An application of calibration estimator in the Italian consumer survey[1]

Luciana Crosilla, Maria Rita Ippoliti, Raffaella Sonego [2]

Abstract

The European Union Harmonised Consumer Survey for Italy, carried out by the Italian National Institute of Statistics - Istat, relies on an estimation process not exactly reflecting the current sampling design (weights are not currently used). In order to deal with this problem, Istat decided to introduce the calibration estimator in data processing. Frequency, balance and Consumer Confidence Indicator series were rebuilt with the calibration estimator for the period 1998-2018. The outcomes of the renewed data processing show that the calibrated estimates are very similar to the unweighted ones. Differences are however present and statistically significant, proving the efficacy of the introduced calibration system. Finally, the calibrated and unweighted Confidence Indicator exhibits equal short-term volatility and similar features in tracking Italian private consumption.

Keywords: Auxiliary information, calibration, consumer confidence, weighting.

1 Some results excerpt from this article were presented at the 9th joint European Commission/OECD workshop on Business and Consumer surveys, Paris, November 2019. Authors wish to thank Claudia De Vitiis, Alessio Guandalini and Marco Dionisio Terribili for their precious help in the application of calibration estimator. A special thank to Giancarlo Bruno that encouraged us to write the paper. Although the paper reports the results of a joint research of all the authors, part of Section 2 and Appendix may be attributed to Maria Rita Ippoliti, part of Section 2 and Section 3 to Raffaella Sonego, Sections 1, 4, 5 and 6 to Luciana Crosilla.

2 Luciana Crosilla (lcrosilla@istat.it); Maria Rita Ippoliti (marippoliti@istat.it); Raffaella Sonego (rsonego@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics – Istat.

1. Introduction

The European Union Harmonised Consumer Survey for Italy (CS), currently carried out by the Italian National Institute of Statistics - Istat, in various occasions was modified with reference both to the survey sample and data collection mode (Martelli, 1998). By contrast, the core of data processing has substantially remained unchanged mostly relying on an underlying stratified sample with a constant sampling fraction (that is equal probability selection method of the sampling units) that gives a self-weighting sample (in data processing weights are not used) (Cochran, 1977). This weighting system, not exactly reflecting the current sampling design, could lead to a not perfect inference for the estimates. This problem may be partially solved introducing the real inclusion probabilities in the estimation process. Indeed, using only the real inclusion probabilities may lead to biased estimates due to the data collection mode. As a matter of fact, although stratification of the current sample design ensures a relatively reliable sample, the use of CATI (Computer Assisted Telephone Interviewing) technique as data collection mode may introduce some biases: a bias mainly may arise regarding extreme age classes (young and elderly people), gender (women are easier to be contacted than men), occupation (working people are more eluding) (Martelli, 2008); another bias is introduced using a frame affected by under-coverage problems as the list of subscribers to landline employed in CS. To deal with these biases affecting the estimates of CS, Istat decided to fully reconsider the estimation process.

As a first step, Consumer Confidence Indicator (CCI) and balances were estimated both with the Horwitz-Thompson estimator (HT) (where sample weights are the inverse of real inclusion probability of the units) and the calibration estimator (Deville and Särndal, 1992; Särndal, 2007) where the sampling weights are adjusted to derive estimates consistent with known population totals, preventing the bias due to data collection mode. Initially, the study adopted a calibration estimator that takes into account 10 auxiliary totals, such as population by gender, 4 geographical areas (North-East, North-West, Centre, South) and 4 age classes (18-29, 30-49, 50-64, 65 and more) known from the population register. Other calibration systems were applied[3]

3 In addition, calibration systems with 24 (population for geographical areas and age classes and for geographical areas and gender) and 17 auxiliary totals (population by gender, geographical areas, age classes, occupation

and outcomes were compared. The system with 10 constraints resulted as the most appropriate system producing less variable weights and more reliable calibration domains (De Vitiis *et al.*, 2015). In a second phase of the study, due to a revision of the questionnaire in 2018[4] and in order to improve the calibration process for age classes, seven age classes were considered in the calibration system increasing the number of total constraints from 10 to 13. On the basis of the test done, validity of the calibration system with thirteen constraints was also confirmed[5].

