the data on y_t for the panel households and on the changes between the two surveys, adjusted using a regression estimator to take account of the difference between the total sample and the panel part of the sample. The two estimators are weighted in inverse proportion to their contribution to the overall variance of the combined estimator.[21]

Lastly, the final weights are modified to reproduce the same characteristics as the population with regard to sex, age group, geographical area and size of municipality of residence:

(5) $w_j^{(2"')} = w_j^{(2")}\gamma_j$ **post-stratification weights,**

where γ_j is the post-stratification factor for each of the j^{th} post-stratification classes.[22]

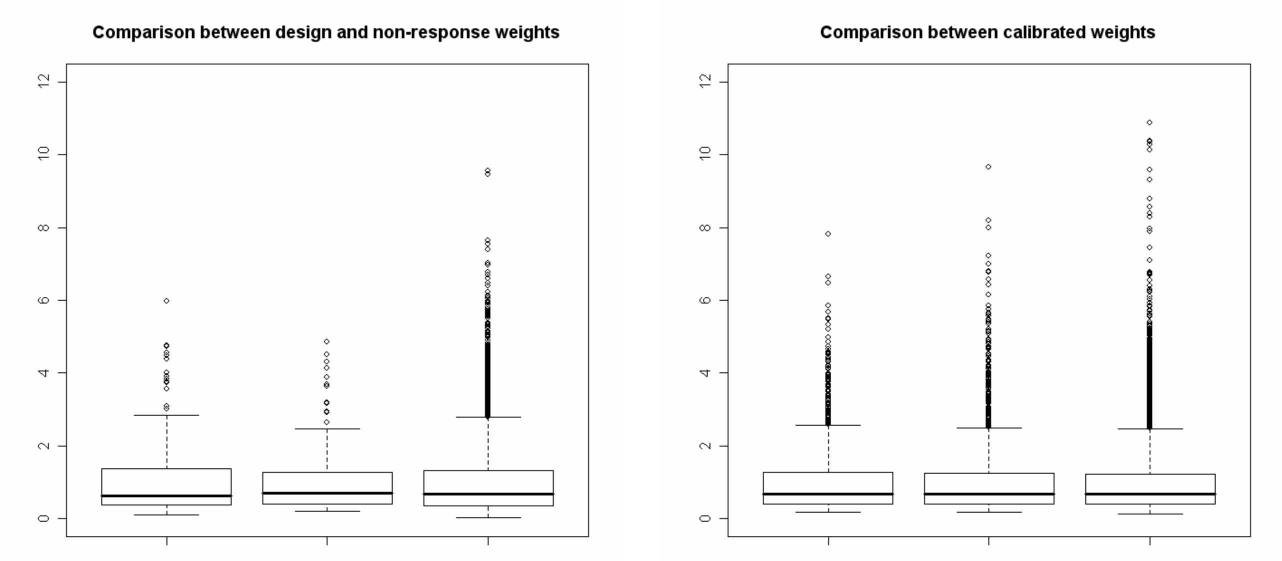
[20] The part of estimator (3) in square brackets is the estimator of the mean of the panel sample only, adjusted using a regression estimator that expands the relation between $\bar{y}_t^p$ and $\bar{y}_{t-1}^p$ to the whole of the sample. The correlation coefficient ρ is used in place of the bivariate regression coefficient on the assumption that the variations in y are constant over two consecutive surveys. See L. Kish (1965), Chapter 12.

[21] Composite estimators are used in the literature on small-area estimation to combine direct and indirect estimates, thus minimising the mean square error. For an application of these estimators with repeated measurements over a period of time, see Chapter 9 of Särndal et al. (1992). The rationale of the procedure is reported in Kish (1965) Chapter 12.4.

[22] Post-stratification uses a *raking-ratio* procedure to align weights to the population marginal distributions with respect to sex, age (4 classes), geographical area (3 classes) and municipality size (4 classes).

Fig.1

Distributions of the weights at various stages of the weighting process
(*box-plots of normalised weights*)



	Design weight	Non-response weight (RHGs)	Non-response weight (logit model)
Coeff. of Variation	93.2	87.4	96.2

	Attrition-adjusted weight	Autocorrelation-adjusted weight	Post-stratified weight
Coeff. of Variation	88.7	90.6	96.2

The impact of the different stages of the weighting process on the variability of weights is shown in Figure 1 (where weights are re-scaled to their mean to sum-up to sample size).

Looking at the first pane of Figure 1, it is apparent that using the RHG method to adjust for non-response decreases the variability of the weights. By contrast, the weights corrected modelling a non-response function (in particular, we applied to 2004 data the method presented in D'Alessio and Faiella, 2002 and extensively used in Brandolini et al., 2004) show a higher degree of variability, with weights that can be 6 to 10 times the average. Therefore, we subsequently use a response adjustment based on the RHGs technique.

The second pane of Figure 1, illustrates the effect of the different calibration phases (applied sequentially) on the distribution of SHIW weights. In the end, the final weights are more dispersed. An option to limit the variance of the weight is to trim extreme weights in some way.[23] This procedure, as pointed out in Potter (1990), has the advantage of reducing variance considerably, but it always involves the risk of introducing bias in the weighted estimator.[24]

[23] As illustrated by Kalton and Flores Cervantes (2003), this is actually equivalent to collapsing cells in the case of cell adjustment.

[24] Chantala (2001) presents a procedure to find a set of adjusted weights that minimizes both the variance and the bias in estimates based on ranking the set of trimmed weights according to their impact on bias and variance.

3.3 Impact of weighting on SHIW estimates

The bias reduction property of weighting produces an efficiency loss (L_w), defined as the percentage increase in the variance due to weighting; in the hypothesis of independent sampling and constant variance (Kish 1965, 1992):

$$(6) \qquad V_w = V_{unw} * \left(1 + cv_w^2\right), \text{ then } L_w = \frac{V_w - V_{unw}}{V_{unw}} = cv_w^2$$

where V_{unw} and V_w are, respectively, the variances of the unweighted and of the weighted estimators and cv_w are the coefficient of variation of the survey weights. With a two-stage stratified cluster sample as in the SHIW, this relationship is only an approximation, but can nevertheless be used as an empirical method to assess the contribution of the weights to the variance of the estimators employed. Using the data in the bottom part of Figure 1, according to this formula weighting will increase the variance of unweighted SHIW estimates by between 76 and 93 per cent. But this refers to the variability of the weights and not to the variance of the weighted estimator. In fact, if the outcome of interest is related to the weighting cells, the overall variability is reduced (Little and Vartivarian, 2005).

A more complete approach to assessing the effect of weighting is based on the contemporaneous analysis of both the gain in terms of bias reduction and the loss due to the increased variance of the weighted estimator. In order to do this, we evaluate the overall effect of weights on an estimator by considering its mean square error (MSE).[25] Given that the weighted estimator is unbiased, we assume that its MSE is equal to its variance. The estimated MSE of the unweighted estimator is the sum of an unbiased estimate of the variance and an unbiased estimate of the squared bias. Following Little et al. (1997), we obtain an empirical counterpart of the MSE, necessary to evaluate different weighting options and their impact on variance and bias of the estimates,[26] using the following formula:

$$(7) \qquad M\hat{S}E_{unw} = \hat{V}_{unw} + \max\{\hat{B}^2 - \hat{V}_{\hat{\theta}_{unw} - \hat{\theta}_w}, 0\},$$

thus correcting the measure of the squared bias with the variance of the difference between the unweighted and the weighted estimator[27] and avoiding negative values.

---

[25] The MSE of an estimator – the sum of the variance and the square of the bias – reflects both the variation about the average and the bias of an estimator.

[26] Both the bias and the variance are characteristics of an estimator and not of one of its possible realisations (i.e. the estimate). In what follows we adhere to the interpretation of Särndal et al. (1992), p.41, indicating for *biased estimate* “an estimate calculated from an estimator that is biased”.

[27] As emphasised by Little et al. (1997) this measure of the MSE corrects for the possible “overestimation” of the empirical measure of the squared bias. In fact, $E\left(\hat{\theta}_{unw} - \hat{\theta}_w\right)^2 = \hat{B}^2 + \hat{V}_{\hat{\theta}_{unw} - \hat{\theta}_w}$, where the last term is the variance of the

To obtain a measure of the variance of the weighted estimator of a collection of variables available in the SHIW and of equation 7, we use a set of 325 *jackknife* replicates determined according to the sampling design and applied to the 6 possible weighting schemes (no weights, design weights, non-response weights and the set of three calibrated weights).[28]

The results in Table A2 show that for all the variables considered the MSE of the weighted estimates are by and large a fraction of the MSE of the unweighted estimator. The gain of weighting is lower for variables with limited range of variability, but is evident for the key variables of the survey (income, expenditure and wealth).

This result holds even when we consider domain estimates. In fact, in the presence of a large sample sizes, the MSE may be dominated by the bias term, while when sample sizes are small, the variance may be a greater cause for concern. To assess this potential drawback of weighting in the SHIW we compare the empirical MSE computed using both the non-response adjusted weight and the final SHIW weight with that of the unweighted estimator for a set of 7 domains of study routinely used for data analysis. Results in Table A3 show that the MSE of the weighted estimator is seldom above the MSE of the unweighted.

