ISTITUTO DI STUDI E ANALISI ECONOMICA

Re-thinking the ISAE Consumer Survey Processing Procedure

by

Flora Fullone

ISAE, Piazza dell'Indipendenza, 4, 00185 Rome
e-mail: f.fullone@isae.it

and

Bianca Maria Martelli

ISAE, Piazza dell'Indipendenza, 4, 00185 Rome
e-mail: b.martelli @isae.it

Working paper n. 92
February 2008

The Series "*Documenti di Lavoro" of the Istituto di Studi e Analisi Economica* – Institute for Studies and Economic Analyses (ISAE) hosts the preliminary results of the research projects carried out within ISAE. The diffusion of the papers is subject to the favourable opinion of an anonymous referee, whom we would like to thank. The opinions expressed are merely the Authors' own and in no way involve the ISAE responsability.

The series is meant for experts and policy-makers with the aim of submitting proposals and raising suggestions and criticism.

La serie "Documenti di Lavoro" dell'Istituto di Studi e Analisi Economica ospita i risultati preliminari di ricerche predisposte all'interno dell'ISAE: La diffusione delle ricerche è autorizzata previo il parere favorevole di un anonimo esperto della materia che qui si ringrazia. Le opinioni espresse nei "Documenti di Lavoro" riflettono esclusivamente il pensiero degli autori e non impegnano la responsabilità dell'Ente.

La serie è destinata agli esperti e agli operatori di politica economica, al fine di formulare proposte e suscitare suggerimenti o critiche.

Stampato presso la sede dell'Istituto

ISAE - Piazza dell'Indipendenza, 4 – 00185 Roma.
Tel. +39-06444821; www.isae.it

ABSTRACT

The Consumer Sentiment Index (CSI), calculated from consumer confidence survey data, often represents one of the most important factors – though not the only one – underlying consumption decisions. However, at times, in short term cyclical analysis, the relationship between Italian consumption and the CSI entailed critical evaluations.

The main idea of this work is that the reason of those sceptical views might be due to the survey estimates, which are no longer efficient as they are stemming from an outdated estimation process.

While major revisions and recurrent updates featured the ISAE Consumer Survey sampling design and the collecting technique, the core of the EDP processing has substantially remained unchanged, mostly relying on an underlying well-established equal probability of selection method of the sample units, which definitely allowed for satisfactory results.

However, over time, new statistical and EDP tools became available, and in 2005 ISAE decided to carry out a complete revision of the consumer survey processing procedure in order to improve the quality of its estimates.

The interaction between sample design, data collection mode and non-response handling was examined from a statistical point of view over an eleven-year period, as these factors may affect sample efficiency. For calculating sample weights and obtaining more reliable estimates, ISAE decided to adopt the calibration methodology, based on both the inclusion probability of selected respondents and the balancing of structural variables on the universe values.

From the EDP point of view, the new processing system utilises new technologies and data maintenance support. The renewed procedure allows wider aggregation capabilities, offering new sectoral/regional breakdowns. Finally, the new microdata database, which is part of the complete EDP project, also enables to carry out microdata analysis.

Outcomes seem encouraging inasmuch they confirm the survey quality. Moreover, they do not support the above-mentioned critical viewpoints. In fact, the differences between calibrated and unweighted series are very small even they are not randomly distributed.

Keywords: Complex sample design, weighting, calibration, post-stratification, non-responses.

JEL codes: C42 (Survey Methods), C81 (Methodology for Collecting, Estimating, and Organizing Microeconomic Data).

REVISIONE DELLA PROCEDURA DI ELABORAZIONE DELL'INCHIESTA ISAE PRESSO I CONSUMATORI

SINTESI

L'indice di fiducia dei consumatori (CSI), calcolato dai dati dell'inchiesta presso i consumatori rappresenta spesso, anche se non esclusivamente, un importante fattore sottostante le decisioni di consumo. Talvolta, tuttavia, nell'analisi congiunturale, la relazione fra i consumi italiani e l'indice di fiducia non è apparsa così evidente ed univoca.

L'idea principale di questo lavoro è che gli stimatori provenienti dall'inchiesta abbiano perduto efficienza a causa di un processo di elaborazione divenuto obsoleto. Mentre il disegno di campionamento e le tecniche di raccolta dei dati sono state sottoposti a sostanziali revisioni e aggiornamenti periodici, il nucleo del processo di elaborazione infatti è rimasto sostanzialmente invariato basandosi sulla natura auto ponderante del campione, che garantiva risultati soddisfacenti in termini di efficienza. Nel 2005, dato che nel frattempo si erano resi disponibili nuovi strumenti informatici e tecniche statistiche, l'ISAE ha quindi deciso di intraprendere una profonda revisione dell'inchiesta per migliorare la qualità degli stimatori.

Nel lavoro sono esaminate accuratamente le interrelazioni fra il disegno di campionamento, le tecniche di raccolta dei dati e il trattamento delle mancate risposte per un periodo di undici anni, dato che queste componenti influenzano la qualità delle stime. Per calcolare i pesi e migliorare ulteriormente i risultati, l'ISAE ha deciso di adottare la tecnica della calibrazione, basata sia sulla probabilità di inclusione dei consumatori selezionati per il campione sia il bilanciamento delle variabili strutturali sulle corrispondenti dell'universo.

Da un punto di vista informatico, la nuova procedura utilizza nuove tecnologie anche per la gestione dei dati. La procedura permette maggiori possibilità di aggregazione offrendo nuovi dettagli settoriali e territoriali. Infine, il nuovo database dei microdati, che è parte dell'intero progetto informatico, rende possibile l'effettuazione di analisi microeconomiche.

I risultati del lavoro sembrano incoraggianti in quanto confermano la qualità dell'inchiesta. Inoltre tali risultati non offrono appiglio alle critiche sopracitate, infatti le differenze fra le serie calibrate e non calibrate sono molto piccole, anche se non casuali.

Parole Chiave: Disegno campionario complesso, ponderazione, calibrazione, post-stratificazione, mancate risposte.

Classificazione JEL: C42, C81.

CONTENTS

1 INTRODUCTION[1]

This paper presents the first outcomes of the renewed ISAE Consumer Survey processing procedure. A complete weighting system was introduced, which comprises both the first-order inclusion probabilities and the final post-stratification weights, calculated using the calibration techniques proposed by Särndal and developed by INSEE[2].

Time series were recalculated since 1995, when major innovations were introduced, namely when the survey turned into a consumer survey, implying a change in the sampling unit, but also when Computer-Assisted Telephone Interviewing (CATI) became the new data collection mode and regional information started to be collected.

A transition phase (Jan.1995 – Jan.1997) was necessary to gradually re-design a more efficient sample thanks to the adoption of CATI, which allowed to reduce the stages.

The re-thinking of the whole procedure led to focus on the influence of sample design and data collection techniques on the final estimates, and to better analyse the role of non-responses in the entire process and their effects on estimates.

The new procedure now meets integrity and safety requirements for the original data, performs estimations according to up-to-date statistical theory, and offers more flexibility in computing estimates with sectoral/regional breakdowns. An important by-product of the whole work is also the setting up of a new database for the original microdata, which assures strong control over time of source information, allows safer information storing, and is also the base for carrying out longitudinal analyses.

Findings are presented both for the entire time span analysed, and detailed for the initial transition phase, when the survey design might have affected the outcomes, as well as, for the intermediate period of relative economic stability and for the period from 2002 onwards (characterised by a dramatic fall in confidence), being for both the survey design unaltered. Some sectoral

---

1 The author wish to thank for the helpful suggestions the participants to the XXVIII Ciret Conference (Rome, September 2006) where a preliminary version of this paper has been presented. The opinions expressed in this paper are the authors' own and do not reflect those of ISAE.
Though the paper is a joint effort of both authors, for administrative reasons we remind that sec. 6 is to be attributed to Flora Fullone and sec.1, 2, 3, 4, 5,7, and 8 to Bianca Maria Martelli.

2 The authors wish to thank Philippe Scherrer, Jérôme Accardo and Pascal Ardilly for the methodological support in applying the calibration technique. The responsibility of the outcomes remains of course with the authors.

breakdowns are also considered (namely by occupation, age, and education), where the risk of sample biasedness is usually more likely to occur.

Section 2 briefly introduces the structure of the ISAE Consumer Survey, while Section 3 focuses on the sample design, showing how it could affect estimates, mainly regarding the interaction between strata, clusters and quota sampling. Also the 1997 sample updating was considered, because it could have represented a possible break in the resulting time series. Section 4 deals with the non-response features, focusing on how they are connected to both the sample design and the data collection method applied (CATI), and analyses their effects in calculating estimates. Section 5 considers the weighting features, examining the initial weights, which stem from the inclusion probability, as well as the final weights, which derive from the calibration technique applied in the first place. A brief, more technical description of the new procedure is then presented, including explanations for the choices made (Section 6). Section 7 provides comparisons between the “old” unweighted Consumer Sentiment Indicator (CSI) and the calibrated one, estimated with the new procedure, both calculated according to the Italian definition. Comparisons, focusing on differences, are also presented for the overall national CSI as well as for the breakdowns by occupation, age, and education, for which the introduction of calibrated weights seems to have produced major effects. Lastly, the paper also envisages some open questions and further research topics.

The adoption of a more reliable new processing procedure, which allows avoiding the risk of biased estimates, shows that there are no remarkable differences between the old and the newly calculated CSI. This outcome is encouraging with regard to both the quality of the selected sample design and the marginality of the biases introduced by the selection method constraints which were added. The preliminary results also support the view that in 1997 the sample design revision did not create any significant break in the time series.

2 THE ISAE CONSUMER SURVEY

Since the very beginning, within the European Harmonised Survey Programme launched in May 1972[3], ISAE (former ISCO) has been carrying out the Consumer Survey for Italy[4], firstly three times yearly and, since January 1982[5], on a monthly basis, considering a sample of about 2,000 respondents. Given its "harmonised" origin, the ISAE Survey fully complied with the Programme recommendations, with only minor additions connected to country specific features.

Regarding the questionnaire, the Survey comprises fifteen qualitative harmonised questions, characterised by three-to-five reply options regarding three main topics, such as opinions on the overall situation and on the household situation and plans to purchase durable goods, cars or homes. Only since 2000, ISAE has been expanding its Survey by adding some questions on "relative poverty", asking consumers about their main concerns over the future and about their major difficulties in affording everyday expenditures. Occasionally, further specific questions were added, related to topics of temporary relevance (such as the February 2002 changeover to the euro). Moreover, the ISAE survey includes a set of structural questions – useful for both processing and carrying out the breakdowns requested by the Commission – referring to the respondents as well as to the composition of households (namely gender, age occupation, and professional status of respondents; households' components, children, number of income perceivers, and overall household income[6]). Further information, like geographical features (region and size of the respondents' municipalities), is also available from the frame.

Originally, the data collecting mode was the face-to-face technique, while the processing procedure, due to the equal probability of selection method (EPSEM) of the sample units (see Section 5), simply focused on the calculation of the relative frequencies and on the related weighted balances. The ISAE Survey sample and processing procedure remained substantially unchanged

---

3 Decision by the Permanent Committee on September 15, 1970.

4 The ISAE Consumer Survey was started in 1973. In 1972 a pilot survey was carried out on 5,000 respondents. ISAE was one of the five countries (with Belgium, Germany, France, and the Netherlands) which originally joined the project.

5 Up to 1994, the survey comprised eleven monthly waves per year (excluding August). Subsequently, it was carried out also in August.

6 For sensitive (quantitative) questions, such as income or, to a lesser extent, age, since they often imply a high non-response rate, ISAE collects data by asking the interviewed persons not to give punctual information, but to choose between predefined bracket ranges.

over the years up to 1994, both in terms of sampling design and of processing technique.

However, the following two major innovations took place in 1995: a) the change in the data collecting mode, and b) the implementation of the Commission Recommendation to modify the survey from a "household" to a "consumer" one. With regard to a), the rapid improvement in computer aided tools for surveys led ISAE to adopt the Computer-Aided Telephone Interviewing (CATI) technique, replacing the face-to-face technique used until then[7]. CATI offers several positive advantages, ranging from lower data collecting costs (with respect to personal interviews), to fewer constrains in setting up the sample and (if well conducted) to a lower item non-response rate. The new technique might, however, affect response outcomes[8].

Referring to b), the 1995 Commission Recommendation invited to select as sampling unit a single full-aged person within the household, and not, as in the past, the whole household to which the interviewed person belongs[9]. For ISAE, this implied a deep revision of its sampling design (see Section 3), which had to be performed along with the change in the data collecting mode.

Even though no dramatic differences emerged from the comparison between the two data collecting modes[10], all the above-mentioned innovations prompted to consider the year 1995 as a possible break year for the consumer series, and to use this year as a starting point for the examined time span.

---

7 For a recent review of data collecting modes, see, e.g., Martelli (2005).

8 With the telephone technique, personal situation answers could be biased toward pessimism, whereas those for the overall situation could be biased towards optimism. However, it is worth remembering that in short-term surveys the crucial aspect is represented by the changes and not the absolute level of the phenomenon.

9 However, also in the "new" Consumer Survey several questions still investigate the consumer's household situation, namely: Q1 - Assessments on households' financial situation; Q2 - Expectations on households' financial situation; Q12 - Current households' financial situation.

10 See De Cindio (1995).

3 THE CONSUMER SAMPLE

Excluding 1995, the Commission Recommendation mainly regarded the timing and the question wording, while the Research Institutes were let free to select the sampling design, the data collection mode and the processing techniques to adopt, thus assuring the desired precision of estimates. In this Section we will describe the changes brought about in the ISAE sample since the Seventies.

3.1 The original sample design (1973-1994)

The original ISAE consumer sample adopted up to 1994, with the necessary yearly updates, was an EPSEM clustered (multi-staged) stratified random sample of about 2,090 units[11]. It was built proportionally to the households' universe, in three stages, stratified in the first stage by six geographical areas and seven sectors related to the demographic width of municipalities[12]. The randomly selected sample units varied in each survey wave, since new consumers were contacted every month.

Clustering or multi-stage sampling is adopted in survey sampling for practical reasons. Mainly, if the sampling frame units cover two or more survey units (e.g. consumers within households), clustering is the only practical way of selecting a sample of the required units (this was the case of the last sample revision - see Section 3.3 -), or if one has to divide the sample into manageable workloads for interviewers, while using the face-to-face data collecting mode (this was the original ISAE sample). The multi-stage feature, while shortening travelling time and consequently reducing costs, implies a larger sampling error than for unclustered samples, as only part of the universe is monitored.