In a second step, outcomes (estimates for the total population and for sectoral breakdowns considered in dissemination of CS results by Istat, i.e. geographical areas, age classes, gender, education level and occupation) stemming from the implementation of the two estimators were compared. Estimates obtained with HT presented a sampling variance slightly slower but biased by age classes and gender. On the contrary, estimates stemming from calibration estimator were characterised by a slightly higher sampling variance and they were coherent with the population distribution by age classes and gender. This was interpreted as higher guarantee for the accuracy of the estimates. Generally, an increase in the variance of the estimator does not necessarily imply less accuracy, since the increase could be more than offset by a lesser bias of the estimator and, consequently, the total error could decrease. All these considerations led to choose the calibration estimator as the best estimator for the CS estimation process (Guandalini, 2019). Frequency and balance series were estimated again with the calibration estimator for the period 1998-2018. Consequently, the overall summary indicator of the results of the survey, the Consumer Confidence Indicator (CCI), was recalculated.

The aim of this work is to present the outcomes of the renewed CS data processing. The differences between Calibrated and Unweighted CCI as well as between its balance composing series are analysed. Furthermore, main

---

and education) were evaluated too. The derived calibration systems pointed out problems of convergence of the algorithm. These outcomes leaded to discard the above-mentioned auxiliary totals.

4 In the questionnaire, interviewees are asked the age class to which they belong to. Until March 2018, the age classes included in the questionnaire were: 18-20, 21-29, 30-39, 40-49, 50-59, 60-64, 65 and more; in April 2018 the age classes considered in the questionnaire were modified as follow: 18-20, 21-29, 30-39, 40-49, 50-59, 60-64, 65-70, 71-75, 76 and more.

5 More specifically, in rebuilding the series with the calibration estimator, until March 2018 four age classes were considered (18-29, 30-49, 50-64, 65 and more) and starting from April 2018, seven age classes were adopted (18-29, 30-39, 40-49, 50-59, 60-64, 65-70, 71 and more) as constraints in the calibration system (constraints for gender and geographical areas were unchanged).

sectoral breakdowns of CCI (by occupation, education, and age) are presented. The performance of new CCI in tracking Italian private consumption growth is compared with that of the unweighted CCI. In terms of methodology, the performing relies on correlation analysis, in-sample and out-of-sample tests, ability to track directional change and a volatility analysis (European Commission, 2018). Including a calibration system in the estimation process of CS diminishes also the bias of the estimates due to the under-coverage of the frame used for the selection of the households that is the list of subscribers to landline. Nevertheless, the quality of the frame is however an important requirement of the survey: here we want to underline the difficulty in currently maintaining a reliable frame given the increasing of mobile phones, privacy constraints etc. This work does not deal with this aspect: the complexity of the problem deserves a deepened analysis in another paper.

The paper goes as follows: in Section 2 a brief description of Italian consumer survey is provided. Section 3 shows the calibration estimator used in data processing and Section 4 describes and remarks some main results. In Section 5 performance of calibrated CCI in tracking private consumption was evaluated and compared with that of the unweighted CCI. Finally, concluding remarks are in Section 6.

2. The Italian Consumer Survey

The Consumer Survey started in Italy in 1973[6] within the Joint Harmonised European Union Programme of business and consumer surveys. The survey was carried out by the Institute for short term studies (ISCO) until 1998; starting from January 1999 the Institute for studies and economic analyses (ISAE) replaced ISCO in the management of the survey. Finally, Istat has conducted the survey since 2011. Up to 1981, the survey was carried out on a quarterly basis and subsequently on a monthly basis. The aim of the survey, which remained unchanged over time, is to collect information on households' spending and savings intentions and to assess their perception of the factors influencing these decisions.