We can then finish this section relying on empirical evidence that the increasing variability induced by using weighted estimators is compensated by the bias reduction even when performing analysis on a sample subset (domain).[29] Nevertheless we suggest that domain analysis include all the auxiliary information regarding domain characteristics (e.g. using post-stratification) in order to maximise the bias reduction function of weighting.

4. A set of longitudinal weights for the SHIW

In the previous paragraphs we dealt with cross-sectional weights. But, as we have seen, almost half of the SHIW sample is composed of panel households that can be used for longitudinal analysis. In this

difference between the unweighted and the weighted estimator. In our computations $\hat{\theta}_w$ is the estimate using the final calibrated weight ($w^{(2''')}$)

[28] The choice of *jackknife* repeated replications (JRR) as a tool to obtain variance estimates for the SHIW is explored in detail in Faiella (2007) and documented in Bank of Italy (2006), appendix A.

[29] As a rule of thumb, the standard errors for the estimates of the domains can be approximated by $Stderr_g = Stderr * \sqrt{n} \big/ \sqrt{n_g}$, where *Stderr* is the standard error of the estimate of the whole sample *n* and $Stderr_g$ is the standard error of the estimate of the g[th] subset of n_g units (Kish, 1965).

case, cross-sectional weights present some limits.[30] Firstly, when using only panel data we are referring only to a subset of the sample, possibly not representative of the entire population;[31] secondly, selection process and response patterns are different for panel units with respect to households interviewed just on one occasion.[32] Furthermore, when dealing with longitudinal analysis we have to account for the variation of the target population that arises between one survey and another. For example, when considering the transition of households' income between two surveys, in the construction of the weights it is not correct to refer either to the population at the starting period or to the one at the end. It is instead necessary to account for the fact that during the period studied some households exit while others enter the population. In order to account for all these factors we propose the construction of a set of longitudinal weights.

4.1 A proposal to build a set of longitudinal weights in the SHIW

In this paragraph we build a new set of weights in order to obtain survey estimates for the panel population.[33] In fact, when dealing with panel data, the population we refer to is represented by a dynamic concept as we have to consider both "deaths" and "births" of families. In practice, longitudinal weights have been built as follows. At time t_{i1}, which is the first time household i has been interviewed, its longitudinal weight is equal to the cross-sectional weight (which is the final weight that accounts for inclusion probabilities, non-response and post-stratification). For each of the following waves t ($t>t_{i1}$), household longitudinal weight ($w_{i,t}^{(L)}$) is the product of its longitudinal weight at the previous wave[34] and an adjusting factor a, which summarises all the reasons behind the changes in the panel-sample between the two periods (Verma, 1995):[35]

[30] We showed that cross-sectional weights are somehow adjusted to account for panel attrition (equation 3). But since the adjustment in each wave uses information on the previous one only, we consider this correction mainly finalised to the autocorrelation adjusted estimator, more than a means of properly introducing longitudinal information in the weights.

[31] This limit can be overcome using some post-stratification techniques to align certain characteristics of the panel domain to known population totals.

[32] In the SHIW, for example, the rate of cooperation of panel and non-panel sample households is very different (also in terms of patterns): during the last 3 waves (2000-2002-2004) the average response rate of the total sample was around 40 per cent, while the cooperation of the panel component was something between 66 and 75 per cent. See Ernst (1989) for a review of the issues regarding longitudinal weighting.

[33] For a more detailed definition of the longitudinal population see Folsom et al. (1989).

[34] We take the interval between two waves to be the basic unit of time. For the SHIW, this has been generally equal to two years, with the exception of 1998 survey, which was delayed by one year.

[35] As in the case of cross-sectional weights, each member of the household has the same longitudinal weight, that is equal to the family one.

(8) $$w_{i,t}^{(L)} = w_{i,t-1}^{(L)} * a_{i,t-1\Rightarrow t}$$

In particular, *a* can be decomposed into three factors that can be ascribed to inclusion probabilities ($a^{(pr)}$), non-response ($a^{(nr)}$) and post-stratification adjustments ($a^{(ps)}$):

(9) $$a_{i,t-1\Rightarrow t} = a_{i,t-1\Rightarrow t}^{(pr)} * a_{i,t-1\Rightarrow t}^{(nr)} * a_{i,t-1\Rightarrow t}^{(ps)}$$

In order to compute $a^{(pr)}$ we need to analyse all the information regarding the selection process of panel units in the SHIW. At each wave all previous survey panel units are contacted again. Nevertheless, some families will leave the panel (*attrition*) because they refuse to participate, are not found at home or are no more in the survey population (*non-eligibility*, e.g. none of the family members are still alive, the family has moved to another country, etc.). In order to compensate for the loss of sample units due to the attrition in the panel, a random subset of families interviewed for the first time in the last wave is added (*sample refreshing*). In symbols, the *target panel sample* at time *t* (P_t^*) is thus equal to the sum of all panel units interviewed in the previous wave (P_{t-1}) plus a fraction f_t of non-panel units of the same survey (Q_{t-1}):[36]

(10) $$P_t^* = (P_{t-1} \cup f_t Q_{t-1})$$

This means that the probability of selecting a household is inflated by the fraction of new families included each year in the target panel through sample refreshing.[37] In particular, the probability of selecting, in the target panel, a household interviewed in the last wave is equal to: $\pi_{i,t-1\Rightarrow t}^{(ref)} = pr_{S_{t-1}}(i \in P_t^*) = pr_{S_{t-1}}(i \in P_{t-1}) + pr_{S_{t-1}}(i \in f_t Q_{t-1}) = \dfrac{p_{t-1} + f_t q_{t-1}}{n_{t-1}}$, while without refreshing the probability is: $\pi_{i,t-1\Rightarrow t}^{(noref)} = pr_{S_{t-1}}(i \in P_t^*) = pr_{S_{t-1}}(i \in P_{t-1}) = \dfrac{p_{t-1}}{n_{t-1}}$, where $p_{t\text{-}1}$ and $q_{t\text{-}1}$ are, respectively, the number of panel and non-panel units interviewed in the previous wave and $n_{t\text{-}1} = p_{t\text{-}1} + q_{t\text{-}1}$ is the total number households interviewed in wave t-1.[38] This means that the probability of selection in the case of refreshing is inflated by a factor $\dfrac{p+fq}{p} > 1$ as $\pi_{i,t-1\Rightarrow t}^{(ref)} = \dfrac{p+fq}{p}\pi_{i,t-1\Rightarrow t}^{(noref)}$. In order to correct for this selection mechanism, longitudinal weights must be reduced by the inverse of the same factor, therefore:

---

[36] It follows that the total sample in *t*-1 ($S_{t\text{-}1}$) is given from the union of panel and non panel units: $S_{t-1} = (P_{t-1} \cup Q_{t-1})$.

[37] In fact, because the panel households are representative of the same domain, they can enter the sample both through the original and the additional selection (Verma, 1995).

[38] The symbol pr_A specifies that the probability is conditional to the event "inclusion to the set *A*". For example $pr_{S_{t-1}}(i \in P_t^*) = pr(i \in P_t^* \mid i \in S_{t-1})$.

(11) $$a^{(pr)}_{i,t\Rightarrow t-1} = \frac{p}{p+fq}.$$

After the selection of the target panel, the *actual panel* at time *t* (P_t) will be composed of all the eligible households successfully contacted and which accepted to participate in the survey.

$$P_t = (E_t \cap C_t \cap M_t) = M_t \qquad \text{as } M_t \subseteq C_t \subseteq E_t \subseteq P_t^*$$

where E_t, C_t and M_t are the subsets of eligible, contacted and cooperative households, respectively.

The response probability is thus equal to the probability of household *i* cooperating in the interview, given that it has been contacted and that it is eligible.

This probability can be estimated using data available from previous interviews. In this case, a parsimonious approach must be adopted by choosing the smallest set of variables that are most related to the specific probability, without redundant information.[39] The literature suggests different sets of variables that can be useful in estimating this probability, like household characteristics, socio-demographic class, economic status and financial situation, geographical location, income and main source of income.[40] Once we have estimated the overall response propensity as $\hat{\pi}^{(r)}_{i,t-1\Rightarrow t}$, the non-response adjustment factor $a^{(nr)}_{i,t-1\Rightarrow t}$ will be equal to $1/\hat{\pi}^{(r)}_{i,t-1\Rightarrow t}$.

In the case of SHIW data we estimate the response probability using a logistic model based on the 2002-04 response patterns. In particular, as suggested by the theory, the variables included in the model account for, respectively, non-response due to cooperation habits (waves of past cooperation, past survey experience, such as information provided by the interviewer on the climate of the interview) and non-contact patterns (old age of all the components of the family and city size, also considered in combination with the geographical area).[41] The results of the logistic model (Table A4)

---

[39] Several approaches have been proposed to make profitable use of information recalled from previous surveys in order to reduce the bias due to non-response in longitudinal data. Lepkowski (1989) provides a review of some approaches based on the Missing at Random (MAR) hypothesis, Fay (1989) advocates the use of the casual models to deal with non-ignorable non-response processes. Finally, Bailey (2005) assesses the bias due to the use of improper methods to correct for non-response when the underlying hypothesis is not true.