On the contrary, *stratification,* that is the technique of subdividing the universe into several subsets according to variables related to - but not being themselves – the subject of investigation, reduces the sampling error. The identified strata should minimise, as far as possible, the variances within each stratum and maximize those between them, since they are selected in order to

[11] Even though the paper focuses on a later period (from 1995 onwards), the original sample design is here briefly recalled since its features also affect subsequent updates.

[12] The geographical areas are: 1) North–west, 2) North-centre, 3) North-east, 4) Centre, 5) South, and 6) Islands. The classes of municipalities according to the demographic width are: 1) up to 5,000 inhabitants, 2) from 5,001 up to 10,000, 3) from 10,001 up to 20,000, 4) from 20,001 up to 50,000, 5) from 50,001 up to 100,000, 6) from 100,001 up to 500,000, and 7) over 500,000.

comprise units as similar as possible to each other within - and different between - the strata. It is worth noticing that each stratum represents an independent sample on which further different selection techniques can be applied.

As clustering tends to reduce the precision of population estimates, while stratification acts in the opposite direction - though to a lesser extent -, the combined effect of applying both techniques generally causes a modest reduction in the precision of estimates.

In the original ISAE consumer sample, the first-stage primary sampling units (PSUs) were formed by the municipalities (184), selected within each stratum with probability proportional to size[13]. The municipalities with more than 500,000 inhabitants (six) were all included in the sample. Excluding the latter ones (always considered in the sample), the other municipalities were updated yearly with a rotation criterion every other month. Within the municipalities, the second-stage sampling units were represented by the electoral rolls of the related municipality.

In the third stage, households corresponding to voters, which were randomly selected from the electoral rolls within each stratum, made up the final sampling units. The head of the family was interviewed and he/she reported for the whole household. The number of households to be interviewed within each stratum was determined proportionally to the households' universe, so as to get an approximately constant sampling fraction and thus an EPSEM sample. This occurrence allowed avoiding further weighting when processing the results, since this kind of sample is representative of the universe. The final sample size amounted to 2,090 households (interviews), due to the necessity to round up to a multiple of ten the number of visits for each interviewer[14]. As the interviewers were charged to perform the desired number of interviews within each stratum, households refusing to participate were replaced, as far as possible, with similar ones belonging to the same electoral roll. Therefore, more properly, the sample also showed a substitution feature.

[13] The determination of the first stage size had to take into account both the need to include as many PSUs as possible and the quickly increasing costs of spreading face-to-face interviews over different municipalities.

[14] Each of them, in fact, performed ten, or a multiple of ten, personal visits.

3.2 The 1995-1997 transition sample design

ISAE decided to gradually improve its sampling design to minimise occurrences of dramatic breaks in the time series. When firstly passing to CATI and to the consumer as the sampling unit, in 1995 ISAE maintained the multi-stage structure of its sample[15]. The first stage (municipalities) remained unchanged, while, in order to implement the Commission Recommendation, the second and third stages were updated and a fourth one was added. Table 1 shows the 1995 first-stage universe (municipalities) and the related population stratification, while Table 2 lists the selected municipalities (PSUs).

The second-stage sampling units (electoral rolls) were replaced by “homogeneous micro-areas” from the telephonic archives (about 150,000 for the whole country) for which the correspondence with the Census sections was established[16].

The third stage still comprised households, this time selected among the telephone subscribers of the “homogeneous micro-areas” with a systematic selection technique. The total sample size summed up to exactly 2,000 units. Table 3 reports the 1995 desired interviews’ distribution for each stratum. However, following the Commission Recommendation, a major innovation regarding final sampling units was introduced, namely for the first time a single consumer[17], instead of the whole household to which he belonged, was selected.

In effect, a drawback of the CATI technique, at least for the Harmonised Survey purposes, is the fact that it is based on use of telephone registers as frame, which have a closer link with households than with consumers, leading to the introduction of a further stage in sample building and thus lowering sample efficiency: within the household (third stage) the selection of the consumer (fourth stage). The final size - still totalling 2,000 units - stemmed from a strata selection proportional to the population universe (rather than the households’

---

15 In June 1994, a pilot CATI Consumer Survey wave was conducted alongside the traditional face-to-face one. Comparisons were presented in De Cindio (1995).

16 The project, called "Geo-referencing System", is an exclusive property of the ATESIA Company, which materially carries out the telephone interviews. Each of these "micro-areas" included on average 130 households corresponding to about 120 telephonic addresses. They were available for municipalities with more than 30,000 inhabitants (about 50% of the consumers). In each of the remaining municipalities, the telephone contract holders were ordered by telephone numbers and grouped in units of about 120. In this way "homogeneous geographical areas” were created, similar to those related to the Census sections. In 1995, the penetration rate of telephone subscribers amounted to about 93% and remained constant up to the recent years, when the above-mentioned problems emerged.

17 The consumer is intended as a full-aged person belonging to the household corresponding to the selected telephone number and who contributes, also in non-monetary terms, to the family income.

universe), in order to maintain the constant sampling fraction and consequently the EPSEM structure of the sample. In this case too, the desired size was achieved through quota sampling: that is consumers refusing to collaborate were replaced with similar ones within each stratum.

Tab. 1 **1995 - Italian Municipalities**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001 - 10,000	10,001 - 20,000	20,001 - 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	1,338	94	44	31	8	1	2	1,518
	2.8%	1.1%	1.1%	1.7%	1.0%	0.2%	2.8%	10.6%
North-centre	1,167	224	96	43	12	3	1	1,546
	3.8%	2.0%	2.3%	2.%	1.4%	0.7%	2.3%	15.6%
North-east	987	292	131	45	10	14	-	1,479
	3.7%	3.6%	3.0%	2.2%	1.3%	4.5%	-	18.2%
Centre	651	162	96	66	20	6	1	1,002
	2.3%	2.0%	2.3%	3.6%	2.4%	1.9%	4.6%	19.2%
South	1,215	274	163	93	37	7	1	1,790
	4.3%	3.3%	4.0%	4.9%	4.3%	2.2%	1.8%	24.7%
Islands	505	126	68	49	12	5	1	766
	1.8%	1.6%	1.6%	2.5%	1.3%	1.8%	1.2%	11.8%
Total	5,865	1,171	598	326	99	36	6	8,101
	18.7%	14.3%	14.3%	17.1%	11.6%	11.3%	12.7%	100%

Source: ISTAT, Population Public Registry.
Note: Percentages refer to total population.

Tab. 2 **1995 - First-Stage Municipalities' Stratification and PSU's Sizes**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001 - 10,000	10,001 - 20,000	20,001 - 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	7	3	3	3	3	1	2	22
North-centre	8	5	4	5	3	2	1	28
North-east	8	7	5	4	2	12	-	38
Centre	5	4	5	7	4	5	1	31
South	9	6	7	9	6	5	1	43
Islands	4	3	3	5	2	4	1	22
Total	41	28	27	33	20	29	6	184

Source: ISAE elaborations on ISTAT data.

To improve the quality and to lower the non responses, CATI interviews are carried out in the evening hours (6-9 p.m.) of the first ten working days of each month. On average about fifteen, carefully trained, interviewers are committed to the survey. Once a year phone numbers are randomly selected within each stratum, five for each interview and for each month, that is 120,000 addresses. Remaining unused numbers selected for one month are never reused in subsequent waves.

Tab. 3 **1995 - Third Stage Interviews' Distribution**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001-10,000	10,001-20,000	20,001 - 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	66	24	24	32	22	8	74	250
	3.3%	1.2%	1.2%	1.6%	1.1%	0.4%	3.7%	12.5%
North-centre	76	50	44	46	30	16	62	324
	3.8%	2.5%	2.2%	2.3%	1.5%	0.8%	3.1%	16.2%
North-east	76	68	54	40	22	116	-	376
	3.8%	3.4%	2.7%	2.0%	1.1%	5.8%	-	18.8%
Centre	48	38	46	66	40	48	100	386
	2.4%	1.9%	2.3%	3.3%	2.0%	2.4%	5.0%	19.3%
South	86	60	66	84	60	50	38	444
	4.3%	3.0%	3.3%	4.2%	3.0%	2.5%	1.9%	22.2%
Islands	36	30	32	46	20	34	22	220
	1.8%	1.5%	1.6%	2.3%	1.0%	1.7%	1.1%	11.0%
Total	388	270	266	314	194	272	296	2,000
	19.4%	13.5%	13.3%	15.7%	9.7%	13.6%	14.8%	100%

Source : ISAE elaborations on ISTAT data.

Furthermore, while using CATI, a gender bias is often present, as women are easier to contact than men[18]. To avoid this distortion while passing to a consumer survey, an additional constrain was added in the last stage: the male/female percentage share of the interviews had to equal the one of the universe (51.5% and 48.5%, respectively). However, both substitution and gender quotas may negatively affect the survey (see Section 0).

A major drawback of CATI is bound to the frame. The telephone coverage has some limits due to "red lines", that is numbers that, for privacy reasons, do not appear in the telephone books, negatively affecting the frame quality. This problem, which was not relevant in the past, will however increase in the near future, as recent Italian legislation allows telephone contract holders to request not to be contacted. In addition, more recently, the rapid growth of mobile phones surely influenced people's decision of maintaining fixed lines, thus probably increasing frame representativeness problems. Furthermore, the recent liberalisation of the Italian telecommunication market led to the birth of several telephone companies, and now complete subscribers' lists for all contract holders are not fully available and reliable. In this paper we will not deal with these drawbacks, as up to now they have been barely relevant and were partially overcome by the Geo-referencing system and by giving the sample a structure proportional to the population universe. However, these problems will surely become an important topic of investigation in the near future.

[18] Without this restriction, practical experience showed that the ratio of female/male respondents is about 2:1.

3.3 The new sample design (1998-2006)

The adoption of the CATI technique had further relevant consequences for the sampling design. While with personal interviews a territorial subset of the universe (municipalities) needs to be selected, as the interviewer has to physically visit the selected households and it would be difficult (too expensive) to pick households randomly spread all over the country, with CATI this obstacle is eliminated, as phone contacts allow complete flexibility. From the sampling theory viewpoint, this factor allowed to exclude both the first-stage step of the former sampling design (the selection of a subset of municipalities within each stratum) and the second-stage step (the choice of "micro-areas" as proxies for the former Census sections), directly selecting the phone numbers within the strata, thus increasing sample efficiency. However, the additional stage of selecting consumers within households had to be maintained.

Tab. 4 **2005 ISAE Consumer Survey Universe (a)**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001-10,000	10,001-20,000	20,001 – 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	1,387,322	563,704	542,560	840,236	461,464	87,544	1,294,282	5,177,112
	2.9%	*1.2%*	*1.1%*	*1.7%*	*1.0%*	*0.2%*	*2.7%*	*10.7%*
North-centre	1,833,144	1,439,938	1,283,756	1,274,677	598,926	361,636	1,090,470	9,265,489
	3.8%	*3.0%*	*2.6%*	*2.6%*	*1.2%*	*0.7%*	*2.2%*	*16.3%*
North-east	1,698,954	1,758,486	1,852,352	1,122,949	681,870	2,186,133	-	9,300,744
	3.5%	*3.6%*	*3.8%*	*2.3%*	*1.4%*	*4.5%*	*0.0%*	*19.2%*
Centre	1,070,003	1,005,288	1,245,157	1,842,198	1,179,103	1,008,926	2,125,871	9,476,546
	2.2%	*2.1%*	*2.6%*	*3.8%*	*2.4%*	*2.1%*	*4.4%*	*19.5%*
South	1,884,477	1,440,842	1,854,007	2,317,911	1,998,780	987,711	779,358	11,263,086
	3.9%	*3.0%*	*3.8%*	*4.8%*	*4.1%*	*2.0%*	*1.6%*	*23.2%*
Islands	843,824	691,760	687,315	1,231,205	610,174	779,965	539,093	5,383,336
	1.7%	*1.4%*	*1.4%*	*2.5%*	*1.3%*	*1.6%*	*1.1%*	*11.1%*
Total	8,195,023	6,461,606	7,146,270	8,261,651	5,905,300	5,113,726	7,399,796	48,483,370
	16.9%	*13.3%*	*14.7%*	*17.0%*	*12.2%*	*10.5%*	*15.3%*	*100%*

Source: ISAE estimations on ISTAT data.

(a) Full-aged population.

According to these assumptions, a substantial revision of the sample was performed and applied since 1998, thus obtaining a stratified two-stage (telephone subscribers/consumers) random sample of exactly 2,000 units, which was applied, with recurrent updates, up to now. The reference universe in the new sampling design is represented by the full-aged (18 years and over) population, stemming from the Census Survey, which is updated yearly with

demographic statistics outcomes. The stratification is still the original one: population is grouped in 42 strata according to six geographical partitions and seven classes of demographic width of municipalities.

Tab. 5 **2005 - ISAE Consumer Survey Frame (Italian Households)**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001-10,000	10,001-20,000	20,001 – 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	725,021	280,685	273,587	427,993	243,628	44,854	732,821	2,728,589
	3.1%	*1.2%*	*1.2%*	*1.8%*	*1.1%*	*0.2%*	*3.1%*	*11.7%*
North-centre	872,930	673,909	610,655	621,847	310,654	198,547	667,114	3,955,656
	3.7%	*2.9%*	*2.6%*	*2.7%*	*1.3%*	*0.9%*	*2.9%*	*17.0%*
North-east	802,940	804,447	855,862	548,825	349,593	1,191,979	-	4,553,646
	3.4%	*3.5%*	*3.7%*	*2.4%*	*1.5%*	*5.1%*	-	*19.5%*
Centre	512,687	466,571	565,038	855,164	576,838	513,151	1,044,000	4,533,449
	2.2%	*2.0%*	*2.4%*	*3.7%*	*2.7%*	*2.2%*	*4.8%*	*19.5%*
South	895,595	639,658	800,244	989,941	866,547	458,766	371,157	5,021,908
	3.8%	*2.7%*	*3.4%*	*4.3%*	*3.7%*	*2.0%*	*1.6%*	*21.5%*
Islands	404,269	321,955	316,132	562,833	280,207	391,664	240,296	2,517,356
	1.7%	*1.4%*	*1.4%*	*2.4%*	*1.2%*	*1.7%*	*1.0%*	*10.8%*
Total	4,213,442	3,187,225	3,421,518	4,006,603	2,627,467	2,798,961	3,055,388	23,310,604
	18.1%	*13.7%*	*14.7%*	*17.2%*	*11.3%*	*12.0%*	*13.1%*	*100%*

Source: ISAE estimations on ISTAT data.