6 More specifically, in 1972 a pilot survey was carried out on 5,000 respondents.

Until 1994, the sample was a clustered stratified random sample proportionally built to the households' universe. It was based on a multi-stage sampling design, stratified in the first stage by six geographical areas and seven sectors related to the demographic width of municipalities[7]. The sample size amounted to 2,090 households (interviews); the data collection mode was the face-to-face technique. In 1995 one important change was introduced in the sampling design: following the European Commission recommendation, the consumer[8] replaced the household as the sampling unit; at the same time the CATI technique was adopted as data collection mode. The sample maintained the multistage structure and another stage was added in order to select the consumer. The impact of CATI technique on the sampling design was significant and in the years 1995-1997 sampling design underwent revision. The revised sampling design was adopted starting from 1998 and it is currently applied. It is stratified into two stages, by geographical area and size of the municipality of residence. The stratification variables are still the original one: six geographical partitions (North west, North centre, North east, Centre, South and Islands) and seven classes of demographic width of municipalities (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+).

Allocation of the units in the strata is proportional to the universe of the adult resident population of Italy (for the number of units within strata see Table 1). The list used for the extraction of the names is made of the list of subscribers to landline; the first stage unit is the subscriber, while the second stage unit is composed of the consumer. The selection technique of the units is systematic in the first stage, and by quotas according to gender in the second

7 The first-stage primary sampling units were formed by the municipalities, selected within each stratum with probability proportional to the size. The determination of the first stage size had to take into account both the need to include as many primary sample units as possible and the quickly increasing costs of spreading face-to-face interviews over different municipalities. The municipalities with more than 500,000 inhabitants where all included in the sample. The other municipalities were updated yearly with a rotation criterion every other month. The households, corresponding to the voters, were randomly selected from the electoral rolls of the selected municipalities within each stratum. 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 (Equal Probability of Selection Method) sample (for more details see Martelli, Fullone 2008).

8 The consumer is a full-aged person belonging to the household identified by the selected telephone number. The individual has to contribute, also in non-monetary terms, to the family income.

stage[9]. The sample size amounted to 2,000 units and the reference universe is represented by the full-aged (18 years and over) population. Even if stratification ensures a relatively reliable sample, introducing CATI technique as data collection mode and the consumer as sampling unit strongly weakened the aim of maintaining an EPSEM structure[10]. This implies that the currently used data processing (no weights in the estimation process) does not reflect the sample design structure: weights should be used in the estimation process. In the survey, non-response unit is handled by substitutions within the same stratum. In order to guarantee the monthly planned number of interviews (that is 2,000), a sample of substitute households equal to four times the base sample size, for a total of 10,000 units, the latter is selected by systematic sampling from the subscribers list of the telephone land-line[11].

The implemented questionnaire includes both qualitative questions harmonised at European level[12] (characterised by three or five ordered reply options relating to the Italian economic situation and to the financial situation of the interviewed and of his family) and additional questions aimed at satisfying a need for information at national level.

Answers obtained from the survey are aggregated in the form of weighted balances calculated, for each question, as the difference between the percentages of respondents giving positive and negative responses. The weighted balances are aggregated to build the CCI, an overall summary indicator of the results of the survey aimed at evaluating the optimism/pessimism of Italian consumers. More specifically, the CCI in the Istat definition, is the average of nine balance series, namely: Q1 - Assessments on households' financial situation; Q2 - Expectations on households' financial situation; Q3 - Assessments on the general economic situation; Q4 -

9 Quota sampling does not allow to define exactly inclusion probability of an individual because of the selection process cannot be verified. This occurrence leads to lose the features of a random sample (sample error cannot be calculated). In the Italian survey, as the unit selection is performed according to random techniques (and not in an arbitrary way as it often happens with quota samples), calculation of inclusion probability with a good approximation is made possible and representativeness of data collected with purposive sampling holds. In Section 3 a brief description of the inclusion probability used in the calibration estimator is provided.

10 The consumers (sampling units) do not have equal inclusion probability because of their probability of inclusion depends on the probability to be selected within the household.

11 In this case, the (approximated) inclusion probability of the household i in the stratum h of the sample is calculated as f_h/F_h, where f_h is the number of households in the stratum h of the sample and F_h is the number of households in the stratum h of the population. This inclusion probability is used in the calibration estimator, as described in Section 3, footnote 14.

12 For more details see European Commission (2019).

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[13].