[40] In particular, for each probability it can be useful to refer to a particular set of information. Lepkowski and Couper (2002) provide a recent review and some applications regarding the estimation of non-response probabilities in longitudinal surveys.

[41] It is worth noting that this set of variables is different from the one usually used to estimate non-response probability in a cross-sectional context as this model refers only to panel families that have already been interviewed. In this case, the subject of the model is thus the probability of response given past cooperation, which is often uncorrelated with other variables, such wealth and income, which are instead crucial in explaining response for families that collaborate for the first time. For further details see Giraldo et al. (2001) and Cannari and D'Alessio (1992).

show that the presence of old age for all the household's components increases the non-response probability, possibly due to non-eligibility patterns. The response rate due to cooperation is lower for households living in big cities. As expected, the non-response probability is negatively associated with past participation in more surveys and with the "climate"[42] of the interview.

By using these model estimates we can derive, for each household interviewed at time t, a response propensity. For each year, estimated response probabilities to participate in the survey are calibrated to the yearly average response rates. Thus, for each wave subsequent to the first one we are able to estimate a factor $\hat{a}^{(nr)}_{t-1 \Rightarrow t}$ equal to the inverse of this probability in order to compensate for the households that did not participate in the survey.

Finally, the adjusting factor $a^{(ps)}$ has to compensate for the transition in population characteristics from one wave to another and to rebalance weights after the adjustments for selection probability and for attrition in order to improve the representativeness of the panel over time. As for cross-sectional weight, this aims to ensure that the weighted marginal distribution of the sample, with respect to some relevant characteristics, matches that of the population. The final weight $w^{(L)}_{i,t}$ is therefore calibrated to the same population total used in the third stage of calibration of the cross-sectional weight (equation 5).

Figure 2 depicts the distribution of longitudinal weights in different stages (weights are re-scaled to their mean to sum-up to sample size).

The first pane of Figure 2 shows how correcting for the differential effect of panel refreshing greatly increases the variability of the weights. As expected, the non-response corrected weights and the post-stratification, as pane 2 illustrates, also further increase the dispersion of the weights.

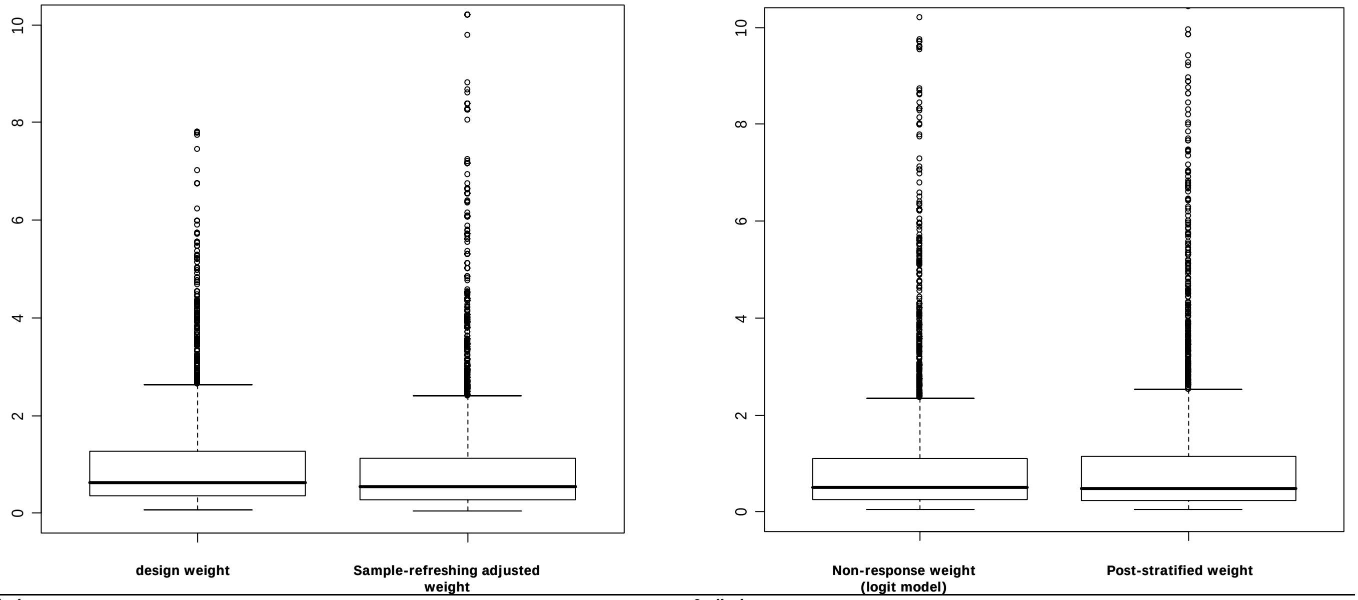
It is worth noting that this set of longitudinal weights has a large impact on the variance of the estimators. Using the Kish rule of thumb, the variance of the final longitudinal weight estimator is almost three times that of the unweighted estimator (L_w =2.97). Nevertheless, these weights yield a more accurate representation of the population dynamics as they exploit all the available information regarding the composition and the "age" of the SHIW panel.

[42] The "climate" refers to the collaboration provided by the household and in general to the environment in which the interview has conducted.

Fig.2

Distributions of longitudinal weights at various stages of the weighting process

(*box plots of normalised weights*)



Coeff. of Variation	102.3	148.6	Coeff. of Variation	164.9	172.3

4.2 The validation of longitudinal weights on SHIW estimates

Given their enhanced description of the "panel population", the longitudinal weights obtained from the process previously described can be expected to be better suited to perform longitudinal analysis. To assess their impact on SHIW estimates we computed different longitudinal statistics of the degree of mobility of income and wealth and of the autocorrelation coefficient of income and wealth with respect to their components for the 2002-04 panels (first part of Table A5).

The use of longitudinal weights results in a higher persistency of the analysed phenomena: in the transition matrices of both income and wealth the percentage of households on the main diagonal increases (the relative Shorrock index[43] decreases from 0.574 to 0.534 for income and from 0.587 to 0.570 for wealth). The autocorrelation between income and its components increases. The same result applies on average for wealth (second part of Table A5).

With respect to the same statistics we computed the relative MSE, using cross-section weights as a benchmark, and we calculated the empirical MSE for both the estimators using cross-section weights post-stratified according to the same distributions considered for cross-sectional and longitudinal weights (tables A6-A7).[44] In many cases longitudinal weights perform better, in terms of

[43] The relative Shorrock's index computed on a transition matrix T presenting *k* classes is equal to $k\text{-}tr(T)/k\text{-}1$. It ranges between 0 (no transition) and $k/(k-1)$ (perfect mobility).

[44] We use equation 7, considering the estimators using longitudinal weights unbiased.

MSE, with respect to cross-section weights with post-stratification, despite their higher degree of variability. Longitudinal weights determine a major loss in terms of MSE when estimating the autocorrelation of wealth with liabilities and financial assets. This result can be due to the small number of families holding these kinds of assets. The high variability of this phenomenon is additionally inflated by the use of longitudinal weights overcoming the reduction in bias.

In general, these preliminary results suggest that, even if longitudinal weights are more appropriate in performing longitudinal analyses, they can be less efficient. The efficiency of this set of weights decreases when dealing with rather dispersed study variables (such as financial assets).

5. Conclusions

In this paper we presented the weighting scheme of the SHIW and assessed its impact on the bias and variance of some estimators. The empirical analysis shows that even if the effect of the weights at various stages is to increase the variance, the final outcome on the estimates is usually a reduction of the mean square error that is considerable for the key survey variables (i.e. income, expenditure and wealth). This result seems to hold even when we consider only sample subsets (domain).

A new longitudinal weight, to be used when performing longitudinal analysis, is also presented. These weights are more appropriate to perform longitudinal analyses but are not always more efficient, in particular when dealing with heavily skewed study variables.

As a practical suggestion, we can therefore conclude that weighted estimators perform definitely better than the unweighted ones in terms of MSE in the case of cross-section analysis. Even in domain analysis, the increasing variability induced by using weights is more than compensated by their bias reduction property. Nevertheless, in this case we suggest including all the auxiliary information regarding domain characteristics (e.g. using post-stratification) in order to maximise the bias reduction function of weighting.

For the production of longitudinal statistics, there is no unambiguous evidence that the use of longitudinal weights always performs better than cross-sectional weighing in terms of MSE.

References

Bailey, L. (2005), "Weighting Alternatives to Compensate for Longitudinal Nonresponse in the Survey of Income and Program Participation", U.S. Census Bureau, Research Report Series, Statistics, No. 2005-03.

Bank of Italy (2006), "Italian Household Budgets in 2004", in I. Faiella, R. Gambacorta, S. Iezzi and A. Neri (eds.), *Supplementi al Bollettino Statistico* (*indagini campionarie*), Bank of Italy, No. 7, January.