Tab. 6 **2005 - ISAE Consumer Survey Sample**

Demographic with / Partitions	Up to 5,000 inhabitants	5,001-10,000	10,001-20,000	20,001 – 50,000	50,001 - 100,000	100,001 - 500,000	500,001+	Total
North-west	57	23	22	35	19	4	53	214
	2.9%	*1.2%*	*1.1%*	*1.7%*	*1.0%*	*0.2%*	*2.7%*	*10.7%*
North-centre	76	59	53	53	25	15	45	325
	3.8%	*3.0%*	*2.6%*	*2.6%*	*1.2%*	*0.7%*	*2.2%*	*16.3%*
North-east	70	73	76	46	28	90	-	384
	3.5%	*3.6%*	*3.8%*	*2.3%*	*1.4%*	*4.5%*	-	*19.2%*
Centre	44	41	51	76	49	42	88	391
	2.2%	*2.1%*	*2.6%*	*3.8%*	*2.4%*	*2.1%*	*4.4%*	*19.5%*
South	78	59	76	96	82	41	32	465
	3.9%	*3.0%*	*3.8%*	*4.8%*	*4.1%*	*2.0%*	*1.6%*	*23.2%*
Islands	35	29	28	51	25	32	22	222
	1.7%	*1.4%*	*1.4%*	*2.5%*	*1.3%*	*1.6%*	*1.1%*	*11.1%*
Total	360	285	308	356	228	223	240	2,000
	18.0%	*14.2%*	*15.4%*	*17.8%*	*11.4%*	*11.2%*	*12.0%*	*100%*

Source: ISAE estimations on ISTAT data.

The frame is made up of the telephone subscribers list, ordered by region, administrative district, municipality, and zip code. The primary sampling unit is

the telephone subscriber, selected within each stratum with a systematic random selection technique. The second-stage sampling unit is the consumer, as already defined in 1995. Also gender quotas were maintained. The inclusion probabilities remained constant with respect to the population universe, and this occurrence allowed, at least to some extent, to continue using the pre-existing processing procedure. The universe of full-aged population, the frame and sample structure (updated up to 2005) are shown in Tables 4, 5 and 6. It is worth noting that the strata sample size exactly reflects the universe proportion, but not the households' one. This occurrence will affect the calculation of the inclusion probability, as shown in Section 0

This new sample design allows increasing the precision of (qualitative) estimates, thus reducing the theoretical percentage of a simple random sample size of 2,000 units from 2.2 per cent to 1.38 per cent (2005 average) for a self-weighting stratified sample.

4 NON-RESPONSE

The importance of non-response relies on the assumption that non-responding people may be different from those who agree to participate, that is they may have different opinions and attitudes or activities and behaviours with regard to the variables being studied. In this case, an important part of the sample would be missed, a fact which would hinder the possibility to make accurate generalizations to the population. This is the main reason why the response rate (RR) and its complementary non-response rate (NR) are direct indications of effectiveness and diligence in designing and implementing a survey. A low response rate is often a sign of poor survey design and it negatively affects the meaning of the findings.

The response rate is generally calculated as the ratio of all respondents to those who "choose not to respond". As this second group of people is often not univocally defined, a remarkable bias may arise for this indicator. Recent literature suggests to clearly report which kind of RR (NR) is being used and how it is calculated[19]. In its Standard Definitions referring to household surveys, AAPOR (2004) presents the following general definition of RR according to several sources: *the response rate is the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample.*

[19] See, e.g., the AAPOR recommendation (2004) and OECD (2005).

With the aim of minimising NR, literature advises to carry out all preliminary efforts in using updated and reliable frames as well as in implementing the data collection techniques. The ISAE CATI foresees up to seven contact attempts and personal call-backs, before giving up the interview.

For the purpose of this paper, we will deal with two kinds of NR, namely: a) Unit non-response (which refers to the impossibility of collecting the interview) and b) Item non-response (when only part of the interview is missing). In this second case, a further differentiation can be introduced depending on whether Item non-response refers to structural variables (like age, education, etc.) or to harmonised questions, since the distinction is important in the weighting step.

In this Section, both NRs are calculated in their unweighted form, also excluding the inclusion probability, in order to provide indications on the quality of the first step of the survey (data collection step).

Table 7 reports a six-month average of the interviews' structure of the ISAE consumer sample. As ISAE considers persons' willingness not to

Tab. 7 ISAE Consumer Survey Telephone Interviews' Structure

Outcomes (AAPOR Classification)	Number of telephone calls [a)]	Rates / Total (%)
I – (1.1) Completed interviews	1,929	38.3
P – (1.2) Partial interviews [c)]	71	1.4
Eligible cases that are not interviewed (non-response)		
R – (2.10) Refusals	768	15.3
NC – (2.20) Unreachable	151	3.0
NC – (2.21) Automatic replier	96	1.9
Cases of unknown eligibility (non-contacts)		
UH – (3.12) Telephone busy	60	1.2
UH – (3.13) No answer	1,202	23.9
Cases that are not eligible		
(4.20) Fax	13	0.3
(4.30) Wrong number	104	2.1
(4.50) Out of target (others)	39	0.8
(4.30) Duplicate	2	0.0
(4.80) Exceeding quotas	596	11.8
Total	5,030	

a) Average January-June 2006. b) Break-offs are included. c) Only for harmonised questions.

communicate structural features as refusals[20], two major effects arise: namely, the refusal quotas increase, a fact which also eases the subsequent weighting procedure and restricts the Item NR to the harmonised – and less sensitive – questions, where non-response percentages are often negligible.

It is worth reminding that telephone numbers are selected from telephone directories, therefore some of the problems stemming from the random-digit dial technique (RDD) do not arise. Table 7 shows different kinds of NR rate calculations.

4.1 Some remarks on the ISAE sample design and on non-response

In setting up a sample design, one usually has to balance theory with practical constrains. These limitations may affect the RR.

The ISAE Consumer Survey maintained over time the feature of achieving the desired sample size through the substitution of non-respondents. In the CATI supported version, also a gender selection was added. Quota sampling assumes that survey participants have the same characteristics, attitudes, behaviors, etc. as non-participants and this occurrence is all but not true[21].

In fact, techniques of replacing non-respondent units with similar ones willing to respond within the same stratum more properly belong to the group of non-probability samples. These techniques, on the one hand, always allow achieving the desired sample size and avoiding the occurrence of unit non-responses, but, on the other, lead to loose the characteristic of pure random sampling. As the inclusion probability for them is unknown, it is theoretically impossible to assess the sampling error and to project the sample characteristics to the population.

In the ISAE survey, as the subject selection is performed according to random techniques (and not in an arbitrary way as it often happens with quota samples), the sample adopted is rather a substitution sample[22]. Furthermore, quotas by gender (as the unit selection is randomly performed) can be seen as a further stratification step within the partition/size strata.

---

[20] Non-response is allowed only for income. This variable, however, is not used for weighting purposes.

[21] More drastically, Dillman (1978), quoting Deming (1953), warns against the substitution sampling to reduce non-response: "Substitution does not help; it is only equivalent to building up the size of the initial sample, leaving the bias of non-response undiminished".

[22] Practical experience has shown that, even cautiously, the usual statistical methods can still be applied.

However, a unit non-response bias, though partially compensated by the accurate stratification applied, is definitely still present.

4.2 Unit non-response

The general definition of Unit NR could be the following: “the percentage ratio between the number of responses and the total size of the sample, under the assumption that a non-response (of whatever kind) is a refusal”. However, this definition does not adequately fit the ISAE Consumer Survey[23], and it could be interesting to analyse the frame underlying each survey wave. According to Table 8, RR ranges from 45%, in the most unfavourable case (where only completed interviews and all cases of unknown eligibility are considered), to 66%, considering all interviews (completed and partial) and thus excluding from the ratio the cases of unknown eligibility. It is difficult to assess “with certainty” whether these rates are large enough to be acceptable. However, ISAE can act only in the preliminary step, where most of the recommended efforts are made (use skilled interviewers, make more contacts, create interest, involve in the topics, etc.), while other tools are not applicable. ISAE cannot assess the characteristics of non-respondents who are randomly selected for only one wave, nor can ISAE “double” the sample, since the monthly frequency of the survey does not leave enough time. As the sample is not a panel, no longitudinal information is available for estimating non-respondents’ features.

Tab. 8 ISAE Consumer Survey Outcome Rates (a)

Response rate	%	Cooperation rate	%	Refusal rate	%	Contact rate	%
RR1 = I / (I+P+R+NC+UH)	45	COOP1=I / (I+P+R)	70	REF1= R / (I+P+R+NC+UH)	18	CON1 = (I+P+R) / (I+P+R+NC+UH)	65
RR2 = (I+P) / (I+P+R+NC	47	COOP2= (I+P) / (I+P+R)	72				
RR6 = (I+P) / (I+P+R+NC)	66			REF3 = R / (I+P+R+NC)	25	CON3 = (I+P+R) / (I+P+R+NC)	92

(a) Average January-June 2006. According to AAPOR (2004). Only the existing items are considered.

Due to the high frequency of the survey and the continuously changing sample units, follow-up surveys[24] cannot be used. According to the literature[25], a RR

[23] Simply applying this definition, the RR would equal 100%, which would be misleading.

[24] Johnson (1991) notes: "It may not be possible to create the ideal instrument or to complete enough follow-ups to obtain a perfect return rate".

[25] E.g., see Curtin (2000), also reported by McKenzie (2005).

falling in a range of about 60% to 65% is acceptable, but we have no means for assessing how to treat UH cases, so as to achieve a more precise result. In addition, the reported outcome rates show quite encouraging results, and indirectly assess the validity of the frame and the skill of interviewers.

4.3 Item non-response

The occurrence of non-responses mainly concerns the two groups of collected information: the "structural one" and the one referring to the harmonised (and ISAE) qualitative questions.

With reference to the former group, the data collection mode of the ISAE Consumer Survey considers as complete/partial carried out interviews only those where people fully respond to the structural set of questions. Consequently no need arises to perform adjustments for these variables, which are essential in the weighting step (Section 0). Only for income ISAE admits non-responses, which account for about 12% of the total 2,000 monthly interviews. For this variable, ISAE applies a deterministic approach by imputing the average value of respondents within the same stratum and with similar structural features[26] (age / household composition / occupation / gender / education).

Regarding the latter harmonised set, non-responses are included in the "I don't know" reply option, with no distinction from the effective no-opinion answers. The ratio of Item non-responses for assessing the interview as "partial" is arbitrary[27]. In the ISAE Consumer Survey, it is however a limited occurrence. In the 2005 average, referring to each single question, Item NR (calculated assuming the "I don't know" as non-response) summed up to 3.56% of the 2,000 sample units, whereas the NR regarding assessments amounted to 1.25%, and uncertainty relatively strongly affected those regarding forecasts, which summed up to 5.86%. Furthermore, in the 2005 average, the NR occurrences affected no more then 1-2 questions per interview, signalling, also in this case, the skill of interviewers who try to avoid these occurrences. Table 7 reports NR occurrences for completeness, but they actually have nearly no influence on outcomes. As in this case the information collected is qualitative, only stochastic methods would be suitable (like Hot-deck) to adjust for non-responses, thus

[26] We do not use, however, this variable in the weighting procedure, as reliable and updated universe data are not available.

[27] See AAPOR (2004). We found more easily operative indications for the distinction between Unit and Item NR. They show that over 30-40% of Item non-responses in the interviews should be considered as a Unit (complete) non-response, as missing data imputation would be problematic (Caron, 2005).

increasing the variance of estimates. Therefore ISAE has currently decided not to perform any action, while duly reporting the "I don't know / non-response" rate for each question when disseminating the results.

5 WEIGHTING

The importance of weighting relies on the fact that often a sample may misrepresent total population, and thus the resulting unweighted estimates are biased, if they are considered as estimates for the population[28].

Also the use of CATI as data collection mode may introduce some biases. While using CATI for carrying out a consumer (and not a household) survey, an underreporting bias mainly arises regarding extreme age classes (young and elderly people), where people are more difficult to contact and less confident in participating, and genders, as women are easier to contact than men. Also regarding occupation, working people are more eluding, because the time they spend home is limited.

Even if stratification still assured a relatively reliable sample, the aim of maintaining an EPSEM structure, however, weakened when the consumer became the sampling unit and CATI was introduced, leading ISAE to fully reconsider the weighting process.

Weights are mainly of the following three kinds: probability weights, related to the sample design, (Unit) non-response weights, and post-stratification (or calibration) weights. As the occurrence of applying non-response weights does not arise in the ISAE Consumers Survey, since it deals more properly about Item non-responses (as pointed out above), this Section focuses on the two remaining kinds of possible weights, namely probability weights (bounded to the inclusion probabilities) and post-stratification weights (related to the universe structure of the population). While probability (sample) weights do not usually change (being equal to the sample design), post-stratification weights may be altered to reflect new and better information becoming available.

[28] See, e.g. Afkami - ESDS (2005).

5.1 Probability weights

Non-equal probabilities of selection should be dealt with by applying weights proportional to the inverse of the probability of selection. When dealing with a two-stage stratified sample, a theoretical problem arises regarding the first-order inclusion probability, which cannot be simply calculated as the following ratio:

$$\pi = \frac{n}{N} \tag{1}$$

where *n* is the sample size and *N* is the universe (population) size; an EPSEM structure implies constant π in all strata.

We have to consider the composite inclusion probabilities[29] of both stages. As the first stage is now based on telephone subscribers, the first-stage inclusion probability should be properly calculated on households as their proxy[30]. Then, a second-stage inclusion probability related to the composition of households has to be taken into account to calculate the final composite probability, as the probability of selecting a consumer within households of varying size is not constant.

Let:

$$\pi_h(i) = \frac{n_h}{F_h} \tag{2}$$

being:

i =1,.. n_h the i^{th} first-stage sample unit within the h^{th} stratum;

h=1,..*H* the strata;

n_h the sample size within the stratum *h*;

N_h the population universe size within the stratum *h;*

F_h the households universe size within the stratum *h;*

This probability is constant for each sample unit *i* belonging to stratum *h*, but it varies between the strata. It is worth remembering that the stratum size n_h is proportional to the population (consumers) universe, being the total sample size *n*=2000. That is $n_h = 2000*N_h/N$. This implies that the stratum probability, while being constant with respect to the population, varies with respect to

---

29 We are dealing here more properly with the first order inclusion probabilities, referring to the composite probability of including a single sample unit of a multi stage sample. Second order inclusion probabilities, not referring to stages, regard the composite probability of including two sample units and are applied to evaluate variances.

30 The limits of this approximation were already pointed out. A possible alternative could be the use of dwelling statistics, see Cristine *et al*. (2003).

households, for which the stratum size F_h depends on the family composition. Wherever households are larger (/smaller) – hence F_h is smaller(/larger) – the probability becomes larger(/smaller).

The second-stage inclusion probability stems from the selected household size, precisely the probability of selecting the j^{th} full-aged component (consumer) of the i^{th} household within the h^{th} stratum is:

$$\pi_h(j \mid i) = \frac{1}{Fcomp_{i,h}} \tag{3}$$

This probability is simply calculated as the reciprocal of the household size. It stems from the structural question on the family composition directly collected in each survey wave. It ranges from 1 (households with only one full-aged consumer) to 1/4 (households with four or more full-aged consumers) and it is computed for each sample unit[31]. Thus this probability varies from month to month, as always different persons are contacted.