Table 1 - The sample

Demographic width/ Geographical 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	55	22	23	34	17	3	50	205
North centre	69	63	56	56	26	15	45	328
North east	62	68	82	56	19	96	0	384
Centre	41	38	53	82	51	41	95	401
South	75	58	78	97	78	44	31	460
Islands	34	25	30	51	27	32	22	222
Total	336	274	321	376	218	231	243	2,000

Source: Istat

3. The calibration estimator

The calibration estimator proposed by Deville and Särndal (1992) has been used to derive survey estimates for the Italian CS. The principle underlying the calibration method is to adjust samples through re-weighting individuals using auxiliary information, strongly correlated with study variables, for which population totals are available for instance from a register or administrative data. The main principles of calibration are here summarised.

Given a population U of N 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:

$$\hat{Y}_{HT} = \sum_{k\in s} (\frac{1}{\pi_k}) y_k = \sum_{k\in s} d_k y_k$$

13 The Q (n) refers to the numeration of the questions within the European Commission questionnaire (see European Commission, 2019).

where π_k represents the inclusion probability of unit k[14]. Let X_1, X_2......, X_j … X_J be the J 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:

$$\sum_{k \in s} w_k x_{jk} = X_j \quad \forall j = 1, \ldots. J$$

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

$$Min_{w_k} \sum_{k \in s} d_k G(w_k / d_k)$$

The solution of this problem provides the calibrated weights used in the estimation process. The calibration estimator of the total for the variable of interest will then be[15]:

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

In order to estimate data until March 2018, the calibration estimator used in CS takes into account 10 auxiliary totals, such as population by gender, 4 geographical areas (North-East, North-West, Centre, South) and 4 age classes (18-29, 30-49, 50-64, more than 65) known from the population register. Starting from April 2018[16], the estimation process adopts 13 auxiliary totals: seven age classes (18-29, 30-39, 40-49, 50-59, 60-64, 65-70, 71 and more), gender and geographical areas unchanged with respect to the previous period[17].

14 For CS, the inclusion probability of unit k (π_k) was calculated according to the following formula: $\pi_k = \pi_{hj}\, \pi_{k|j}$, where $\pi_{hj} = f_h / F_h$ is the proportion of the sample households on the total of households in stratum h of the population. Since the number of households included in the list of subscribers to landline is not available and considering that the list is affected by under-coverage problems, the proportion is a proxy of the real probability of the household j to be selected from the frame in the stratum h to which the individual k belongs to. $\pi_{k|j} = 1/($*no. family members>18 years*$)$ is the probability that individual k is interviewed, conditioned to the selection of the household j.

17 The calibration weights, sampling variance and confidence intervals of the estimates were calculated by ReGenesees (R Evolved Generalised Software for Sampling Estimates and Errors in Surveys), an R software for design-based and model-assisted analysis of complex sample surveys, developed in Istat (Zardetto, 2015).

4. Comparison of calibrated and unweighted CCI

Frequency and balance series were rebuilt with the calibration estimator for the period 1998-2018[18]. Consequently, CCI was recalculated. The calibrated estimates were recalculated for both the aggregate and the breakdowns disseminated.

This Section reports a comparison of calibrated and unweighted CCI, as well as of new and old composing series. Furthermore, the differences between the indicators are analysed. The results of the main sectoral breakdowns (by occupation, education, age) are also presented.

4.1 CCI and composing series

Starting from the CCI (for definition see Section 2), Figure 1 shows the calibrated series plotted against the unweighted one. A first graphical inspection highlights the two time series are likely to record very close monthly changes in both series. Peaks and troughs do not present shift and the R^2 correlation coefficient is very high (see Table 2).