Bethlehem, J.G. (2002), "Weighting Nonresponse Adjustments Based on Auxiliary Information", in R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little (eds.), *Survey Nonresponse*, New York, Wiley, pp. 275-87.

Brandolini, A., L. Cannari, G. D'Alessio and I. Faiella (2004), "Household Wealth Distribution in Italy in the 1990s", Bank of Italy, Temi di Discussione, No. 504, December.

Cannari, L. and G. D'Alessio (1992), "Mancate interviste e distorsione degli stimatori", Bank of Italy, Temi di discussione, No. 172, June.

Chantala, K. (2001), "Constructing Weights to Use in Analyzing Pairs of Individuals from Add Health Data", Carolina Population Center, University of North Carolina at Chapel Hill, August.

Cochran, W.G. (1977), *Sampling Techniques*, third edition, New York, Wiley.

D'Alessio, G. and I. Faiella (2002), "Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth", Bank of Italy, Temi di discussione, No. 462, December.

Deaton, A.S. (1997), *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*, Baltimore and London, Johns Hopkins University Press for the World Bank.

Deville, J.C. and C.E. Särndal (1992), "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, Vol. 87, pp. 376-82.

Duncan, G.J. and G. Kalton (1987), "Issues on Design and Analysis of Survey Across Time", *International Statistical Review*, Vol. 55, No. 1, pp. 97-117.

Ernst, L.R. (1989), "Weighting Issues for Longitudinal Household and Family Estimates", in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.

Faiella, I. (2007), "Accounting for sampling design in the SHIW", Bank of Italy, mimeo.

Fay, R.E. (1989), "Estimating Nonignorable Nonresponse in Longitudinal Surveys though Casual Modeling", in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.

Fitzgerald, J., P. Gottschalk and R. Moffitt (1998), "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics", *The Journal of Human Resources*, Vol. 33, No. 2, pp. 251-99.

Folsom, R., L. LaVange and R.L. Williams (1989), "A probability sampling perspective on panel data analysis" in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley, pp. 108-38.

Gelman, A. and J.B. Carlin (2002), "Poststratification and Weighting Adjustments", in R.M Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little (eds.), *Survey Nonresponse*, New York, Wiley, pp. 289-302.

Giraldo, A., E. Rettore and U. Trivellato (2001), "Attrition Bias in the Bank of Italy's SHIW", *Proceedings of the International Conference on Quality in Official Statistics*, Stockholm, May 14-15 2001.

Groves, R.M, F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer and R. Tourangeau (2004), *Survey Methodology*, New York, Wiley.

Hansen, M.H., W.G. Madow and B.J. Tepping (1983), "An Evaluation of Model-Dependent and Probability –Sampling Inferences in Sample Surveys", *Journal of the American Statistical Association*, Vol. 78, No. 384, pp. 776-93.

Kalton, G. and I. Flores Cervantes (2003), "Weighting Methods", *Journal of Official Statistics*, Vol. 19, No. 2, pp. 81-97.

Kish, L. (1965), *Survey Sampling*, New York, Wiley, reprinted in 1995.

Kish, L. (1987), *Statistical Design for Research*, New York, Wiley, reprinted in 2004.

Kish, L. (1992), "Weighting for Unequal Pi", *Journal of Official Statistics*, Vol. 8, No. 2, pp. 183-200.

Lepkowski, J. M. (1989), "Treatment of Wave Nonresponse in Panel Surveys", in D. Kasprzyk, G.J. Duncan, G. Kalton and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.

Lepkowski, J.M. and M.P. Couper (2002), "Nonresponse in the Second Wave of Longitudinal Household Surveys", in R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. (eds.), *Survey Nonresponse*, New York, Wiley.

Little, R.J.A. (1989), “Survey inference with weights for differential sample selection or Nonresponse”, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 62-71.

Little, R.J.A. and S. Vartivarian (2005), “Does Weighting for Nonresponse Increase the Variance of Survey Means?”, *Survey Methodology*, Vol. 31, No. 2, pp. 161-68.

Little, R.J.A., S. Lewitzky, S. Heeringa, J.M. Lepkowski and R.C. Kessier (1997), “Assessment of Weighting Methodology for the National Comorbidity Survey”, *American Journal of Epidemiology,* Vol. 146, No. 5, pp. 290-300.

Lohr, S.L. (1999), “Sampling: Design and Analysis”, Pacific Grove, CA, Duxbury Press.

Oh, H.L. and F.J. Scheuren (1983), “Weighting Adjustment for Unit Nonresponse”, in W.G. Madow, I. Olkin and D.B. Rubin (eds.), *Incomplete Data in Sample Surveys*, Vol. 2, New York, Academic Press.

Potter, F.J. (1990), “A Study of Procedures to Identify and trim Extreme Sampling Weights”, American Statistical Association, *ASA Proceedings of the Section on Survey Research Methods* (Alexandria, VA), pp. 225-30.

Purdon, S. and K. Pickering (2001), “The use of sampling weights in the analysis of the 1998 Workplace Employee Relations Survey”, National Institute of Economic and Social Research, July.

Särndal, C.E. and S. Lundström (2005), *Estimation in Surveys with Nonresponse*, New York, Wiley.

Särndal, C.E., B. Swensson and J. Wretman (1992), *Model Assisted Survey Sampling*, Springer Series in Statistics, New York, Springer-Verlag, reprinted in 2003.

Smith, T.M.F. (1991), “Post-Stratification”, *The Statistician*, Vol. 40, pp. 315-23.

Verma, V. (1995), “Longitudinal Weighting”, *ECHP Document PAN 51*.

Verma, V. (2000), “Advanced Sampling Methods”, *Manual for Statistical Trainers*, Statistical Institute for Asia and the Pacific, Tokyo, pp. 6.13-6.21.

Appendix A - Statistical tables

Table A1.

DESIGN AND RESPONSE ADJUSTED WEIGHTS

Stratum id	**HH interviewed** m_h (1)	**HH contacted** n'_h (2)	**HH in the population** N_h (3)	**Design weight** w^0 (4)=(3)/(2)	**Response rate** m_h/n'_h (5)	**Response-adj. weight** w^1 (6)=(4)/(5)
11	271	594	960,895	1,650.6	0.46	3,618.0
12	179	420	234,370	566.4	0.43	1,328.9
13	319	1,109	642,201	581.0	0.29	2,019.7
31	218	531	2,222,994	4,191.9	0.41	10,210.6
32	120	407	424,867	1,056.5	0.29	3,583.3
33	505	1,710	1,133,845	706.9	0.30	2,393.8
41	109	264	271,879	1,083.4	0.41	2,624.1
43	42	94	86,831	924.0	0.45	2,068.1
51	227	580	1,086,994	1,904.4	0.39	4,865.9
52	195	472	248,096	531.2	0.41	1,285.8
53	156	514	418,286	814.0	0.30	2,682.1
61	144	377	304,917	812.9	0.38	2,128.2
62	47	75	30,119	401.7	0.63	641.0
63	59	144	151,995	1,174.6	0.41	2,866.8
71	111	214	242,924	1,140.6	0.52	2,199.0
72	89	170	73,051	446.6	0.52	853.0
73	171	1,049	377,343	358.1	0.16	2,196.8
81	180	414	800,210	1,967.6	0.43	4,525.5
82	121	373	178,335	494.4	0.32	1,524.0
83	377	1,119	842,350	774.9	0.34	2,299.9
91	133	260	596,769	2,348.0	0.51	4,590.0
92	85	191	235,837	1,247.5	0.45	2,803.2
93	423	1,201	702,423	597.6	0.35	1,696.7
101	157	292	136,045	471.4	0.54	876.8
102	64	164	61,411	374.7	0.39	960.2
103	62	113	122,955	1,105.9	0.55	2,015.6
111	158	386	283,300	739.4	0.41	1,806.5
112	85	256	98,976	386.9	0.33	1,165.2
113	146	428	165,009	393.2	0.34	1,152.6
121	105	210	477,713	2,286.0	0.50	4,572.0
122	23	53	211,011	3,981.3	0.43	9,174.4
123	297	1,476	1,388,179	1,057.1	0.20	5,253.7
131	102	291	422,499	1,472.1	0.35	4,199.7
132	109	180	100,436	558.9	0.61	922.9
133	108	318	145,244	468.1	0.34	1,378.2
151	121	204	780,406	4,108.8	0.59	6,927.3
152	59	182	313,130	1,732.8	0.32	5,345.3
153	446	1,587	833,175	541.9	0.28	1,928.4
161	67	136	526,195	3,943.8	0.49	8,005.2
162	66	134	301,681	2,324.5	0.49	4,719.5
163	317	760	584,709	817.4	0.42	1,959.6
181	178	348	667,226	1,944.2	0.51	3,801.0
182	22	54	78,054	1,445.5	0.41	3,547.9
183	119	247	221,224	941.1	0.48	1,953.4
191	43	114	602,109	5,281.7	0.38	14,002.5
192	88	201	334,720	1,688.9	0.44	3,857.5
193	459	1,346	795,208	576.5	0.34	1,690.5
201	149	300	362,683	1,229.6	0.50	2,475.8
202	108	254	99,199	399.4	0.43	939.3
203	73	251	128,337	534.2	0.29	1,836.8
Total	**8,012**	**22,018**	**22,508,364**	**1,117.79**	**0,36**	**2,883.87**

Table A2.