The final joint first-order probability of the j^{th} full aged consumer to be selected within the i^{th} household of the h^{th} stratum is then:

$$\pi_h(i, j) = \pi_h(i) * \pi_h(j \mid i) = \frac{n_h}{F_h * Fcomp_{i,h}} \tag{4}$$

and may differ from consumer to consumer within and between the strata[32].

The reciprocal of (4) stands for the starting sample weight to apply for the initial calibration processing and roughly indicates how many consumers each interview represents. The number of consumers varies from interview to interview[33]

$$d_{h,i,j} = 1/\pi_h(i, j) \tag{5}$$

---

[31] More properly, ISAE collects this information with more details, for up to 6 or more components. However, as the average Italian household size (of full-aged consumers) is just above 2 (more precisely, the average value in 2005 was 2.3) and is also showing a downward trend, we preferred to bind this distribution in order to achieve similar probability weights to apply in the initial calibration step.

[32] Theoretically, if $Fcomp_{i,h}$ were those of the universe, we would obtain:

$$\sum_{h=1}^{H} F_h * \sum_{i=1}^{h} Fcomp_{i,h} = N$$

going back to the usual denominator of the EPSEM sample.

[33] More properly, they depend on the sample design. In the ISAE Consumer Surveys, they range within 168 values stemming from all possible combinations of the 4 first-stage inclusion probabilities multiplied by the 42 strata households' second-stage inclusion probabilities.

5.2 Population weights

The use of these weights, also known as post-stratification or calibration weights, is recommended to achieve reliable results, avoiding biases in population estimates (stemming from non EPSEM sample surveys).

Post-stratification is one of the most common statistical techniques applied to obtain these weights. It assures that, within each post-stratum, the estimated population size equals a pre-specified value. Post-stratification assumes that *h* variables suitable for stratification (like age, gender, education) are not known until data are collected. Supposing the stratum size N_h can be derived from official statistics, the units can be classified into the strata, but only after the sample data are known.

Instead of using the sample estimate $\overline{y}$ we use $\overline{y_w} = \sum_h W_h \overline{y_h}$ where $\overline{y_h}$ is the mean of the sample units that fall in stratum *h* and $W_h = N_h / N$. This method is almost precise as the proportional stratified sampling provided that the sample is reasonably large (say >20) in each stratum and the effects of errors in weights W_h can be ignored. This method can also be applied to a sample that is already stratified by other variables, like for instance geographical regions, provided that the weights W_h are known separately within each region[34]. This method is more properly called *complete post-stratification*.

Also an alternate technique, the ratio estimation, which assures that for a given auxiliary variable (a proxy for the study variable) the estimated total equals a given value[35], requires that the auxiliary variables are known for each cell.

Whenever only the marginal distributions of the auxiliary variables are known, but the cross classification cell counts are lacking or are unreliable, or the size of the cells is extremely small, other approaches have to be applied, namely the class of *incomplete post-stratification* methods. The most utilised among them are the regression estimator, which introduces multiple post strata indicator variables, and the widely applied raking ratio (or iterative proportional fitting) method, according to which weights are computed so as to satisfy marginal constrains in a cross tabulation. The latter method can be extended to log linear models.

The *calibration* method can be regarded as a more general method where ratio, regression, and raking ratio estimators are seen as special cases[35]. The principle underlying the calibration method, proposed by Deville and Särndal

---

[34] Cochran (1977), Section 5A.9.

[35] Vanderhoeft (2001).

(1992) in their leading article, is to adjust samples through re-weighting individuals using auxiliary information stemming from a set of *X* (available) information referred to as calibration variables. The main principles of calibration are here synthetically reported following Sautory (1993 and 2003).

Given a population *U* of *N* of individuals from which a sample *s* of size *n* has been selected and being *Y* a variable of interest, for which we want to estimate the total in the population: $Y = \sum_{k \in U} y_k$, the usual Horvitz-Thompson estimator is: $\hat{Y}_{HT} = \sum_{k \in s} \frac{1}{\pi_k} y_k = \sum_{k \in s} d_k y_k$.

Let $X_1, ...X_j, ...X_J$ be the *J* the auxiliary variables, available in the sample, for which the population values are known: $X_j = \sum_{k \in U} x_{jk}$

The calibration weights w_k , as similar as possible to the original d_k weights, should verify the following calibration constrains:

$$\sum_{k \in s} w_k x_{j,k} = X_j \qquad \forall j = 1, ...J \tag{6}$$

on the basis of a selected distance function G as to:

$$Min_{w_k} \sum_{k \in s} d_k G\left(w_k / d_k\right) \tag{7}$$

The solution is given by: $w_k = d_k F\left(x_k' \lambda\right)$ where $x_k' = \left(x_{1,k} .....x_{J,k}\right)$, λ is a vector of *J* Lagrange multiplier associated with the constrains, and *F* is the calibration function defined as the reciprocal of the derivate of *G*. Vector λ results from the solution of the non linear system of *J* equations in *J* unknowns resulting from the calibration equation :

$$\sum_{k \in s} d_k F\left(x_k' \lambda\right) x_k = X \tag{8}$$

The calibrated estimator of the total for the variable of interest will then be:

$$\hat{Y}_w = \sum_{k \in s} w_k y_k \tag{9}$$

Four *G* distance functions are considered, namely linear, raking ratio, logit, and truncated linear.

6 PROCESSING THE RESULTS

ISAE faced several problems when it decided to re-think its consumer survey procedure. First of all, a preliminary but essential step was considered, namely the setting up of an efficient and flexible historical database for microdata. Subsequently, ISAE re-engineered the processing procedure as to integrate the methodological updates and introduce further requirements of flexibility in performing elaborations. The researchers' aim to easily have at their disposal results also at sectoral/local levels implied the setting up of a further front-end database. Security and integrity, of course, had to be guaranteed.

All those instances could be accomplished by switching from the old EDP language, used up to then, to consolidated, reliable fourth-generation software and to other updated EDP tools, now available at ISAE.

6.1 The microdata database

Security and integrity issues are fundamental in setting up an efficient database. In the past, the lack of adequate software tools exposed to this kind of risks.

The new project led to built up a database using the IBM DB2 system running in a UNIX/AIX environment. This tool allows contemporaneous sharing of microdata between users, thanks to the availability of concurrency control tools. The system is reliable, as back-up mechanisms are foreseen. It guarantees safety, since only authorised users are allowed to access the data and perform predetermined actions. It is efficient, because of its multi-access feature; in fact, several users can access the database simultaneously, receiving results in a reasonable time. Setting up a DB2 database assures independence from the software used for processing, in our case SAS (which also offers its database) and Java. Should ISAE in the future decide to change its processing software[36], the database would maintain its efficiency and validity.

The up-to-date consumer database now comprises all microdata since January 1995. The choice of the time span is based upon the following two considerations: in January 1995, the CATI data collection mode was started and, at the same time, the regional breakdown was performed and recorded (while previously it was only performed at partition level, due to the different

[36] For example, to open source and therefore costless languages (like R).

sample design)[37]. The setting up of the database implied a careful control of each of the 2,000 interviews for every monthly wave, because, in the original recording, fields and codes changed over time. This long and patient work also represented an opportunity to further thoroughly control the initial information. Furthermore, the database easily allowed controls on both structural and qualitative questions and was the basis for starting longitudinal microdata analysis[38].

6.2 The new procedure

The renewed procedure foresees all the steps which constitute the so-called back-end processing phase, namely:

1. performance of accurate coherency controls over the original input data;
2. reorganisation of the structural data information to meet the needs of weighting calculations and of desired detailed estimations;
3. data adjusting for non-response;
4. calibration, which implies computing the individual (probability and post-stratification) weights;
5. data estimation (both for the fifteen harmonised questions and for the national ones) according to the selected classification (structural) variables, applying individual weights and computing balances and confidence indicators;
6. setting up an aggregate database from which the final front-end application would get the requested series.

Finally, ISAE set up a front-end procedure, which allows a user-friendly and flexible examination of the estimated time series and, at the same time, secures against embargo breaking by unauthorised personnel.

Calibration and estimation steps, as they represent the core of the whole processing procedure, are hereafter presented in detail[39].

---

[37] ISAE is currently working to extend backwards the time span, at least up to the beginning of the nineties.

[38] Even though microdata stem from changing respondents, the Deaton's pseudo-panel techniques can be applied.

[39] Technical internal notes are set up for the other steps.

6.2.1 Calibration

ISAE firstly applied calibration by using the CALMAR software (*CALage sur MARges*)[40]. For an easier application, i.e. to shorten the iterative process of convergence, CALMAR requires as starting weights the effective probability weights, which ISAE calculated according to the procedure described in Section 5.1 (formula 5). Since the sample size amounts to 2,000 and the universe reaches about 48,000,000, each interview may represent from about 6,000 to 24,000 individuals, depending on how many consumers belong to the contacted households.

As variables liable for calibration, ISAE decided to consider, besides the regional ones, those that are usually more affected by the sampling design. More precisely, the selected calibration variables are the following:

- age (4 classes: up to 29 years, 30-49, 50-64, 65, and over);
- regional full-aged population (in 19 Administrative Regions: *Piemonte-Valle D'Aosta, Lombardia, Liguria, Trentino-Alto Adige, Veneto, Friuli, Emilia-Romagna, Marche, Toscana, Umbria, Lazio, Campania, Abruzzo, Molise, Puglia, Basilicata, Calabria, Sicilia, and Sardegna*);
- demographic width of municipalities (seven classes): 1) up to 5,000 inhabitants, 2) from 5,001 up to 10,000, 3) from 10,001 up to 20,000, 4) from 20,001 up to 50,000, 5) from 50,001 up to 100,000, 6) from 100,001 up to 500,000, and 7) over 500,000);
- occupation (4 categories: Independent workers, Dependent workers, Unemployed, and Inactive people);
- education (3 categories: Primary, Secondary, University and over).

We excluded gender (already considered in the quota interviews) and the stratification variable partition (already comprised in the more detailed region variables). The aim was to balance the need to have as many details as possible with a reasonable convergence throughout the whole considered time span.

These variables (total population, regional population, age, occupation) were calculated for each year of the considered period, or as often as available (for education, stemming from Census data, only two revisions were set). It was not possible to use income, because the universe data were not quickly

[40] It is freely downloadable from the INSEE website (www.insee.fr). Several EDP tools for weighting the survey results are freely available; among them is GENESEES, the one proposed by the Italian Statistical Institute (ISTAT, 2005). Our choice has fallen on CALMAR mainly for harmonisation reasons, as recommended by the European Commission (Mac Kenzie, 2005).

available and reliable. Only for the period 1995-1997, however, we had to change regions back to partitions, because for some months the multistage sample design did not comprise the municipalities for one region (*Basilicata*).

As the selected universe variables were not available at cell (stratum) level, ISAE more properly performed an incomplete post-stratification.

Also with the support of INSEE, among the four possible models (distance functions) offered by CALMAR for calculating weights, ISAE chose to apply the CALMAR logit model, as linear models permit negative values and the raking ratio allows fairly high extreme values. The logit method is based on a logistic distance function and needs to have the domain (lower and upper bounds) of the function defined. This setting is rather subjective, and ISAE faced the task of automating somehow this definition for an eleven year period, with twelve monthly waves for each year. ISAE then set up a procedure that first applied the raking ratio method for each month. The resulting bounds, identified by the raking ratio, were used as starting points to set the domain of the logit method, gradually decreasing the upper (increasing the lower) bound *t* times as to reach values for which convergence was no more reachable. Then, the bounds indentified in the *t-1* attempt were the desired ones.

In the ISAE experience (Table 9), the lower (L) and upper (U) margins emerging in the whole time span considered were, on average, L=0.41 and U =1.55. Namely, the most over-represented consumers contributed to the final outcome for only 41% of their initial weights, while those who were under-represented increased their weights about one and a half times.

Tab. 9 **ISAE CALMAR Logit Bounds**

Periods	Lower bound (average)	Upper bound (average)	Difference max	Difference min	Differences s.e.
1995-2006	0.41	1.55	1.95	0.75	0.18
1995-1997	0.32	1.50	1.70	0.90	0.16
1998-2001	0.51	1.62	1.40	0.85	0.16
2002-2006	0.38	1.51	1.95	0.75	0.20

Source: ISAE.

These bounds resulted rather constant in time: in the 1995-1997 transition period, their averages amounted to L=0.32 and U=1.50 respectively, as in the considered period the less efficient sample design might have been compensated by the collapsing of regional margins into partitions. In the period 2002-2006, the lower and upper bounds were worth 0.38 and 1.51, respectively,

since in this case the volatility of the differences was slightly higher than in the previous periods.

The final calibrated weights thus resulted ranging from about 2,500 to 37,000, being a composite result of the initial weights and of the post-stratification process. For each monthly wave, these weights, once associated with their respective interview, were divided by their average, thus summing up to *n*=2000 (the sample size).

Table 10 reports the average weighting structure of the sample for the whole period considered (1995-May 2006) and for three sub-periods of major

Tab. 10 ISAE Consumer Survey Weights (%)

Periods	1995-2006			1995-1997			1998-2001			2002-2006		
Sample Structure	W U	CW	*UW-CW*	UW	CW	*UW-CW*	UW	CW	*UW-CW*	UW	CW	UW-CW
Occupation												
Independent workers	9.10	12.58	*-3.48*	9.09	12.47	*-3.37*	10.10	12.50	*-2.40*	8.12	12.75	*-4.63*
Dependent workers	36.54	32.28	*4.25*	36.38	30.80	*5.58*	37.13	31.98	*5.15*	36.06	33.70	*2.36*
Unemployed	3.12	4.62	*-1.49*	4.12	5.30	*-1.18*	3.15	5.03	*-1.88*	2.35	3.70	*-1.35*
Inactives	51.24	50.52	*0.72*	50.41	51.43	*-1.03*	49.62	50.50	*-0.88*	53.47	49.85	*3.62*
Education												
Primary	54.27	61.96	*-7.69*	55.27	61.96	*-6.69*	53.95	61.96	*-8.01*	53.85	61.96	*-8.11*
Secondary	37.33	29.47	*7.86*	35.39	29.47	*5.92*	37.96	29.47	*8.49*	38.16	29.47	*8.69*
University	8.40	8.57	*-0.17*	9.34	8.57	*0.77*	8.09	8.57	*-0.48*	7.99	8.57	*-0.58*
Age												
Up to 29 y	14.94	18.12	*-3.18*	19.12	18.12	*1.00*	15.42	18.12	*-2.70*	11.32	18.12	*-6.80*
30-49 y	39.78	36.46	*3.32*	41.77	36.46	*5.31*	42.01	36.46	*5.55*	36.05	36.46	*-0.41*
50-64 y	26.42	22.46	*3.96*	23.23	22.46	*0.77*	25.62	22.46	*3.16*	29.61	22.46	*7.15*
≥ 65 y	18.87	22.96	*-4.09*	15.88	22.96	*-7.08*	16.95	22.96	*-6.01*	23.02	22.96	*0.06*

Source: ISAE Consumer Survey.