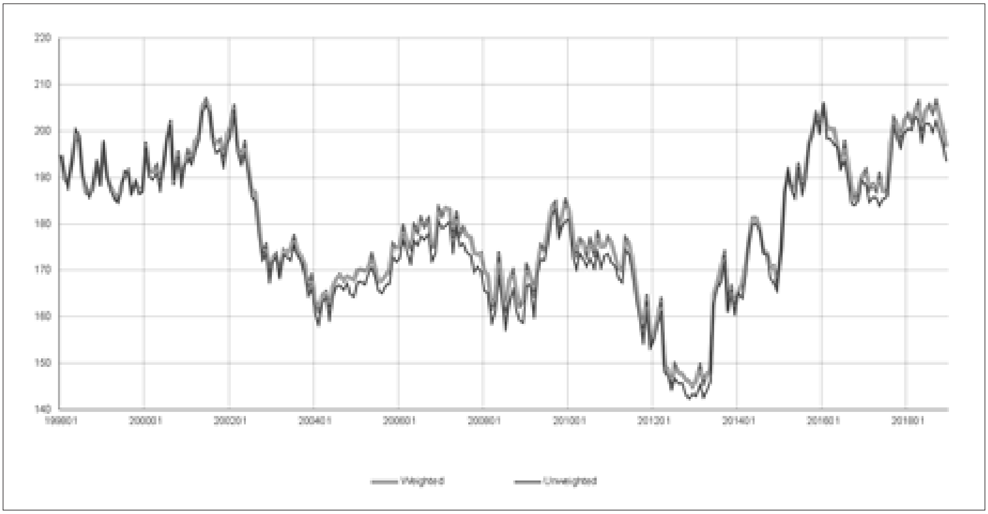
The weighted indicator presents, on the average, less variability than the current one, proved both by the standard deviation and the coefficient of variation (see Table 2). On the contrary, the mean is higher (-19.99 against -22.08) suggesting that the new CCI is, on average, always higher than the old one: calibration estimator generally produces more optimistic estimates.

This occurrence may be due to the correction by age classes introduced in the sample by the calibration estimator. Calibration indeed fixes the under-reporting of younger people, usually more optimistic, and the over reporting of elder people, with common negative opinions. Moving to CCI balance composing series, graphic inspection highlights very similar paths of calibrated and unweighted series: R^2 values suggest a very close and widespread correspondence between the signals stemming from the calibrated and unweighted balance series (Table 2). This outcome can be considered as an indirect support for the sample quality which correctly reflects the universe

18 This time span was chosen because throughout the years the sample design is consistent with the calibration estimator adopted.

structure. Moreover, for all the calibrated composing series the mean is higher than that of the unweighted series and variability is generally lower, as occurs for the CCI.

Figure 1 - Consumer Confidence Indicator



Source: Istat

Moving to analyse the differences between CCIs, Figure 2 shows that calibration had statistical effects on the estimates. More specifically, the calibration process introduced a systematic increase of the estimates (supported by the mean values that are always higher for the calibrated series). Moreover, differences exhibit an upward trend until 2008 and after 2015 onwards; in the 2009-2014 time span a downward trend was recorded.

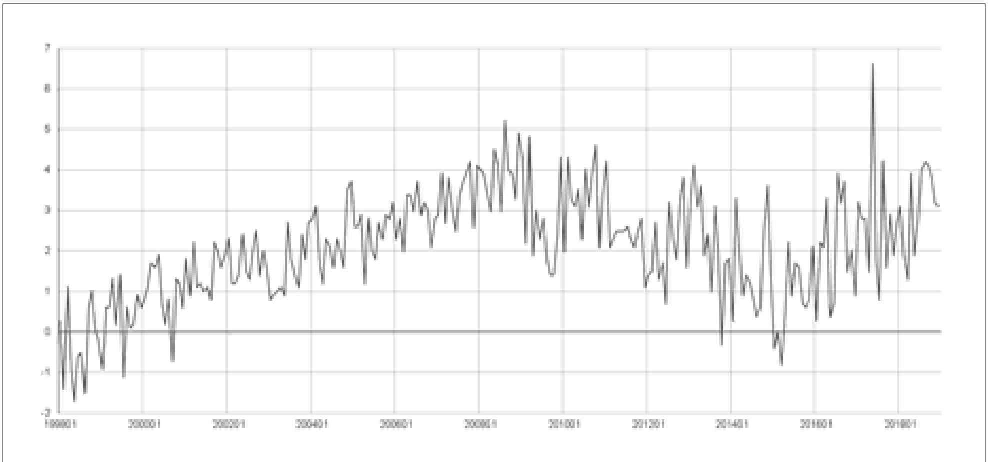
It seems that in a cyclical positive phase, calibration process produced higher estimates than the unweighted ones with growing differences between the two series while in a negative evolution of the economy it produced higher estimates with decreasing differences.