VARIABILITY AND MSE OF UNWEIGHTED AND WEIGHTED ESTIMATES

(*SHIW 2004*)

	Unweighted	Design weight (w^0)	Non-response adjusted weight (w^1)	Calibrated- weight (w^2): Attrition-adjusted weight ($w^{2'}$)	Calibrated- weight (w^2): Autocorrelation-corrected weight ($w^{2''}$)	Calibrated- weight (w^2): Final weight ($w^{2'''}$)
			Estimates			
Age of the HH	55.7	56.2	56.0	55.6	55.8	53.8
Household size	2.569	2.527	2.505	2.510	2.529	2.577
HH employee	42.2	40.9	41.9	42.9	42.7	46.4
HH self-employed	12.4	12.1	11.9	12.3	12.4	13.2
Total income	29,866	28,802	29,310	28,976	29,203	29,484
Wages	11,182	10,646	11,053	11,105	11,165	11,998
Income from self-employment	4,495	4,174	4,163	4,254	4,287	4,526
Transfers	8,012	7,985	7,888	7,549	7,605	6,939
Net wealth	213,817	200,284	204,379	199,896	204,537	199,943
Tangible assets	199,119	187,158	190,453	186,667	190,710	187,416
Financial assets	23,397	20,431	21,363	20,865	21,648	21,226
Liabilities	8,700	7,305	7,437	7,635	7,821	8,699
Expenditure	22,390	21,596	21,993	21,800	21,920	22,139
			Coefficient of variation (percentages)*			
Age of the HH	0.386	0.593	0.531	0.535	0.517	0.564
Household size	0.638	1.059	0.942	0.959	0.894	0.938
HH employee	1.387	1.804	1.596	1.581	1.629	1.608
HH self-employed	3.142	4.921	4.855	4.791	4.563	4.356
Total income	0.906	1.093	0.965	0.962	0.969	0.980
Wages	1.502	2.119	1.832	1.820	1.820	1.797
Income from self-employment	4.339	4.938	4.728	4.744	4.729	4.768
Transfers	1.432	1.759	1.618	1.624	1.647	1.823
Net wealth	1.738	2.333	2.143	2.091	2.082	2.023
Tangible assets	1.790	2.294	2.111	2.068	2.096	2.048
Financial assets	3.423	3.794	3.616	3.574	3.403	3.319
Liabilities	9.038	9.626	9.179	9.259	9.088	9.906
Expenditure	0.735	1.026	0.919	0.912	0.907	0.876
			Relative Mean Square Error (percentages)**			
Household characteristics						
Age of the head of household	100.0	154.7	127.1	90.0	105.1	2.4
Household size	100.0	1,147.5	2,097.0	1,877.3	1,046.6	217.8
Head of household employee	100.0	169.5	112.6	71.7	78.6	3.1
Head of household self-employed	100.0	267.8	352.1	216.4	180.5	61.6
Average	**100.0**	**434.9**	**672.2**	**563.8**	**352.7**	**71.2**
Economic variables						
Total income	100.0	291.2	54.8	178.9	84.5	44.4
Wages	100.0	277.4	137.9	124.2	109.0	6.9
Income from self-employment	100.0	409.4	434.3	301.8	258.5	122.4
Transfers	100.0	96.0	79.1	33.4	39.8	1.4
Net wealth	100.0	10.9	18.6	8.7	19.6	8.2
Tangible assets	100.0	12.7	16.5	10.7	18.5	10.2
Financial assets	100.0	22.8	11.8	13.6	14.3	9.8
Liabilities	100.0	383.1	321.6	263.7	206.3	120.1
Expenditure	100.0	453.5	78.4	207.9	117.8	50.7
Average	**100.0**	**217.5**	**128.1**	**127.0**	**96.5**	**41.6**

* Percentage standard error estimated on 325 *jackknife* replications, divided by the estimate of the sample. Individual characteristics are those of the head of household, i.e. the person earning the highest income. ** Mean squared errors (MSE) of weighted estimates expressed as a percentage of the unweighted MSE. The MSEs are estimated using 325 *jackknife* replications according to sample design (for details see Faiella, 2007). The MSE corrects the value of the squared bias with a *jackknife* estimate of the variance of the difference between the unweighted and the weighted estimator (see Little et al., 1997). In our computations the unbiased estimator is the weighted average using the final calibrated weights ($w^{2'''}$).

Table A3.

RELATIVE MEAN SQUARE ERROR*

(*SHIW 2004*)

Characteristics	Non-response adjusted weight (w^1)			Final weight ($w^{2'''}$)		
	Income	Expenditure	Wealth	Income	Expenditure	Wealth
Gender						
male	53.2	72.9	32.1	55.6	64.0	12.9
female	109.6	222.0	17.2	79.5	89.0	14.1
Age						
up to 30 years	132.1	237.0	119.5	57.6	163.7	66.7
31 to 40	92.0	40.5	53.3	46.5	24.9	47.5
41 to 50	53.7	40.6	27.0	50.6	41.0	25.9
51 to 65	228.3	269.3	123.9	151.9	181.8	103.7
over 65	152.9	99.1	51.4	176.0	85.9	57.3
Education						
none	37.6	31.7	16.2	38.0	30.8	14.1
elementary school	202.4	156.8	17.1	215.1	136.2	18.6
middle school	30.4	39.9	33.6	26.9	37.0	10.7
high school	134.0	161.3	175.3	74.9	129.7	96.9
university degree	238.7	166.5	76.7	162.7	91.9	35.6
Work status						
Payroll employed	106.5	45.7	69.8	16.8	12.4	9.8
Self-employed	26.4	44.7	29.8	19.2	28.7	25.6
Not employed	66.0	66.9	35.5	60.8	61.1	29.1
Household size						
1 member	61.7	49.1	42.6	58.9	45.1	18.0
2 members	117.1	139.5	98.4	129.3	103.7	60.0
3 members	66.0	92.8	28.5	40.2	62.8	20.0
4 members	47.4	52.3	33.3	31.6	31.3	20.2
5 members or more	50.3	14.8	183.8	32.2	11.6	172.1
Town size						
up to 20,000 inhabitants	73.1	164.0	16.4	49.4	80.2	13.4
from 20,000 to 40,000	128.1	25.9	28.6	100.8	14.9	29.5
from 40,000 to 500,000	107.0	99.7	496.6	90.1	67.6	105.1
more than 500,000	7.4	18.0	144.9	8.6	13.6	73.0
Geographical area						
North	53.9	142.9	34.9	60.9	131.0	13.2
Centre	130.9	187.7	36.6	69.0	117.8	26.4
South and Islands	20.5	21.2	24.9	15.6	15.8	22.0
Total	**54.8**	**78.4**	**18.6**	**44.4**	**50.7**	**8.2**

* Mean squared errors (MSE) of weighted estimates expressed as a percentage of the unweighted MSE. The MSEs are estimated using 325 *jackknife* replications according to sample design (for details see Faiella, 2007). The MSE corrects the value of the squared bias with a *jackknife* estimate of the variance of the difference between the weighted and the unweighted estimator (see Little et al., 1997). In our computations the unbiased estimator is the weighted average using the final calibrated weights ($w^{2'''}$)

Table A4.

LOGIT MODEL FOR NON-RESPONSE PROBABILITIES FOR PANEL UNITS

Parameter	Coefficient Estimate	P-value	Odds Ratio	95% Wald Confidence Limits
Intercept	0.3239	0.5008		
Number previous interviews	-0.168	0.0002	0.845	0.774-0.923
Climate	-0.1242	0.0243	0.883	0.793-0.984
Town with more than 500,000 inhabitants	0.721	0.0146	2.056	1.153-3.667
Town with more than 500,000 inhabitants in the North	-0.1161	0.7531	0.890	0.432-1.836
Town with more than 500,000 inhabitants in the Centre	0.527	0.1688	1.694	0.800-3.588
All family members over 80	0.6352	0.0961	1.887	0.893-3.988

Max-rescaled R-Square=0.06; Percent Concordant=61.8; Percent Discordant=34.7. Sample size, 4.842: 1.129 non-respondents (weighted).

Table A5.