UW: unweighted sample structure; CW: calibrated (universe) weights sample structure.

UW-CW percentage points difference between starting sample structure and final calibrated sample structure.

interest: the initial period (1995-1997), the intermediate one (1998-2001) and the most recent period (2002-2006). Furthermore, the weighting structure is presented for three major breakdowns, namely occupation, education, and age. For each (sub)period, the average unweighted structure (UW), the final

calibrated one (CW) and the differences between them (UW-CW) are reported. Looking at the whole time span, the expected biases do emerge, that is sample under-representation for independent workers, low educated respondents (along with a complementary over-representation of middle educated ones), younger and elderly people. Focusing however at the three sub-periods, some structural changes appear evident, particularly in the most recent time span. In the period 2002-2006, the sample quota of Inactive people, still under-represented up to 2001, results more than three percentage points above the corresponding universe value. Also looking at the age breakdown, the sample structure shows remarkable changes in the last period under consideration. While the presence in the sample of younger people even more weakens also the intermediate age class (30-49 years) appears under-represented (being in the past over-represented instead). The 50-64 year class shows a difference of more than seven percentage points above the corresponding universe value, while the elderly people class completely recovered the sample under-representation of the former periods.

6.2.2 Estimation

The estimation step of the back-end process remained substantially unchanged with respect to the former procedure. Percentage frequencies are still calculated, but they now include the calibration weights. As usual, balances are processed as differences between favourable and unfavourable reply options, giving double weight to extreme options. The "I don't' know"/non-response option is shown as a percentage value for each question, but it is excluded from the balance processing, assuming at present an underlying random distribution.

Detailed breakdowns are now processed. They include – apart from the variables considered for the calibration step (age, education, occupation, and regions) – also gender, income (in quartiles), regions, and the stratification variables (partitions and demographic width[41]). The new procedure also calculates the cross-breakdowns of gender and partitions with all the other classification variables. All results are available in HTLM Tables.

[41] This variable was slightly collapsed to better meet research purposes: the classes were reduced from the original seven to the following four: small villages (up to 20,000 inhabitants,) small towns (from 20,001 to 100,000), medium towns (from 100,001 to 500,000), and big cities (100,001 and over).

7 SOME PRELIMINARY OUTCOMES

For the whole period and for the three sub-periods examined, this Section reports some comparisons between Calibrated and Unweighted CSIs, as well as between composing series. Furthermore, the differences between the indicators are examined and sectoral breakdowns of the results (by occupation, education, and age) are presented.

7.1 CSI and composing series comparisons

Firstly the comparisons between the CSI calculated according the new weighting procedure and the indicator calculated with the previous one (under the assumption of an EPSEM sample) in the ISAE definition[42] are analysed. Graphs showing the comparisons for the CSI composing series are reported in the Appendix. It is worth remembering that the new calibrated series are affected both by the introduction of the sampling probability and by post-stratification weights.

A quick graphical inspection (Graph 1) also shows that the two consumer sentiment indicators are very similar, thus recording very close monthly changes in both series. Furthermore, there is no evidence of shifts in peaks and troughs.

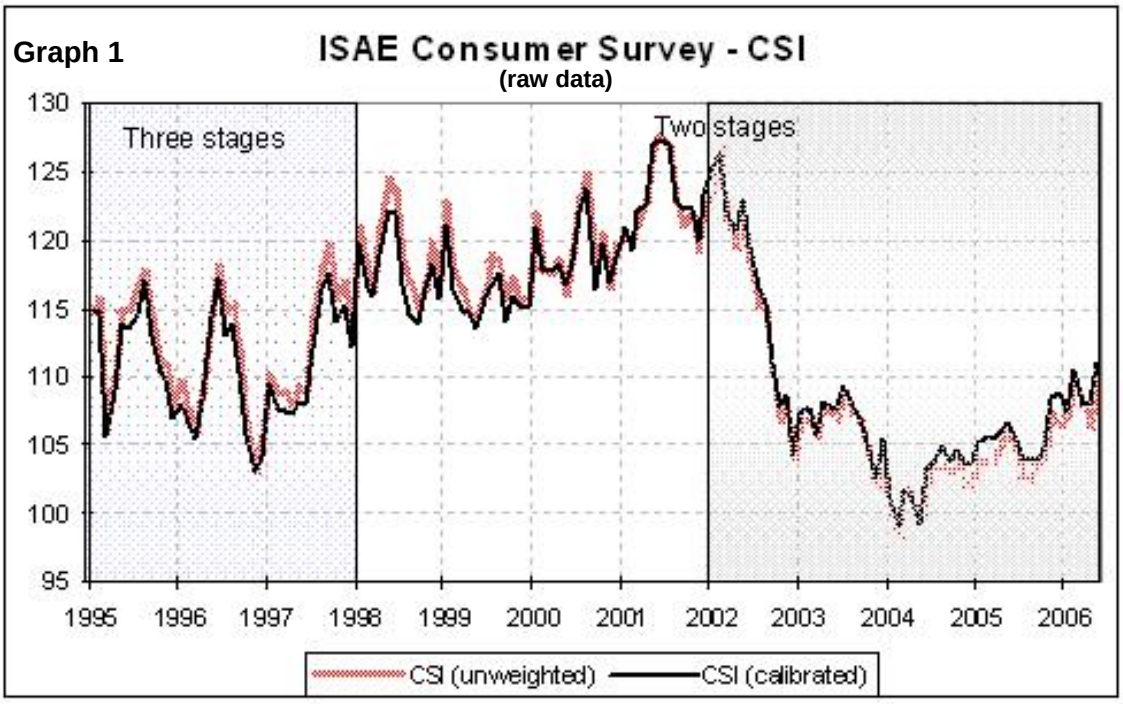
The calibrated CSI series (Tab. 11) showed, on average, for the entire 1995-2006 period slightly less volatility. Its standard deviation (10.999) was lower than the one for the unweighted indicator (11.646). According to the coefficient of variation, this occurrence featured also the first and second period under consideration. In recent years, instead, the volatility of the calibrated indicator became slightly higher. The coefficient of variation (2002-2006) was equal to -0.393, whereas it amounted to -0.369 for the unweighted indicator[43].

---

[42] The ISAE CSI is the result of the average of nine series, namely: Q1 - Assessments on households' financial situation; Q2 - Expectations on households' financial situation; Q3 - Assessments on the general economic situation; Q4 - Expectations on the general economic situation; Q7 - Expectations on unemployment (with inverted sign); Q8 -Assessments on purchases; Q10 - Assessments on saving; Q11 - Expectations on saving; Q12 - Current households' financial situation. The harmonised CSI is instead calculated as the average of the sole questions regarding expectations, namely: Q2, Q4, Q7, and Q11 (EC, 1997).

[43] The coefficient of variation is a dimensionless index of variability expressed as the ratio between standard deviation and mean. It loses however its significance when the average is approximately zero, as in the case of the entire 1995- 2006 period of Tab. 12.

Graph 1 ISAE Consumer Survey - CSI
(raw data)



This result might be seen as the consequence of the uncertainty which characterized consumers' behaviour in recent years and of the better capability of the calibrated CSI of capturing it. Finally, looking at time comparisons, the years between 1998 and 2001 surprisingly showed relatively less stability, as both calibrated and unweighted CSIs recorded the highest coefficients of variation (about two times larger than those of the other sub-periods).

The Pearson's R^2 correlation coefficient between the two indicators was very high, equalling 98.9% for the whole period. The possible 1998 break does not seem to have affected the CSI series, since the correlation coefficient was even higher (99.3% for the period 1995-1997); in this case, a deeper investigation on the effects of strata collapsing should be carried out. A relatively lower correlation, instead, featured the intermediate period (1998-2001), with R^2 = 96.7. In this case, the introduction of a weighting grid has possibly produced a stabilisation effect on the universe, which was not captured by the unweighted indicator. In the most recent period, however, the relationship between unweighted and weighted CSIs strengthened again (R^2= 99.5). Nonetheless, alternative non parametric indicators suggested a relatively less tight relationship between the two indicators, mainly in the intermediate period and, to a lesser extent, in the most recent one[44]. This occurrence might suggest that

[44] Parametric techniques often retain considerable power to detect differences/similarities even when the assumption of normality is violated. If this happens markedly, non parametric alternatives are more likely to detect those features.

statistical factors related to the sample structure did not cause the 2002 fall in confidence, as calibration became more effective (lower correlation) during the second sub-period along with a more detailed sample structure (both during the years of the cyclical recovery as well as after the sharp fall which occurred in third sub-period).

Tab. 11 ISAE CSI (Calibrated and Unweighted) Comparisons

	CSI calibrated			CSI unweighted			Correlation		
Periods	Mean	Standard deviation	Coefficient of variation	Mean	Standard deviation	Coefficient of variation	Pearson R^2	Kendall tau b	Hoeffding
1995-2006	-18.539	10.999	-0.593	-18.333	11.646	-0.635	98.9	90.9	76.4
1995-997	-21.390	6.496	-0.304	-19.693	6.811	-0.346	99.3	93.1	79.7
1998-2001	-8.353	5.730	-0.686	-7.339	5.294	-0.721	96.7	83.1	60.0
2002-2006	-25.828	10.138	-0.393	-27.367	10.107	-0.369	99.5	90.1	70.6

Computed on unbased CSIs.

Furthermore, looking at the CSI composing series (Graph A.1 and Table A.1 in the Appendix), the R^2 outcomes suggest a very close and widespread correspondence between the signals stemming from the balances of the calibrated and of the unweighted series. In fact, in the whole period examined, R^2 varied from a minimum of 97.1 for question Q2 (Expectations on households' financial situation) to 99.7 for question Q7 (Expectations on unemployment). The R^2 results strongly support the evidence that the shapes of the new weighted series were very similar to the previous unweighted ones. This outcome can be considered as an indirect support for the sample quality which correctly reflects the universe structure. This feature also holds true for the first breakdown period (1995-1997), when the sample was built in a less efficient way, as well as for the most recent period of higher uncertainty. Slightly lower correlations characterised, instead, the intermediate period.

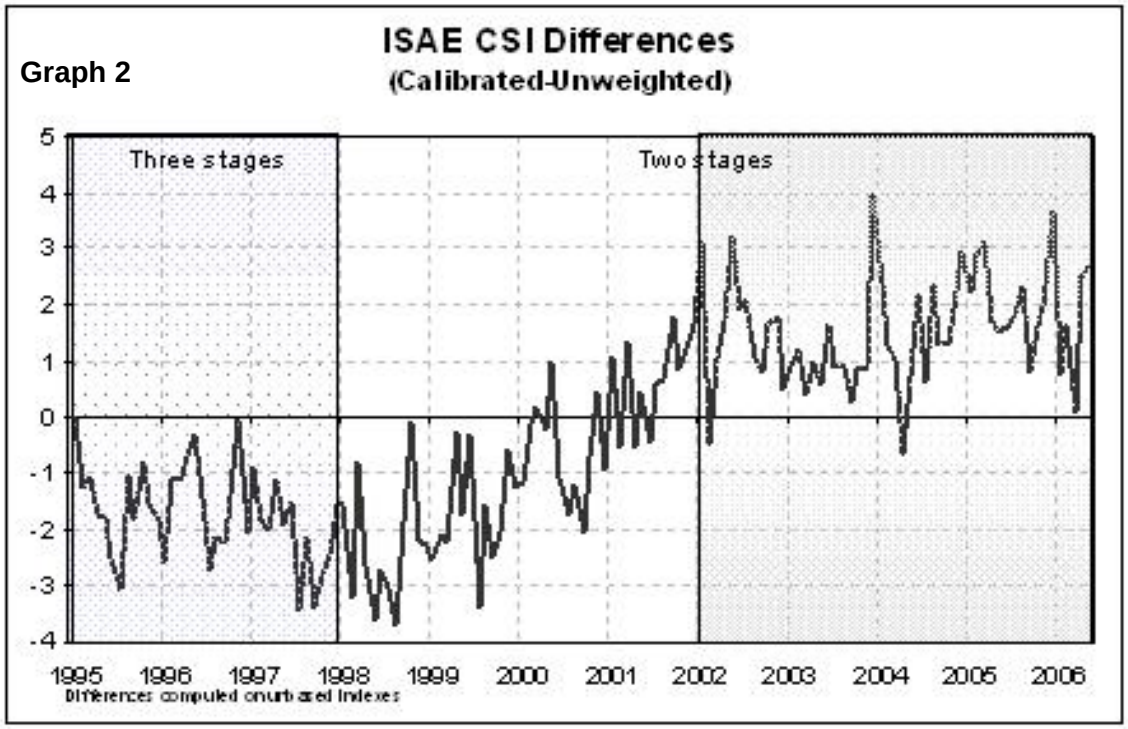
7.2 CSI differences

The introduction of calibration, however, had some statistical effects which appear evident when focusing on differences. Graph 2 shows that differences are not randomly spread throughout the whole considered period, as in the 1995-1997 time span calibration introduced a slight systematic reduction of

estimates, while from 2002 onwards an opposite feature was present. In the intermediate period, instead, differences were characterised by an upward trend. This outcome seems more difficult to explain as these years[45] fall within a cyclical recovery phase and the calibrated outcome seems to slightly underestimate the process in the first half, while overestimating it in the second one.

In Tab. 12 some descriptive statistics of the CSI differences are presented. Referring to the whole period, the differences show approximately a null average and skewness, suggesting a symmetric distribution. However, the negative value of Kurtosis may indicate a "fat tail" risk of deviation from normal distribution. This occurrence is also present for the three sub-periods analysed, though to a lesser extent for the first and the third ones. The three sub-periods show instead different features for their means, which are negative up to 2001 and turn positive in the last period. Variability instead seems to be higher for the intermediate period according to the coefficient of variation.

Graph 2



Through the *t* test analysis, the influence of the introduced weighting system is evident when focusing on sub-periods. We tested the null hypothesis whereby the differences between the calibrated and the unweighted series are negligible, that is H0: mean (calibrated-unweighted) =0. While for the whole

[45] More exactly up to the end of 2000 (See ISAE 2006).

period the CSI associated *p*-value does not support any evidence for rejecting the null hypothesis, for all the three sub-periods considered the differences introduced by calibration seem to be relevant. The p-values are all markedly lower than 0.05, which suggests rejecting the hypothesis of irrelevance of differences at a 5% confidence level and assessing that calibration does influence the results. The non parametric alternative given by the Wilcoxon signed rank test (which does not require assumptions about the form of the distribution) produces similar outcomes.