Table 2 - CCI and balance composing series – Comparisons of calibrated and unweighted Series

Series	Calibrated series			Unweighted series			Correlation between two series
	Mean	Standard deviation	Coefficient of variation	Mean	Standard deviation	Coefficient of variation	R^2
CCI	-19.99	14.841	-0.742	-22.080	15.248	-0.691	0.996
Ass. on the general economic situation (Q3)	-78.327	33.804	-0.432	-80.015	33.961	-0.424	0.998
Exp. on the general economic situation (Q4)	-18.558	21.151	-1.14	-20.329	21.928	-1.079	0.996
Ass. on households' financial situation (Q1)	-37.480	16.170	-0.431	-39.029	16.589	-0.425	0.996
Exp. on households' financial situation (Q2)	-5.514	9.885	-1.793	-7.290	10.641	-1.460	0.993
Expectations on unemployment (Q7)	47.532	29.122	0.613	47.160	28.760	0.610	0.998
Current households' financial situation (Q12)	7.288	14.050	1.928	5.633	14.030	2.491	0.996
Expectations on saving (Q11)	-43.431	18.893	-0.435	-49.538	20.964	-0.423	0.987
Assessments on saving (Q10)	116.427	28.358	0.244	113.756	28.192	0.248	0.997
Assessments on purchases (Q8)	-72.773	23.044	-0.317	-74.737	23.365	-0.313	0.990

Source: Istat
Q(n) refers to the numeration used in the European Commission questionnaire (European Commission, 2019).

Figure 2 - Consumer Confidence Indicator: differences between calibrated and unweighted series



Source: Istat

In Table 3 some descriptive statistics of the differences are described. The average significantly differs from zero (value is 2.0897), skewness coefficient is very low and negative, indicating that data are slightly skewed left and a positive kurtosis coefficient points out a "heavy-tailed" distribution. The influence of the introduced weighting system is corroborated by the paired t-Student test analysis. We tested the null hypothesis whereby the differences between the calibrated and the unweighted series are negligible, that is H_0: mean (calibrated-unweighted) = 0. The outcome highlights the differences introduced by calibration are relevant. The p-value is markedly lower than 0.01, which suggests rejecting the hypothesis of irrelevance of differences at a 1% level of significance 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 distribution) provides analogous results.

Moving to the differences in composing series, looking at the graphs (Figures A.1 - A.9 in Appendix) the shape is similar to that of the CCI for all the series but for the unemployment expectations for which the upward trend until 2009 is not present. Instead, we find an upward trend in 2010-2016 and a downward trend from 2017 onwards. We underline that assessments and expectations on saving show mean higher values while unemployment expectations series exhibits a very high coefficient of variation (see Table 3). The results of t-test and Wilcoxon test show that differences are significant for all the series at 1% level except for unemployment expectations where the level is 10%.

Table 3 - CCI and balance composing series - differences between calibrated and unweighted

Series	N	Mean	Standard deviation	Coefficient of variation	Skewness	Kurtosis	t-Student test	p-value	S-Signed Rank test	p-value
CCI	252	2.0897	1.3458	0.6440	-0.0914	0.1133	24.65	<0.0001	15,313	<0.0001
Ass. on the general economic situation (Q3)	252	1.6877	1.8238	1.0806	0.5231	0.6834	14.69	<0.0001	13,112	<0.0001
Exp. on the general economic situation (Q4)	252	1.7702	1.9775	1.1170	0.1264	0.9456	14.21	<0.0001	12,508.5	<0.0001
Ass. on households' financial situation (Q1)	252	1.5492	1.6057	1.0364	0.524	0.6667	15.32	<0.0001	13,376.5	<0.0001
Exp. on households' financial situation (Q2)	252	1.7758	1.4549	0.8192	0.2335	1.2827	19.38	<0.0001	14,255	<0.0001
Expectations on unemployment (Q7)	252	0.3722	1.9081	5.1262	1.0688	1.6108	3.10	0.0022	1,946.5	0.0796
Current households' financial situation (Q12)	252	1.6556	1.2302	0.7430	-0.0686	0.0365	21.36	<0.0001	14,181	<0.0001
Expectations on saving (Q11)	252	6.1071	3.7889	0.6204	-0.0366	-0.1595	25.59	<0.0001	15,579.5	<0.0001
Assessments on saving (Q10)	252	2.671	2.3087	0.8643	0.4258	1.192	18.37	<0.0001	14,463	<0.0001
Assessments on purchases (Q8)	252	1.9635	3.3282	1.6950	-0.0904	0.392	9.37	<0.0001	9,629.5	<0.0001