HOUSEHOLD INCOME AND WEALTH: COMPARISON BETWEEN 2002 AND 2004

Transition Matrices

Cross-section weights

Household income ↓2002 / 2004➔	up to €10,000	€10,000 - €20,000	€20,000 - €30,000	€30,000 - €40,000	more than €40,000	Total
up to €10,000	52.8	35.3	6.5	1.6	3.7	100.0
€10,000 - €20,000	5.9	60.8	22.2	7	4	100.0
€20,000 - €30,000	1.6	17.1	47.3	23.5	10.5	100.0
€30,000 - €40,000	1.3	7.7	15.6	35.9	39.5	100.0
more than €40,000	0.7	2.5	9.3	13.7	73.8	100.0

Longitudinal weights

Household income ↓2002 / 2004➔	up to €10,000	€10,000 - €20,000	€20,000 - €30,000	€30,000 - €40,000	more than €40,000	Total
up to €10,000	60.6	26.8	6.4	0.7	5.5	100.0
€10,000 - €20,000	5.1	65.8	19.4	7.1	2.6	100.0
€20,000 - €30,000	1.5	21.3	45.5	23.0	8.7	100.0
€30,000 - €40,000	2.1	7.1	17.9	38.0	34.9	100.0
more than €40,000	0.3	2.5	8.4	12.4	76.5	100.0

Cross-section weights

Household wealth ↓2002 / 2004➔	up to €20,000	€20,000 - €50,000	€50,000 - €100,000	€100,000 - €200,000	more than €100,000	Total
up to €20,000	70.3	11.4	6.3	4.9	7.1	100.0
€20,000 - €50,000	15.4	33.1	25.7	17.1	8.7	100.0
€50,000 - €100,000	9.5	6.5	35.2	35.6	13.2	100.0
€100,000 - €200,000	3.2	3.5	8.6	44.5	40.3	100.0
more than €100,000	1.8	0.6	1.6	13.8	82.3	100.0

Longitudinal weights

Household wealth ↓2002 / 2004➔	up to €20,000	€20,000 - €50,000	€50,000 - €100,000	€100,000 - €200,000	more than €100,000	Total
up to €20,000	67.0	10.1	10.9	5.6	6.5	100.0
€20,000 - €50,000	12.5	41.6	26.0	12.7	7.2	100.0
€50,000 - €100,000	8.6	5.7	38.0	35.4	12.3	100.0
€100,000 - €200,000	1.4	3.5	8.7	43.3	43.1	100.0
more than €100,000	1.2	0.8	1.6	14.3	82.1	100.0

Pearson Autocorrelation

	Payroll employment	**Self-employment**	**Transfers**	**Total Income**
Cross-section weights................................	80.7	52.8	81.6	63.4
Longitudinal weights	*82.5*	*66.1*	*84.4*	*71.0*
	Real assets	**Financial assets**	**Liabilities**	**Net wealth**
Cross-section weights................................	74.1	29.2	29.4	73.3
Longitudinal weights	*78.0*	*26.9*	*29.5*	*78.1*

Table A6.

RELATIVE MSE* OF INCOME AND WEALTH TRANSITION MATRICES

(MSE with cross-section weights=100)

Cross-section weights + post-stratification

Household income ↓2002 2004→	up to €10,000	€10,000 - €20,000	€20,000 - €30,000	€30,000 - €40,000	more than €40,000
up to €10,000	141.9	89.8	188.1	80.1	138.8
€10,000 - €20,000	99.8	154.7	153.2	159.9	106.7
€20,000 - €30,000	100.0	105.4	92.8	104.1	148.3
€30,000 - €40,000	96.7	100.7	117.9	96.7	95.8
more than €40,000	110.7	97.3	120.1	83.3	84.6
Average	**109.8**	**109.6**	**134.4**	**104.8**	**114.8**

Longitudinal weights

Household income ↓2002 2004→	up to €10,000	€10,000 - €20,000	€20,000 - €30,000	€30,000 - €40,000	more than €40,000
up to €10,000	51.7	25.0	139.1	7.0	142.8
€10,000 - €20,000	49.1	22.6	40.6	223.3	12.5
€20,000 - €30,000	93.2	42.1	130.2	139.9	55.2
€30,000 - €40,000	132.1	116.8	84.6	145.7	45.1
more than €40,000	6.6	231.8	90.3	67.1	50.5
Average	**66.5**	**87.7**	**96.9**	**116.6**	**61.2**

Cross-section weights + post-stratification

Household wealth ↓2002 2004→	up to €20,000	€20,000 - €50,000	€50,000 - €100,000	€100,000 - €200,000	more than €100,000
up to €20,000	104.4	89.0	104.1	111.5	165.7
€20,000 - €50,000	96.9	91.8	106.1	115.0	105.7
€50,000 - €100,000	127.6	80.9	134.9	104.5	189.8
€100,000 - €200,000	84.0	206.4	104.9	87.0	83.3
more than €100,000	137.6	101.8	106.2	126.6	107.5
Average	**110.1**	**114.0**	**111.3**	**108.9**	**130.4**

Longitudinal weights

Household wealth ↓2002 2004→	up to €20,000	€20,000 - €50,000	€50,000 - €100,000	€100,000 - €200,000	more than €100,000
up to €20,000	89.2	65.4	59.5	139.9	135.7
€20,000 - €50,000	69.7	79.5	256.8	32.7	77.7
€50,000 - €100,000	79.2	79.9	80.9	128.5	103.9
€100,000 - €200,000	3.7	114.5	156.7	150.0	77.9
more than €100,000	21.8	198.8	140.3	169.1	173.3
Average	**52.7**	**107.6**	**138.9**	**124.0**	**113.7**

* Mean squared errors (MSE) of weighted estimates expressed as a percentage of the MSE computed using cross-section weights. The MSEs are estimated using 58 *jackknife* replications according to sample design. In our computations the unbiased estimator is the weighted average using the longitudinal calibrated weights.

Table A7.

RELATIVE MSE* OF INCOME AND WEALTH: PEARSON AUTOCORRELATION

(MSE with cross-section weights =100)

	Payroll employment	Self-employment	Transfers	Total Income
Cross-section weights + post-stratification..	120.9	110.9	92.4	113.2
Longitudinal weights	38.4	29.0	26.0	15.8
	Real assets	**Financial assets**	**Liabilities**	**Net wealth**
Cross-section weights + post-stratification..	109.8	101.2	94.3	116.7
Longitudinal weights	88.8	207.2	131.8	59.0

* Mean squared errors (MSE) of weighted estimates expressed as a percentage of the MSE computed using cross-section weights. The MSEs are estimated using 58 *jackknife* replications according to sample design. In our computations the unbiased estimator is the weighted average using the longitudinal weights.

RECENTLY PUBLISHED "TEMI" (*)

N. 612 – *A neural network architecture for data editing in the Bank of Italy's business surveys* by Claudia Biancotti, Leandro D'Aurizio and Raffaele Tartaglia-Polcini (February 2007).

N. 613 – *Outward FDI and Local Employment Growth in Italy*, by Stefano Federico and Gaetano Alfredo Minerva (February 2007).

N. 614 – *Testing for trend,* by Fabio Busetti and Andrew Harvey (February 2007).

N. 615 – *Macroeconomic uncertainty and banks' lending decisions: The case of Italy,* by Mario Quagliariello (February 2007).

N. 616 – *Entry barriers in italian retail trade,* by Fabiano Schivardi and Eliana Viviano (February 2007).

N. 617 – *A politicy-sensible core-inflation measure for the euro area,* by Stefano Siviero and Giovanni Veronese (February 2007).

N. 618 – *Le opinioni degli italiani sull'evasione fiscale,* by Luigi Cannari and Giovanni D'Alessio (February 2007)

N. 619 – *Memory for prices and the euro cash changeover: An analysis for cinema prices in Italy*, by Vincenzo Cestari, Paolo Del Giovane and Clelia Rossi-Arnaud (February 2007).

N. 620 – *Intertemporal consumption choices, transaction costs and limited participation in financial markets: Reconciling data and theory*, by Orazio P. Attanasio and Monica Paiella (April 2007).

N. 621 – *Why demand uncertainty curbs investment: Evidence from a panel of Italian manufacturing firms*, by Maria Elena Bontempi, Roberto Golinelli and Giuseppe Parigi (April 2007).

N. 622 – *Employment, innovation and productivity: Evidence from Italian microdata*, by Bronwyn H. Hall, Francesca Lotti and Jacques Mairesse (April 2007).

N. 623 – *Measurement of Income Distribution in Supranational Entities: The Case of the European Union*, by Andrea Brandolini (April 2007).

N. 624 – *Un nuovo metodo per misurare la dotazione territoriale di infrastrutture di trasporto*, by Giovanna Messina (April 2007).

N. 625 – *The forgone gains of incomplete portfolios,* by Monica Paiella (April 2007).

N. 626 – *University drop-out: The case of Italy,* by Federico Cingano and Piero Cipollone (April 2007).

N. 627 – *The sectoral distribution of money supply in the euro area,* by Giuseppe Ferrero, Andrea Nobili and Patrizia Passiglia (April 2007).

N. 628 – *Changes in transport and non-transport costs: Local vs global impacts in a spatial network*, by Kristian Behrens, Andrea R. Lamorgese, Gianmarco I.P. Ottaviano and Takatoshi Tabuchi (April 2007).

N. 629 – *Monetary policy shocks in the euro area and global liquidity spillovers*, by João Sousa and Andrea Zaghini (June 2007).

N. 630 – *Endogenous growth and trade liberalization between asymmetric countries*, by Daniela Marconi (June 2007).

N. 631 – *New Eurocoin: Tracking economic growth in real time*, by Filippo Altissimo, Riccardo Cristadoro, Mario Forni, Marco Lippi and Giovanni Veronese (June 2007).

N. 632 – *Oil supply news in a VAR: Information from financial markets*, by Alessio Anzuini, Patrizio Pagano and Massimiliano Pisani (June 2007).