Tab. 12 ISAE Consumer Survey CSI Differences

Periods	N	Mean	Standard deviation	Coefficient of variation	Skewness	Kurtosis	t-Student test	p > | t |	S-Signed Rank test	p > | s |
1995-2006	137	-0.2059	1.8255	-8.866	0.0756	-0.9331	-1.3203	0.1890	-618	0.1708
1995-1997	36	-1.6971	0.8622	-0.508	-0.1544	-0.3058	-11.8117	<.0001	-315	<.0001
1998-2001	48	-1.0142	1.4744	-1.454	0.0691	-0.9051	-4.7658	<.0001	-368	<.0001
2002-2006	53	1.5390	1.0025	0.651	0.1711	-0.0256	11.1762	<.0001	676.5	<.0001

Computed on unbased CSIs.

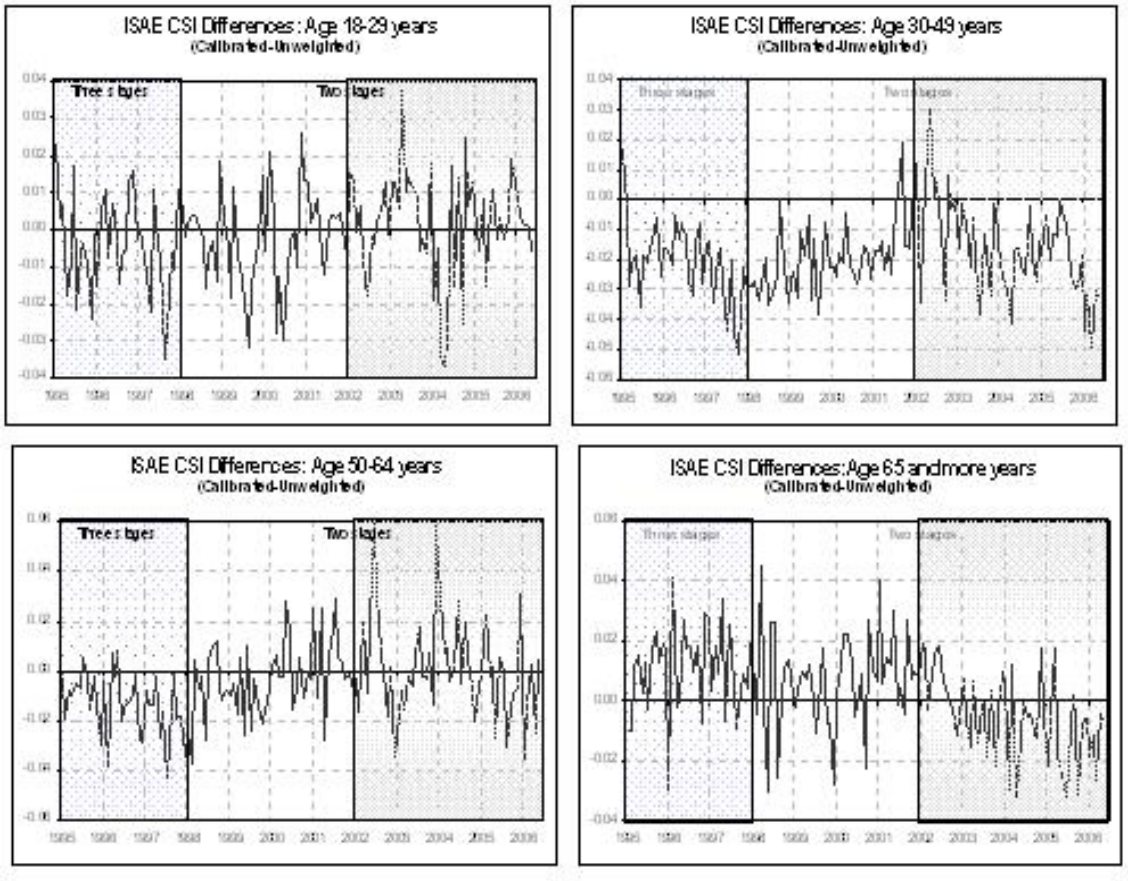
Examining the composing series (Table A.2 in Appendix), for the whole period considered the *t* test supports the evidence of refusing the null hypothesis (with $\alpha = 5\%$) for all the questions based on assessments (Q1, Q3, Q8, Q10, and Q12) and for unemployment expectations (Q7). For the remaining questions based on expectations (that is Q2, Q4, and Q11), the *t* test does not provide sufficient evidence to reject the null hypothesis. Probably these outcomes are related to the better capability of respondents to give more reliable information on their present situation than on the future one. In the assessment case, the changes introduced by calibration are more effective. The significance of the *t* test even in the case of question Q7 could suggest that consumers are deeply concerned by labour market problems and thus respond with greater accuracy. Also the non-parametric rank test produces similar results. Considering the three sub-periods, a widespread increase in weighting significance (the refusal of null hypothesis shows a marked increase) is apparent. However, it is worth noticing that, in the most recent years, questions Q3 (Assessments on the general economic situation) and Q7 (Expectations on unemployment) show very high *t* probabilities. These occurrences might reflect major uncertainty featuring these topics in the recent past.

7.3 Sectoral breakdowns of CSI differences

Similar outcomes for the consumer survey sectoral breakdowns referring to occupation, age and education are presented, in order to investigate where changes introduced by calibration are more effective.

Graph 3 shows CSI differences for the four major kinds of respondents' occupation, in which the universe weights mostly differ from the percentages collected in the sample, whereas Independent workers and Unemployed are underrepresented in the sample. The graphical inspection shows that the joint calibration effects of the others variables act in opposite directions according to different occupational categories, mainly in the most recent period (from 2002 onwards). Differences calculated for Dependent and Independent workers show lower positive values and, above all, a decreasing trend. The higher pessimism showed by these categories is, however, counterbalanced in the aggregate outcomes (Graph 2) by opposite features, such as the positive values and trend showed by Inactive people (which account for about half of the sample) and, to a lesser extent, by Unemployed.

Graph 3 ISAE Consumer Survey CSI Differences: Breakdown by Occupation



For the whole 1995-2006 period, the CSI indicators (Tab. 13) still show widespread high R^2 mainly for Dependent workers (99.3) and relatively lower R^2 for Unemployed (96.6).

Considering the entire period, the *t* test results provide evidence to reject the null hypothesis at the 5% confidence level for all categories. During the first sub-period (1995-1998), correlations are lower with respect to the average, whereas the null hypothesis cannot be refused only for Unemployed.

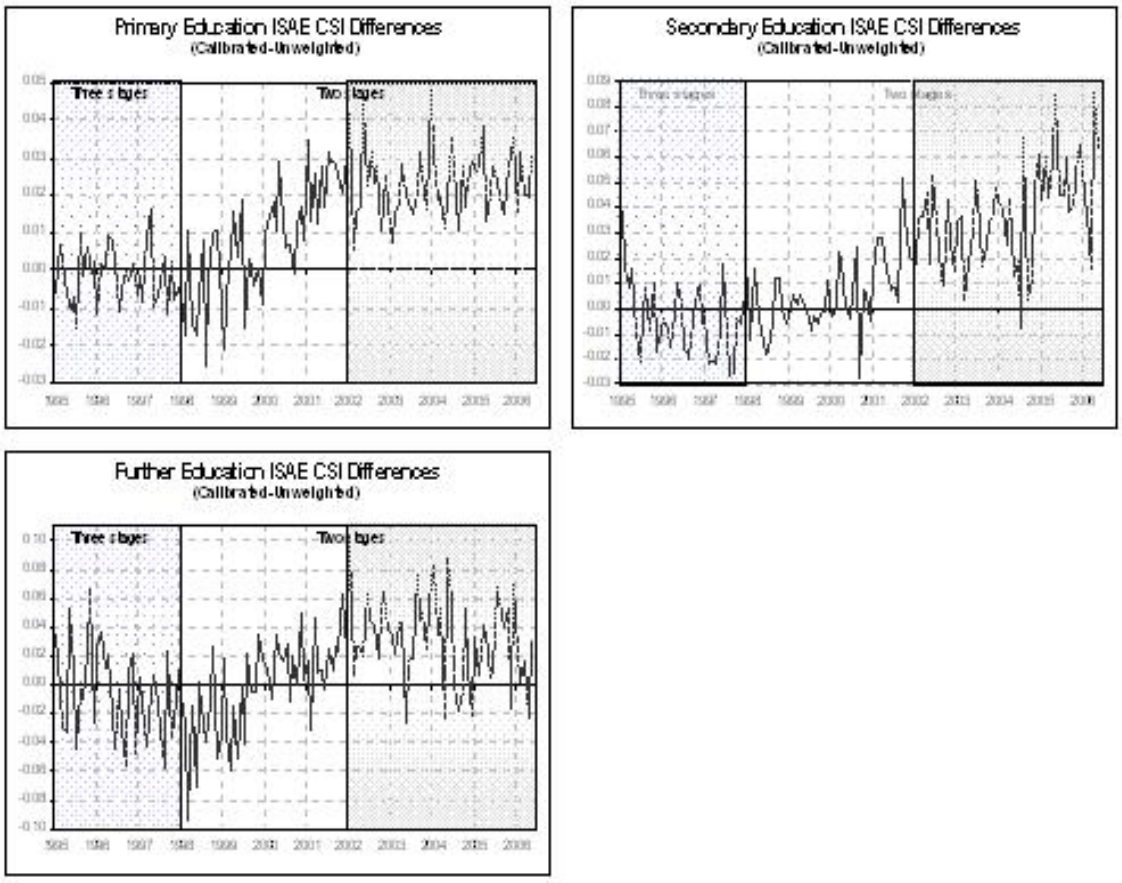
Tab. 13 ISAE CSI (Calibrated and Unweighted) Comparisons: Breakdown by Occupation

	Independent workers			Dependent workers		
Periods	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |
1995-2006	98.6	-5.7050	<.0001	99.3	-4.9032	<.0001
1995-1997	97.6	-5.8532	<.0001	98.8	-7.8459	<.0001
1998-2001	97.4	-3.1754	.0021	99.1	-2.9849	.0036
2002-2006	98.9	-1.7450	.0949	94.4	-0.4021	.6915
	Unemployed			Inactives		
Periods	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |
1995-2006	96.6	-5.4686	<.0001	98.9	2.2794	.0242
1995-1997	97.6	0.8021	.4307	98.8	-5.3341	<.0001
1998-2001	96.7	3.3557	.0012	97.2	-1.5115	.1343
2002-2006	94.4	4.6339	.0001	99.1	9.4970	<.0001

Focusing on the most recent period, the relationship between the two CSIs strengthens for Inactive people (R^2 equalling 99.1) and Independent workers (R^2=98.9), but eases slightly for Dependent workers and Unemployed (R^2 lowering in both cases to 94.4). According to the *t* test, however, differences do not seem relevant for both Dependent and Independent workers, whereas the refusal of the null hypothesis has to be accepted for Unemployed and Inactive people. This may signal the effectiveness of weighting.

The graphical inspection of the educational breakdown of the CSI differences (Graph 4) shows that the first two sub-periods have a shape similar to the aggregate one. In the most recent period, however, this holds true only for the low educated respondents. For secondary educated consumers, instead, calibration introduces not only a widespread increase in the estimates, since the differences are positive, but also a marked upward trend. On the opposite, for post-secondary educated consumers, in the period 2002-2006 the trend is decreasing and differences are, on average, null.

Graph 4 ISAE CSI (Calibrated and Unweighted) Comparisons: Breakdown by Education



The CSI indicators (Tab. 14) still show a high R^2 mainly for Primary (99.0) and Secondary (98.2) educated consumers and a relatively lower one for higher educated respondents (96.8). Considering the whole period, the *t* test results

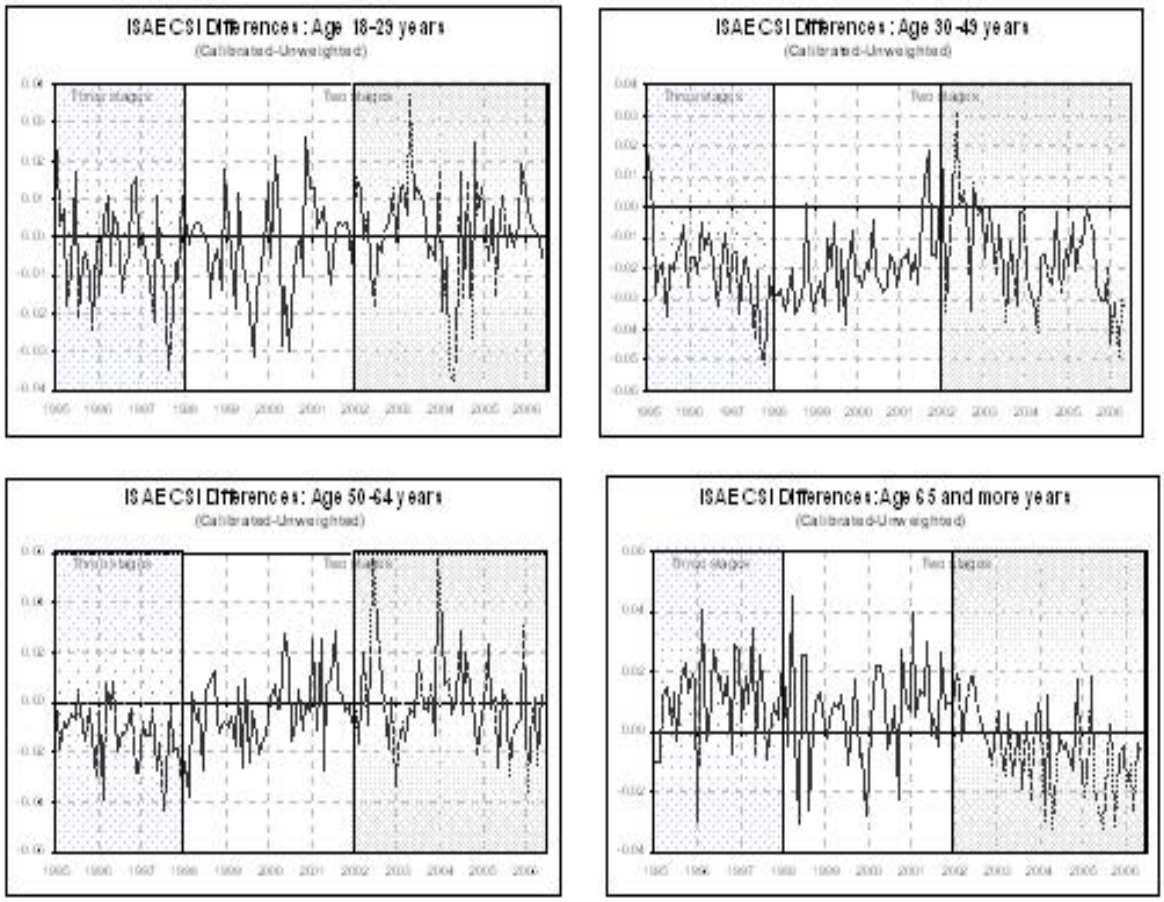
Tab. 14 ISAE CSI (Calibrated and Unweighted) Comparisons: Breakdown by Education

	Primary Education			Secondary Education			Further Education		
Periods	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |
1995 - 2006	99.0	8.7911	<.0001	98.2	7.2958	<.0001	96.8	3.3209	.0012
1995 - 1997	99.3	-1.2503	.2238	98.1	-1.7754	.0891	94.0	-1.1751	.2520
1998 - 2001	96.9	3.6355	.0005	97.0	2.7891	.0065	93.8	-0.3637	.7170
2002 - 2006	99.6	-1.7450	<.0001	98.0	13.5722	<.0001	97.0	7.7165	<.0001

provide evidence of effectiveness of calibration at the 5% confidence level for all categories. This occurrence is confirmed for the most recent sub-period (2002-2006), for part for the intermediate one (with the exception of higher educated respondents), but not for the first one, for which the *t* test results do not appear significant.