Source: Istat
Q(n) refers to the numeration used in the European Commission questionnaire – period: 1998-2018.

4.2 CCI in the main sectoral breakdowns

Outcomes of the CCI main sectoral breakdowns are presented in this section. More specifically, we analyse differences between calibrated CCI and unweighted CCI by age, occupation and education namely the sectoral breakdowns with significant differences between the weighted and unweighted structure of the sample (see Table 4). With regard to breakdown by age, correlation between weighted and unweighted series is high and widespread (see Table A.1 in Appendix). Differences show similar shapes and different levels for the four age groups considered (see Figures A.10 - A.13 in Appendix). More specifically, differences are mostly negative for the 30-49 group and are positive for 18-29 class especially in the last years. In the first case, calibration introduced a decrease in the estimates while in the second group the weighting process causes an increase of the estimates. The results of pair t-test and Wilcoxon-test provide evidence to reject the null hypothesis at the 1% level for all the groups exception made for the 50-64 age class (Table A.1 in Appendix).

Moving to breakdown for education, we highlight that CCI differences in primary education show a very similar path in respect to the aggregate one; secondary education is affected by an increase in the estimates that has become more marked from 2012 onwards (Figures A.14 - A.16 in Appendix). "Further" category does not present trend and the average is close to zero. These outcomes suggest that calibration worked: education, for which constraints were applied, is indeed linked to age. For all the three categories R^2 is high and parametric and not parametric tests provide evidence of effectiveness of calibration at the 1% level (Table A.2 in Appendix).

Finally, graphical inspection of differences by occupation (Figures A.17 - A.20 in Appendix) highlights the employed group's increase in the calibrated estimates especially in some periods while the self-employed present an average close to zero. The unemployed and inactive categories exhibit a marked increase of the calibrated estimates as "indirect" effect of the constraints on the age classes (young people are often affected by the unemployment and elder people are commonly inactive). As for the other examined breakdowns, R^2 is very high and statistical tests on the differences are significant at 1% level (Table A.3 in Appendix) for all the groups.

Table 4 - Unweighted and weighted structure of the sample (%)

Variables	Unweighted Sample* (%)	Weighted sample* (%)	Differences (Unweighted-Weighted)
Age			
18-29	8.6	16.7	-8.1
30-49	31.3	35.9	-4.6
50-64	30.0	23.2	6.8
65+	30.1	24.2	5.9
Total	100.0	100.0	
Gender			
Male	48.1	47.9	0.2
Female	51.9	52.1	-0.2
Total	100.0	100.0	
Geographical area			
North West	26.9	26.8	0.1
North East	19.3	19.2	0.1
Centre	19.8	19.7	0.1
South	34.0	34.3	-0.3
Total	100.0	100.0	
Occupation			
Self Employed	8.3	8.5	-0.2
Employed	32.8	35	-2.2
Unemployed	3.6	4.6	-1
Inactive	55.3	51.9	3.4
Total	100.0	100.0	
Education			
Primary	50.3	44.9	5.4
Secondary	38.0	42.5	-4.5
Further	11.7	12.6	-0.9
Total	100.0	100.0	

Source: Istat
Mean over years 1998 – 2018.