N. 633 – *The reliability of EMU fiscal indicators: Risks and safeguards*, by Fabrizio Balassone, Daniele Franco and Stefania Zotteri (June 2007).

N. 634 – *Prezzi delle esportazioni, qualità dei prodotti e caratteristiche di impresa: un'analisi su un campione di imprese italiane*, by Matteo Bugamelli (June 2007).

N. 635 – *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms*, by Massimo Del Gatto, Gianmarco I.P. Ottaviano and Marcello Pagnini (June 2007).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi – Divisione Biblioteca e pubblicazioni – Via Nazionale, 91 – 00184 Rome (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2004

P. ANGELINI and N. CETORELLI, *Gli effetti delle modifiche normative sulla concorrenza nel mercato creditizio,* in F. Panetta (eds.), Il sistema bancario negli anni novanta: gli effetti di una trasformazione, Bologna, il Mulino, **TD No. 380 (October 2000)**.

P. CHIADES and L. GAMBACORTA, *The Bernanke and Blinder model in an open economy: The Italian case,* German Economic Review, Vol. 5 (1), pp. 1-34, **TD No. 388 (December 2000)**.

M. BUGAMELLI and P. PAGANO, *Barriers to investment in ICT*, Applied Economics, Vol. 36 (20), pp. 2275-2286, **TD No. 420 (October 2001)**.

F. BUSETTI, *Preliminary data and econometric forecasting: An application with the Bank of Italy quarterly model*, CEPR Discussion Paper, 4382, **TD No. 437 (December 2001)**.

A. BAFFIGI, R. GOLINELLI and G. PARIGI, *Bridge models to forecast the euro area GDP*, International Journal of Forecasting, Vol. 20 (3), pp. 447-460,**TD No. 456 (December 2002)**.

D. AMEL, C. BARNES, F. PANETTA and C. SALLEO, *Consolidation and efficiency in the financial sector: A review of the international evidence,* Journal of Banking and Finance, Vol. 28 (10), pp. 2493-2519, **TD No. 464 (December 2002).**

M. PAIELLA, *Heterogeneity in financial market participation: Appraising its implications for the C-CAPM,* Review of Finance, Vol. 8, 3, pp. 445-480, **TD No. 473 (June 2003)**.

F. CINGANO and F. SCHIVARDI, *Identifying the sources of local productivity growth,* Journal of the European Economic Association, Vol. 2 (4), pp. 720-742, **TD No. 474 (June 2003)**.

E. BARUCCI, C. IMPENNA and R. RENÒ, *Monetary integration, markets and regulation*, Research in Banking and Finance, (4), pp. 319-360, **TD No. 475 (June 2003)**.

G. ARDIZZI, *Cost efficiency in the retail payment networks: first evidence from the Italian credit card system*, Rivista di Politica Economica, Vol. 94, (3), pp. 51-82, **TD No. 480 (June 2003)**.

E. BONACCORSI DI PATTI and G. DELL'ARICCIA, *Bank competition and firm creation,* Journal of Money Credit and Banking, Vol. 36 (2), pp. 225-251, **TD No. 481 (June 2003)**.

R. GOLINELLI and G. PARIGI, *Consumer sentiment and economic activity: a cross country comparison,* Journal of Business Cycle Measurement and Analysis, Vol. 1 (2), pp. 147-170, **TD No. 484 (September 2003)**.

L. GAMBACORTA and P. E. MISTRULLI, *Does bank capital affect lending behavior?*, Journal of Financial Intermediation, Vol. 13 (4), pp. 436-457, **TD No. 486 (September 2003)**.

F. SPADAFORA, *Il pilastro privato del sistema previdenziale: il caso del Regno Unito,* Economia Pubblica, 34, (5), pp. 75-114, **TD No. 503** (**June 2004**).

F. LIPPI. and S. NERI, *Information variables for monetary policy in a small structural model of the euro area*, Journal of Monetary Economics, v. 54, 4, pp. 1256-1270, **TD No. 511 (July 2004)**.

C. BENTIVOGLI and F. QUINTILIANI, *Tecnologia e dinamica dei vantaggi comparati: un confronto fra quattro regioni italiane*, in C. Conigliani (eds.), *Tra sviluppo e stagnazione: l'economia dell'Emilia-Romagna*, Bologna, Il Mulino, **TD No. 522 (October 2004)**.

G. GOBBI and F. LOTTI, *Entry decisions and adverse selection: An empirical analysis of local credit markets,* Journal of Financial services Research, Vol. 26 (3), pp. 225-244, **TD No. 535 (December 2004)**.

E. GAIOTTI and F. LIPPI, *Pricing behavior and the introduction of the euro: Evidence from a panel of restaurants*, Giornale degli Economisti e Annali di Economia, 2004, Vol. 63, (3/4), pp. 491-526, **TD No. 541 (February 2005)**.

A. CICCONE, F. CINGANO and P. CIPOLLONE, *The private and social return to schooling in Italy*, Giornale degli economisti e annali di economia, v. 63, 3-4, pp. 413-444, **TD No. 569 (January 2006)**.

2005

L. DEDOLA and F. LIPPI, *The monetary transmission mechanism: Evidence from the industries of 5 OECD countries,* European Economic Review, 2005, Vol. 49, (6), pp. 1543-1569, **TD No. 389 (December 2000)**.

D. Jr. MARCHETTI and F. NUCCI, *Price stickiness and the contractionary effects of technology shocks*. European Economic Review, v. 49, pp. 1137-1164, **TD No. 392 (February 2001)**.

G. CORSETTI, M. PERICOLI and M. SBRACIA, *Some contagion, some interdependence: More pitfalls in tests of financial contagion,* Journal of International Money and Finance, v. 24, 8, pp. 1177-1199, **TD No. 408 (June 2001)**.

GUISO L., L. PISTAFERRI and F. SCHIVARDI, *Insurance within the firm*. Journal of Political Economy, 113, pp. 1054-1087, **TD No. 414** (**August 2001**)

R. CRISTADORO, M. FORNI, L. REICHLIN and G. VERONESE, *A core inflation indicator for the euro area,* Journal of Money, Credit, and Banking, v. 37, 3, pp. 539-560, **TD No. 435 (December 2001)**.

F. ALTISSIMO, E. GAIOTTI and A. LOCARNO, *Is money informative? Evidence from a large model used for policy analysis,* Economic & Financial Modelling, v. 22, 2, pp. 285-304, **TD No. 445 (July 2002)**.

G. DE BLASIO and S. DI ADDARIO, *Do workers benefit from industrial agglomeration?* Journal of regional Science, Vol. 45, (4), pp. 797-827, **TD No. 453 (October 2002).**

R. TORRINI, *Cross-country differences in self-employment rates: The role of institutions,* Labour Economics, V. 12, 5, pp. 661-683, **TD No. 459 (December 2002)**.

A. CUKIERMAN and F. LIPPI, *Endogenous monetary policy with unobserved potential output,* Journal of Economic Dynamics and Control, v. 29, 11, pp. 1951-1983, **TD No. 493 (June 2004)**.

M. OMICCIOLI, *Il credito commerciale: problemi e teorie,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 494 (June 2004)**.

L. CANNARI, S. CHIRI and M. OMICCIOLI, *Condizioni di pagamento e differenziazione della clientela,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 495 (June 2004)**.

P. FINALDI RUSSO and L. LEVA, *Il debito commerciale in Italia: quanto contano le motivazioni finanziarie?,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 496 (June 2004)**.

A. CARMIGNANI, *Funzionamento della giustizia civile e struttura finanziaria delle imprese: il ruolo del credito commerciale,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 497 (June 2004)**.

G. DE BLASIO, *Credito commerciale e politica monetaria: una verifica basata sull'investimento in scorte,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 498 (June 2004)**.

G. DE BLASIO, *Does trade credit substitute bank credit? Evidence from firm-level data.* Economic notes, Vol. 34 (1), pp. 85-112, **TD No. 498 (June 2004).**

A. DI CESARE, *Estimating expectations of shocks using option prices,* The ICFAI Journal of Derivatives Markets, Vol. 2, (1), pp. 42-53, **TD No. 506 (July 2004).**

M. BENVENUTI and M. GALLO, *Il ricorso al "factoring" da parte delle imprese italiane,* in L. Cannari, S. Chiri e M. Omiccioli (eds.), *Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia,* Bologna, Il Mulino, **TD No. 518 (October 2004)**.

L. CASOLARO and L. GAMBACORTA, *Redditività bancaria e ciclo economico,* Bancaria, v. 61, 3, pp. 19-27, **TD No. 519 (October 2004)**.

F. PANETTA, F. SCHIVARDI and M. SHUM, *Do mergers improve information? Evidence from the loan market,* CEPR Discussion Paper, 4961, **TD No. 521 (October 2004)**.

P. DEL GIOVANE and R. SABBATINI, *La divergenza tra inflazione rilevata e percepita in Italia,* Bologna, Il Mulino, **TD No. 532 (December 2004).**

R. TORRINI, *Quota dei profitti e redditività del capitale in Italia: un tentativo di interpretazione,* Politica economica, v. 21, pp. 7-42, **TD No. 551 (June 2005)**.