Finally, when analysing the breakdowns by age, Graph 5 shows different levels and shapes, depending on the considered age bracket.

Graph 5 ISAE CSI (Calibrated and Unweighted) Comparisons: Breakdown by Age



Specifically, the 30-49 year class shows negative differences for all periods and a downward trend in the most recent one, suggesting a lower and decreasing incidence in building the calibrated indicator of this subset of consumers. The same occurrence is present for the period 2002-2006 also for elderly people (65 years and over).

In the most recent period, all age classes have different shapes compared to the aggregate one, where negative average values and decreasing trends are evident, for all but the first class.

The *t*-test outcomes confirm the effectiveness of calibration for all consumers but the younger ones, and seem to be particularly significant also for the three sub-periods of the second class (30-49 years). The permanence of high R2 results (Tab. 15) is widely confirmed also for age breakdowns.

Tab. 15 ISAE CSI (Calibrated and Unweighted) Comparisons: Breakdown by Age

	18-29 years			30-49 years		
Periods	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |
1995-2006	99.3	-1.2848	.2010	99.3	-15.2260	<.0001
1995-1997	98.5	-1.8647	.0750	98.1	-9.4142	<.0001
1998-2001	98.2	-1.5986	.1135	98.3	-11.2458	<.0001
2002-2006	99.0	0.7734	.4475	99.0	-7.2194	<.0001
	50-64 years			65 years and over		
Periods	R^2	Paired t.test	p > | t |	R^2	Paired t.test	p > | t |
1995-2006	98.9	-2.9081	.0042	98.9	2.1660	.0321
1995-1997	98.9	-6.2649	<.0001	97.9	4.2978	.0003
1998-2001	97.4	-1.5655	.1211	96.5	3.0222	.0033
2002-2006	98.3	0.2901	.7745	99.2	-2.8349	.0096

8 FINAL REMARKS

The analysis presented in this paper shows that the calibrated series and indicators are very similar to the unweighted ones, thus assessing the quality of the underlying sample design. Differences are however present and systematic, confirming the effectiveness of the calibration system introduced. Actually, the weighted outcomes values are slightly lower at the beginning of the considered period and higher at the end. Furthermore, calibration affects confidence differently according to the various breakdowns.

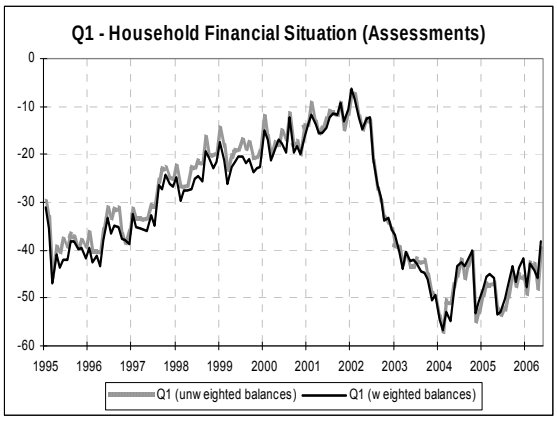
Even though the first outcomes are encouraging, additional work is needed to address some issues.

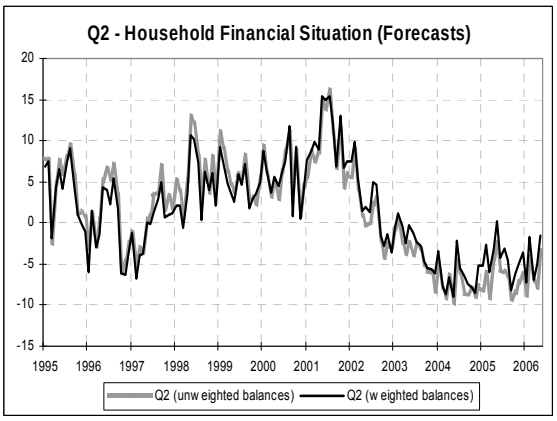
From a statistical point of view, it is necessary to further analyse the variability of estimates and to more deeply investigate the possible improvement to the predictive capability of the calibrated CSI with respect to consumption. From an EDP point of view, the fine-tuning of the front-end procedure is under way, and it is going to allow greater flexibility in processing breakdown estimates.

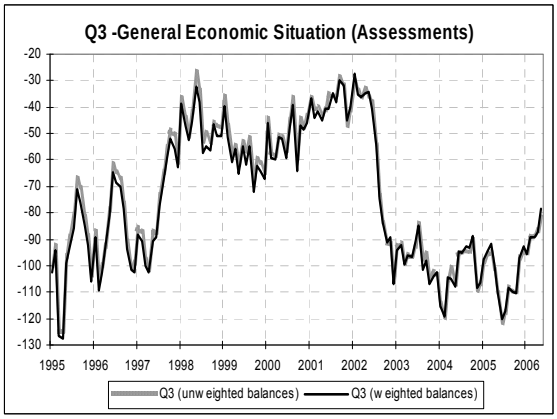
Finally, the following issues (more generally related to the consumer survey) should be further investigated: 1) how to maintain a reliable frame, given the many rapidly increasing concurrent problems related to “hot lines”, privacy constrains, mobile phones, etc.; 2) how to more deeply evaluate the impact of the item non-response rate of qualitative variables on estimates.

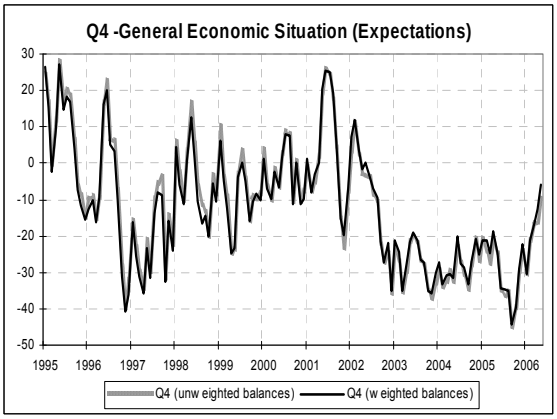
APPENDIX

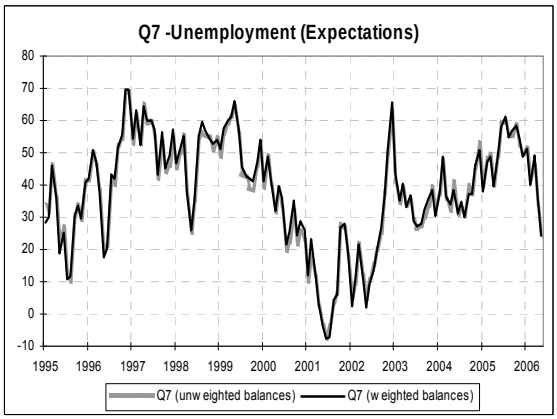
Graph A.1 ISAE Consumer Survey CSI Building Series (raw balances)

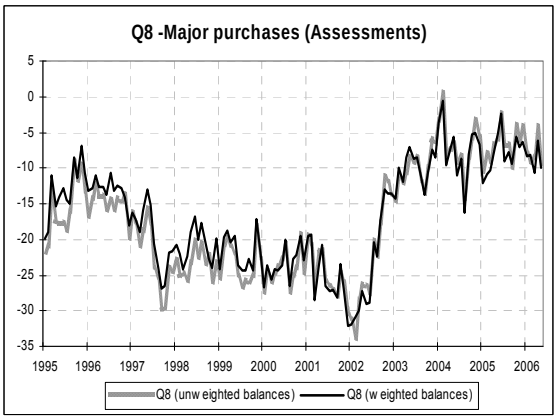




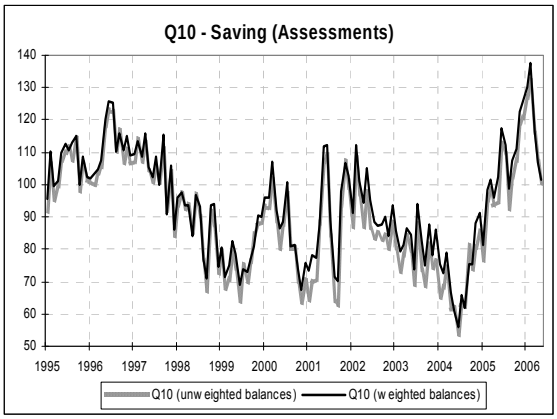


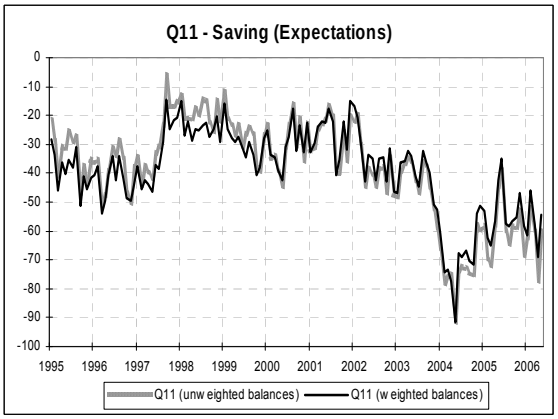


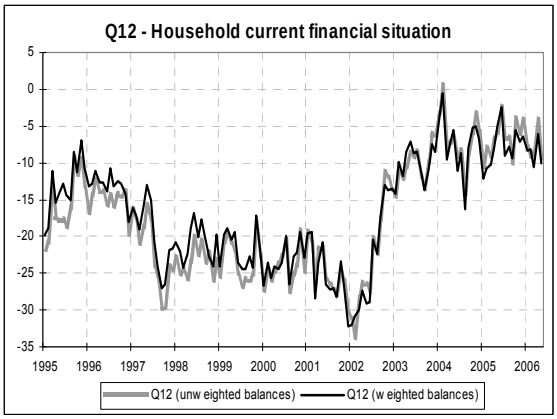




Graph A.1 (continued) ISAE Consumer Survey CSI Building Series (raw balances)







Tab. A.1 ISAE CSI Composing Series (Calibrated and Unweighted) Correlation

CSI Composing Series (Calibrated and Unweighted)	R^2			
	1995-2006(*)	1995-1997	1998-2001	2002-2006(*)
Q1 - Assessments on households' financial situation	99.2	97.9	97.2	99.3
Q2 - Expectations on households' financial situation	97.1	97.9	93.8	97.2
Q3 - Assessments on the general economic situation	99.7	99.8	98.5	99.8
Q4 - Expectations on the general economic situation	99.4	99.8	98.7	99.5
Q7- Expectations on unemployment	99.6	99.6	99.7	99.4
Q8 - Assessments on purchases	99.3	99.0	96.4	98.2
Q10 - Assessments on saving	98.8	97.9	97.8	98.8
Q11 - Expectations on saving	97.2	97.6	89.8	98.8
Q12 - Current households' financial situation	97.6	97.2	88.9	98.6

(*) Up to May.

Tab. A.2 ISAE Consumer Survey CSI's Composing Series Differences

		Mean	std	Skewness	Kurtosis	t-Student test	p > | t |	s-Signed Rank test	p > | s |
1995-2006	Q1 _ Ass. on households' financial sit.	-0.9430	1.7519	0.2737	-0.6379	-6.3004	<.0001	-2545.5	<.0001
	Q2 _ Exp. on households' financial sit.	0.1649	1.5513	0.0348	-0.7694	1.2441	.2156	514	.2710
	Q3 _ Ass. on general economic sit.	-1.5786	2.0321	0.2119	-0.5879	-9.0927	<.0001	-3376	<.0001
	Q4_ Exp. on general economic sit.	-0.2758	1.9611	-0.0516	-0.3765	-1.6464	.1020	-717.5	.1236
	Q7_ Expectations on unemployment	0.5762	1.6344	-0.3523	0.6790	4.1266	<.0001	1940.5	<.0001
	Q8 _ Assessments on purchases	-0.6340	3.1440	0.1406	-0.1038	-2.3603	.0197	-1171.5	.0113
	Q10_ Assessment on saving	3.3188	2.6695	0.0218	0.3791	14.5520	<.0001	4367.5	<.0001
	Q11_ Expectations on saving	-0.5251	4.5476	0.0431	-0.8253	-1.3515	.1788	-568.5	.2232
	Q12_ Current households' financial sit.	0.7150	1.7246	0.0703	-0.7119	4.8523	<.0001	1993	<.0001
1995-1997	Q1 _ Ass. on households' financial sit.	-2.0702	1.1142	-0.0437	-0.3252	-11.1486	<.0001	-331	<.0001
	Q2 _ Exp. on households' financial sit.	-1.0747	0.8757	-0.3499	-0.2631	-7.364	<.0001	-313	<.0001
	Q3 _ Ass. on general economic sit.	-2.8331	1.1966	-0.1686	0.1223	-14.2061	<.0001	-333	<.0001
	Q4_ Exp. on general economic sit.	-1.7902	1.2505	0.3549	2.1539	-8.5897	<.0001	-311	<.0001
	Q7_ Expectations on unemployment	0.3175	1.6931	-0.9252	2.5249	1.1251	.2682	101	.1137
	Q8 _ Assessments on purchases	-2.2533	2.1778	-0.4701	0.4530	-6.2081	<.0001	-290	<.0001
	Q10_ Assessment on saving	2.0075	1.7103	-0.0670	-0.4391	7.0427	<.0001	299	<.0001
	Q11_ Expectations on saving	-4.8281	2.5325	-0.0725	-0.4222	-11.4386	<.0001	-332	<.0001
	Q12_ Current households' fin. sit.	2.1478	1.1768	0.0725	-0.0506	10.9509	<.0001	331	<.0001