M. OMICCIOLI, *Il credito commerciale come "collateral"*, in L. Cannari, S. Chiri, M. Omiccioli (eds.), Imprese o intermediari? Aspetti finanziari e commerciali del credito tra imprese in Italia, Bologna, il Mulino, **TD No. 553 (June 2005)**.

L. CASOLARO, L. GAMBACORTA and L. GUISO, *Regulation, formal and informal enforcement and the development of the household loan market. Lessons from Italy,* in Bertola G., Grant C. and Disney R. (eds.) The Economics of Consumer Credit: European Experience and Lessons from the US, Boston, MIT Press, **TD No. 560 (September 2005)**.

P. ANGELINI and F. LIPPI, *Did inflation really soar after the euro changeover? Indirect evidence from ATM withdrawals,* CEPR Discussion Paper, 4950, **TD No. 581 (March 2006)**.

S. DI ADDARIO, *Job search in thick markets: Evidence from Italy,* Oxford Discussion Paper 235, Department of Economics Series, **TD No. 605 (December 2006)**.

2006

F. BUSETTI, *Tests of seasonal integration and cointegration in multivariate unobserved component models,* Journal of Applied Econometrics, v. 21, 4, pp. 419-438, **TD No. 476 (June 2003).**

C. BIANCOTTI, *A polarization of inequality? The distribution of national Gini coefficients 1970-1996,* Journal of Economic Inequality, v. 4, 1, pp. 1-32, **TD No. 487 (March 2004)**.

L. CANNARI and S. CHIRI, *La bilancia dei pagamenti di parte corrente Nord-Sud (1998-2000),* in L. Cannari, F. Panetta (a cura di), Il sistema finanziario e il Mezzogiorno: squilibri strutturali e divari finanziari, Bari, Cacucci, **TD No. 490 (March 2004).**

M. BOFONDI and G. GOBBI, *Information barriers to entry into credit markets,* Review of Finance, Vol. 10 (1), pp. 39-67, **TD No. 509 (July 2004).**

LIPPI F. and W. FUCHS, *Monetary union with voluntary participation,* Review of Economic Studies, 73, pp. 437-457 **TD No. 512 (July 2004).**

GAIOTTI E. and A. SECCHI, *Is there a cost channel of monetary transmission? An investigation into the pricing behaviour of 2000 firms,* Journal of Money, Credit, and Banking, v. 38, 8, pp. 2013-2038 **TD No. 525 (December 2004)**.

A. BRANDOLINI, P. CIPOLLONE and E. VIVIANO, *Does the ILO definition capture all unemployment?,* Journal of the European Economic Association, v. 4, 1, pp. 153-179, **TD No. 529 (December 2004)**.

A. BRANDOLINI, L. CANNARI, G. D'ALESSIO and I. FAIELLA, *Household wealth distribution in Italy in the 1990s,* In E. N. Wolff (ed.) International Perspectives on Household Wealth, Cheltenham, Edward Elgar, **TD No. 530 (December 2004)**.

P. DEL GIOVANE and R. SABBATINI, *Perceived and measured inflation after the launch of the Euro: Explaining the gap in Italy*, Giornale degli economisti e annali di economia, v. 65, 2 , pp. 155-192, **TD No. 532 (December 2004).**

M. CARUSO, *Monetary policy impulses, local output and the transmission mechanism*, Giornale degli economisti e annali di economia, v. 65, 1, pp. 1-30, **TD No. 537 (December 2004).**

L. GUISO and M. PAIELLA, *The role of risk aversion in predicting individual behavior, In P. A. Chiappori e C. Gollier (eds.) Competitive Failures in Insurance Markets: Theory and Policy Implications, Monaco, CESifo*, **TD No. 546 (February 2005).**

G. M. TOMAT, *Prices product differentiation and quality measurement: A comparison between hedonic and matched model methods*, Research in Economics, No. 60, pp. 54-68, **TD No. 547 (February 2005)**.

F. LOTTI, E. SANTARELLI and M. VIVARELLI, *Gibrat's law in a medium-technology industry: Empirical evidence for Italy*, in E. Santarelli (ed.), Entrepreneurship, Growth, and Innovation: the Dynamics of Firms and Industries, New York, Springer, **TD No. 555 (June 2005).**

F. BUSETTI, S. FABIANI and A. HARVEY, *Convergence of prices and rates of inflation*, Oxford Bulletin of Economics and Statistics, v. 68, 1, pp. 863-878, **TD No. 575 (February 2006).**

M. CARUSO, *Stock market fluctuations and money demand in Italy, 1913 - 2003,* Economic Notes, v. 35, 1, pp. 1-47, **TD No. 576 (February 2006)**.

S. IRANZO, F. SCHIVARDI and E. TOSETTI, *Skill dispersion and productivity: An analysis with matched data,* CEPR Discussion Paper, 5539, **TD No. 577 (February 2006).**

M. BUGAMELLI and A. ROSOLIA, *Produttività e concorrenza estera*, Rivista di politica economica, 3, **TD No. 578 (February 2006).**

R. BRONZINI and G. DE BLASIO, *Evaluating the impact of investment incentives: The case of Italy's Law 488/92.* Journal of Urban Economics, vol. 60, n. 2, pag. 327-349, **TD No. 582 (March 2006).**

A. DI CESARE, *Do market-based indicators anticipate rating agencies? Evidence for international banks*, Economic Notes, v. 35, pp. 121-150, **TD No. 593 (May 2006).**

L. DEDOLA and S. NERI, *What does a technology shock do? A VAR analysis with model-based sign restrictions,* Journal of Monetary Economics, v. 54, 2, pp. 512 - 549, **TD No. 607 (December 2006)**.

R. GOLINELLI and S. MOMIGLIANO, *Real-time determinants of fiscal policies in the euro area*, Journal of Policy Modeling, v. 28, 9, pp. 943-64, **TD No. 609 (December 2006).**

P. ANGELINI, S. GERLACH, G. GRANDE, A. LEVY, F. PANETTA, R. PERLI,S. RAMASWAMY, M. SCATIGNA and P. YESIN, *The recent behaviour of financial market volatility*, BIS Papers, 29, **QEF No. 2 (August 2006)**.

2007

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Development Studies Working Papers 231, Centro Studi Luca d'Agliano, **TD No. 570 (January 2006)**.

F. LOTTI and J. MARCUCCI, *Revisiting the empirical evidence on firms' money demand*, Journal of Economics and Business, v. 59, 1, pp. 51-73, **TD No. 595 (May 2006)**.

L. MONTEFORTE, *Aggregation bias in macro models: Does it matter for the euro area?*, Economic Modelling, 24, pp. 236-261, **TD No. 534 (December 2004)**.

FORTHCOMING

P. ANGELINI, *Liquidity and announcement effects in the euro area*, Giornale degli economisti e annali di economia, **TD No. 451 (October 2002).**

S. MAGRI, *Italian households' debt: The participation to the debt market and the size of the loan*, Empirical Economics, **TD No. 454 (October 2002)**.

G. FERRERO, *Monetary policy, learning and the speed of convergence*, Journal of Economic Dynamics and Control, **TD No. 499 (June 2004).**

M. PAIELLA, *Does wealth affect consumption? Evidence for Italy*, Journal of Macroeconomics, v. 29, 1, **TD No. 510 (July 2004).**

A. ANZUINI and A. LEVY, *Monetary policy shocks in the new EU members: A VAR approach*, Applied Economics **TD No. 514 (July 2004)**.

S. MOMIGLIANO, J. Henry and P. Hernández de Cos, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, **TD No. 523 (October 2004).**

D. Jr. MARCHETTI and F. Nucci, *Pricing behavior and the response of hours to productivity shocks*, Journal of Money Credit and Banking, **TD No. 524 (December 2004).**

R. BRONZINI, FDI Inflows, *Agglomeration and host country firms' size: Evidence from Italy*, Regional Studies, **TD No. 526 (December 2004).**

A. NOBILI, *Assessing the predictive power of financial spreads in the euro area: does parameters instability matter?*, Empirical Economics, v. 31, 4, pp. , **TD No. 544 (February 2005)**.

P. ANGELINI and A. Generale, *On the evolution of firm size distributions*, American Economic Review, **TD No. 549 (June 2005).**

A. DALMAZZO and G. DE BLASIO, *Production and consumption externalities of human capital: An empirical study for Italy*, Journal of Population Economics, **TD No. 554 (June 2005).**

R. FELICI and M. PAGNINI,, *Distance, bank heterogeneity and entry in local banking markets*, The Journal of Industrial Economics, **TD No. 557 (June 2005).**

R. BRONZINI and G. DE BLASIO, *Una valutazione degli incentivi pubblici agli investimenti*, Rivista Italiana degli Economisti , **TD No. 582 (March 2006).**

P. CIPOLLONE and A. ROSOLIA, *Social interactions in high school: Lessons from an earthquake*, American Economic Review, **TD No. 596 (September 2006).**

F. BUSETTI and A. HARVEY, *Testing for trend*, Econometric Theory **TD No. 614 (February 2007).**