Source: ISAE Consumer Survey

Tab. A.2 (continued) ISAE Consumer Survey CSI's Composing Series Differences

		Mean	std	Skewness	Kurtosis	t-Student test	p > | t |	s-Signed Rank test	p > | s |
1998-2001	Q1 _ Ass. on households' financial sit.	-1.6521	1.3586	0.3886	-0.6149	-8.4247	<.0001	-512.5	<.0001
	Q2 _ Exp. on households' financial sit.	-0.3671	1.3674	0.1318	-0.5096	-1.8600	.0692	-177.5	.0681
	Q3 _ Ass. on general economic sit.	-2.4275	1.7827	0.4130	-0.3662	-9.4339	<.0001	-545	<.0001
	Q4_ Exp. on general economic sit.	-0.8679	1.8438	-0.1111	-0.0128	-3.2613	.0021	-292	.0019
	Q7_ Expectations on unemployment	1.3006	1.4072	-0.1314	1.3951	6.4036	<.0001	490.5	<.0001
	Q8 _ Assessments on purchases	-2.4921	2.3266	-0.1257	-0.0281	-7.4211	<.0001	-511	<.0001
	Q10_ Assessment on saving	3.0869	2.6833	0.1031	-0.1884	7.9701	<.0001	533	<.0001
	Q11_ Expectations on saving	-1.9008	3.5360	-0.1186	-0.5935	-3.7243	.0005	-319	.0006
	Q12_ Current households' fin.l sit.	1.0615	1.5059	0.0191	-0.4300	4.8833	<.0001	398.5	<.0001
2002-2006	Q1 _ Ass. on households' financial sit.	0.4649	1.4692	-0.4189	0.3515	2.3036	.0253	272	.0146
	Q2 _ Exp. on households' financial sit.	1.4887	0.9885	0.0231	-1.0132	10.9643	<.0001	702	<.0001
	Q3 _ Ass. on general economic sit.	0.0423	1.5910	-0.2674	0.5164	0.1933	.8474	29	.8001
	Q4_ Exp. on general economic sit.	1.2891	1.2351	0.5110	-0.7871	7.5979	<.0001	636	<.0001
	Q7_ Expectations on unemployment	0.0958	1.5814	0.0510	-0.7256	0.4412	.6609	33.5	.7699
	Q8 _ Assessments on purchases	2.1487	2.2076	0.4519	-0.0628	7.0859	<.0001	609	<.0001
	Q10_ Assessment on saving	4.4196	2.7722	-0.6505	2.0883	11.6062	<.0001	672.5	<.0001
	Q11_ Expectations on saving	3.6436	2.4886	0.2350	0.0605	10.6587	<.0001	688.5	<.0001
	Q12_ Current households' fin.l sit.	-0.5721	1.2558	0.4449	-0.1113	-3.3165	.0017	-353	.0012

Source: ISAE Consumer Survey.

REFERENCES

Afkhami, R., Crockett, A. (2005), “Weighting the Social Surveys”, ESDS Government, available at www.esds.ae.uk

American Association for Public Opinion Research (2004), “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys”, 3rd edition. Lenexa, Kansas, AAPOR.

American Association for Public Opinion Research AAPOR, “Best Practices for Surveys and Public Opinion Research”, available at: http://www.aapor.org

Caron, N. (2005), “La correction de la non réponse par re-pondération et par imputation“, INSEE Document n. M0502.

Christine, M., Davezies, L., Morando, M., Rousseau, S. (2003), “Rénovation de l'enquête CAMME: de l'échantillonnage au calcul de précision, en passant par l'allocation optimale entre les enquêteurs“, INSEE.

Cochran, W. G. (1977), “Sampling Techniques“, New Jersey, John Wiley & Sons.

De Cindio, G.(1995), “Personal versus Telephone Interviews: the ISCO Consumer Survey“ , Paper presented at 22th CIRET Conference, and Rassegna dei lavori dell'ISCO, Year XIII, n.1, ISCO.

European Commission (1997), “The Joint Harmonised EU Programme of Business and Consumer Surveys“, European Economy, Report and Studies n.6.

European Commission (2006), COS/BTS Survey Metadata, available at: http://ec.europa.eu/economy_finance/indicators/business_consumer_surveys/metadata_en.htm

D’Elia, E., Martelli, B.M. (2000), “Households' Incomes, Professional Status, and Inflation in Italy in the 1990s: Evidence from the ISAE Consumer Survey“, Proceedings of the XIV CIRET Conference, London, Ashgate.

Deville, J. C., Särndal C. E. (1992), “Calibration Estimators in Survey Sampling“, Journal of the American Statistical Association, Vol. 87, n.418, June, pp. 376-381.

Dillman D.A., Phelps G., Tortora R., Swift K., Kohrell J., and Berck J., (2001), “Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet“, Paper presented at 2001 AAPOR Annual Conference, Montreal, May, pp. 17-20.

Don A. Dillman site available at:
http://survey.sesrc.wsu.edu/dillman/papers.htm
Golinelli, R., Parigi, G. (2004), "Consumer Sentiment and Economic Activity: A Cross Country Comparison", Journal Business Cycle measurement and Analysis, Vol.1 n.2.
Golinelli, R., Parigi, G. (2005), "Le famiglie italiane e l'introduzione dell'Euro: storia di uno shock annunciato", Rivista di Politica Economica, Anno XXI, N. 2, pp. 201-226.
Groves, R.M., Fowler Jr F.J., Couper M.P., Lepkowski J.M., Singer E. and R. Tourangeau, (2004), "Survey Methodology", New Jersey, John Wiley & Sons.
ISAE (2006), "L'economia Italiana negli anni duemila: un complesso ciclo economico", Quarterly Report, July, pp. 60-64.
ISTAT (1991 and 2001), 13th and 14th Population and Housing Census.
ISTAT, Population public registry, several years.
ISTAT, Labour Force Survey, several years.
ISTAT (2005), GENESEES - GENEralised software for Sampling Estimates and Errors in Surveys, available at:
http://www.istat.it/strumenti/metodi/software/produzione_stime/genesees/index.html
Mac Kenzie, R. (2005), "Assessing and Minimizing the Impact of Non-Response of Survey Estimates". Analysis of Key Issues and Main Findings, OECD Taskforce on Improvement of Response Rate and Minimisation of Respondent Load, Paper presented at the Joint European Commission – OECD Workshop on International Development of Business and Consumer Tendency Surveys, Brussels, 14-15 November, available at: http://www.oecd.org/dataoecd/56/16/356340123.pdf
Malgarini, M. (2005), "Efficient Sample Design and Weighting Methods". Analysis of Key Issues and Recommendations, OECD Taskforce on Harmonisation on Survey Operation and Technical Design, Paper presented at the Joint European Commission – OECD Workshop on International Development of Business and Consumer Tendency Surveys, Brussels, 14-15 November, available at:
http://www.oecd.org/dataoecd/12/37/35493506.pdf
Malgarini, M., Margani P., and Martelli B.M. (2005), "Re-engineering the ISAE Manufacturing Survey", Journal of Business Cycle Measurement and Analysis, Vol. 2 n. 1, Paris.
Malgarini, M., Margani P (2005), "Psychology, Consumer Sentiment and Household Expenditures." A disaggregated Analysis Based on the ISAE Survey, ISAE WP n.58.

Martelli, B.M. (2005), “Relationship between Response Rate and Data Collection Methods“. Analysis of Key Issues and Main Findings, Paper presented at the Joint European Commission –OECD Workshop on International Development of Business and Consumer Tendency Surveys, Brussels, 14-15 November, available at: http://www.oecd.org/dataoecd/55/40/35558806.pdf

Martelli, B.M. (1998), “Le inchieste congiunturali dell’ISCO: aspetti metodologici“, Rassegna di Lavori dell'ISCO, Year XV, n.3, pp. 13-67.

Sautory, O. (1993), “La macro CALMAR - Redressement d'un échantillon par calage sur marges“, INSEE, Document n. F9310, Paris.

Sautory, O. (1993), CALMAR (CALage sur MARges) available at: http://www.insee.fr

Vanderhoeft, C. (2001), “Generalised Calibration at Statistics Belgium“, Statistic Belgium Working Paper n. 3, NIS.

Working Papers available:		
n. 40/04	R. BASILE S. DE NARDIS	Non linearità e dinamica della dimensione d'impresa in Italia
n. 41/04	G. BRUNO E. OTRANTO	Dating the Italian Business Cycle: a Comparison of Procedures
n. 42/04	C. PAPPALARDO G. PIRAS	Vector-auto-regression Approach to Forecast Italian Imports
n. 43/04	R. DE SANTIS	Has Trade Structure Any Importance in the Transmission of Currency Shocks? An Empirical Application for Central and Eastern European Acceding Countries to EU
n. 44/04	L. DE BENEDICTIS C. VICARELLI	Trade Potentials in Gravity Panel Data Models
n. 47/05	M. MALGARINI P. MARGANI B.M. MARTELLI	Re-engineering the ISAE manufacturing survey
n. 48/05	R. BASILE A. GIUNTA	Things change. Foreign market penetration and firms' behaviour in industrial districts: an empirical analysis
n. 49/05	C. CICCONI	Building smooth indicators nearly free of end-of-sample revisions
n. 50/05	T. CESARONI M. MALGARINI G. ROCCHETTI	L'inchiesta ISAE sugli investimenti delle imprese manifatturiere ed estrattive: aspetti metodologici e risultati
n. 51/05	G. ARBIA G. PIRAS	Convergence in per-capita GDP across European regions using panel data models extended to spatial autocorrelation effects
n. 52/05	L. DE BENEDICTIS R. DE SANTIS C. VICARELLI	Hub-and-Spoke or else? Free trade agreements in the "enlarged" European Union
n. 53/05	R. BASILE M. COSTANTINI S. DESTEFANIS	Unit root and cointegration tests for cross-sectionally correlated panels. Estimating regional production functions
n. 54/05	C. DE LUCIA M. MEACCI	Does job security matter for consumption? An analysis on Italian microdata

Working Papers available:		
n. 55/05	G. ARBIA R. BASILE G. PIRAS	Using Spatial Panel Data in Modelling Regional Growth and Convergence
n. 56/05	E. D'ELIA	Using the results of qualitative surveys in quantitative analysis
n. 57/05	D. ANTONUCCI A. GIRARDI	Structural changes and deviations from the PPP within the Euro Area
n. 58/05	M. MALGARINI P. MARGANI	Psychology, consumer sentiment and household expenditures: a disaggregated analysis
n. 59/05	P. MARGANI R. RICCIUTI	Equivalenza Ricardiana in economia aperta: un'analisi dinamica su dati panel
n. 60/05	M. BOSCHI A. GIRARDI	Euro Area inflation: long-run determinants and short-run dynamics
n. 61/05	M. BOVI	Book-Tax Gap. An Income Horse Race
n. 62/06	M. BOVI	The Cyclical Behavior of Shadow and Regular Employment
n. 63/06	G. BRUNO C. LUPI C. PAPPALARDO G. PIRAS	The cross-country effects of EU holidays on domestic GDP's
n. 64/06	M. COZZOLINO F. DI NICOLA M. RAITANO	Il futuro dei fondi pensione: opportunità e scelte sulla destinazione del TFR
n. 65/06	S. LEPROUX M. MALGARINI	Clima di fiducia e spesa delle famiglie in Italia: un'analisi disaggregata secondo il reddito degli intervistati
n. 66/06	M. BOVI	Consumers Sentiment and Cognitive Macroeconometrics Paradoxes and Explanations
n. 67/06	G. ROCCHETTI	Modelli di business nel mercato del software e partecipazione delle imprese italiane al fenomeno open source

Working Papers available:		
n. 68/06	L. CROSILLA	La stagionalità delle inchieste ISAE su imprese e consumatori: aspetti metodologici ed evidenza empirica
n. 69/06	C. OLDANI	Money demand & futures
n. 70/06	R. BASILE S. DE NARDIS A. GIRARDI	Pricing to market of italian exporting firms
n. 71/06	B.M. MARTELLI G. ROCCHETTII	The ISAE Market Services Survey: Methodological Upgrading, Survey Reliability
n. 72/06	M. FIORAMANTI	Predicting sovereign debt crises using artificial neural networks: a comparative approach
n. 73/06	S. ZECCHINI M. VENTURA	Public Credit Guarantees and SME Finance
n. 74/06	G. FERRI M. VENTURA	Macchinari del *made in Italy* e dinamiche dei distretti industriali
n. 75/07	R. BASILE	Intra-distribution dynamics of regional per-capita income in Europe: evidence from alternative conditional density estimators
n. 76/07	M. BOVI	National Accounts, Fiscal Rules and Fiscal Policy Mind the Hidden Gaps
n. 77/07	L. CROSILLA S. LEPROUX	Leading indicators on construction and retail trade sectors based on ISAE survey data
n. 78/07	R. CERQUETI M. COSTANTINI	Non parametric Fractional Cointegration Analysis
n. 79/07	R. DE SANTIS C. VICARELLI	The “deeper” and the “wider” EU strategies of trade integration
n. 80/07	S. de NARDIS R. DE SANTIS C. VICARELLI	The Euro’s Effects on Trade in a Dynamic Setting
n. 81/07	M. BOVI R. DELL’ANNO	The Changing Nature of the OECD Shadow Economy

Working Papers available:		
n. 82/07	C. DE LUCIA	Did the FED Inflate a Housing Price Bubble? A Cointegration Analysis between the 1980s and the 1990s
n. 83/07	T. CESARONI	Inspecting the cyclical properties of the Italian Manufacturing Business survey data
n. 84/07	M. MALGARINI	Inventories and business cycle volatility: an analysis based on ISAE survey data
n. 85/07	D. MARCHESI	The Rule Incentives that Rule Civil Justice
n. 86/07	M. COSTANTINI S. de NARDIS	Estimates of Structural Changes in the Wage Equation: Some Evidence for Italy
n. 87/07	R. BASILE M. MANTUANO	La concentrazione geografica dell'industria in Italia: 1971-2001
n. 88/07	S. de NARDIS R. DE SANTIS C. VICARELLI	The single currency's effects on Eurozone sectoral trade: winners and losers?
n. 89/07	B.M. MARTELLI G. ROCCHETTII	Cyclical features of the ISAE business services series
n. 90/08	M. MALGARINI	Quantitative inflation perceptions and expectations of Italian Consumers
n. 91/08	P. L. SCANDIZZO M. VENTURA	Contingent valuation of natural resources: a case study for Sicily