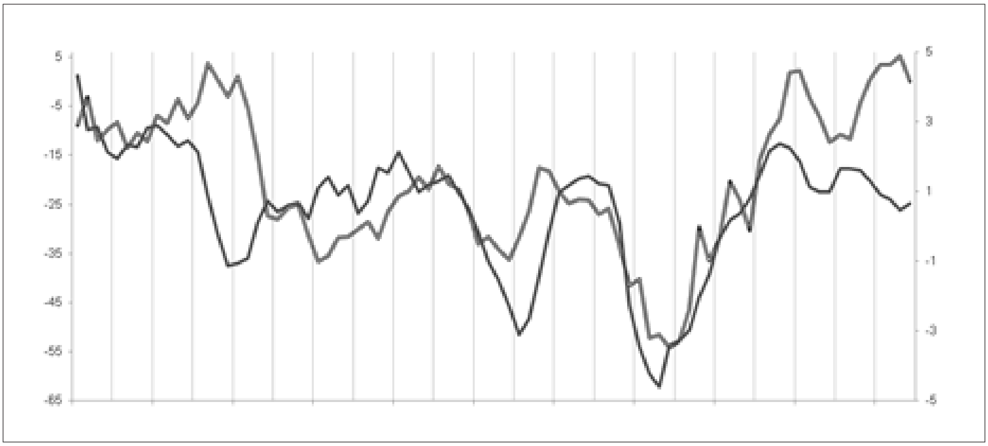
5. Calibrated CCI: performance in tracking private consumption

In this section the performance of calibrated CCI in tracking the reference series, i.e. private consumption, was evaluated and compared with that of the unweighted CCI. First, the calibrated and unweighted indicators were compared with the quarterly seasonally adjusted quantitative reference series namely final consumption expenditure of households on the economic territory, generated by Istat within the National Accounts framework (chain linked-volumes, reference year 2015). Since the seasonality is very weak, CCIs were not seasonally adjusted. As the reference series is available with a quarterly frequency, the monthly CCIs became quarterly by calculating the average indicator of the three months in each quarter.

The reference series underwent a preliminary testing for the presence of unit roots (namely, the test confirming that the time series data are not stationary), using the Augmented Dickey-Fuller (ADF) test. On the basis of the results of the ADF test, the quantitative series was found to be non-stationary and was transformed into quarter-on-quarter changes. The calibrated and unweighted CCI, being the average of balances bounded by construction, are considered "stationary" by construction. On the basis of this last consideration, the indicators were not subject to any transformation. In terms of methodology, the performance is based on correlation analysis, ability to track directional changes, in-sample properties evaluated by the Granger test and an in-sample model, an out-of-sample forecasting exercise and a volatility analysis.

In Table 5 results of performance are shown. As regard to correlation analysis, both coincident and one quarter leading correlations have been calculated. The two indicators present very similar values of correlation both coincident and leading. In addition, two indicators exhibit an alike pattern also in the moving correlation analysis over a period of five years.

Figure 3 - Calibrated CCI and Italian private consumption



Source: Istat

Table 5 - Performance in tracking private consumption of the unweighted and calibrated CCI

CCI	Correlation analysis		Rate of correct indications of change of direction	Granger causality test		MCD
	Coincident correlation	Leading correlation (one quarter)		F-statistic	Probability	
Unweighted	0.64	0.60	0.48	3.16	0.0481	3
Weighted	0.63	0.58	0.47	2.71	0.0731	3
In sample and out of sample analysis						
$c_t = \alpha + \beta CCI_t + \varepsilon_t$						
		Adjusted R^2	t-statistic			RMSE (Out-of-sample)
Unweighted		0.41	7.51			0.33
Weighted		0.39	7.30			0.40

Source: Istat

Referring to the ability of tracking directional changes, the percentage of correct indications of changes was 0.47 for calibrated CCI and 0.48 for unweighted CCI.

Moving to the in-sample properties, the Granger test was applied using the regression model that includes the past values of both the dependent and independent variables. Given that the series were quarterly, a specific decision was made to insert up to 2 lags, thus giving rise to the following equation:

$$c_t = \alpha + \sum_{i=1}^{l} \beta_i c_{t-i} + \sum_{i=1}^{l} \gamma_i CCI_{t-i} + \varepsilon_t$$

where i indicates the delay, c_t the quarter-on-quarter changes of the reference series, α, a constant, β_i and γ_i, respectively, regression coefficients for the past values of the dependent and independent variable CCI (confidence climate indicator), and lastly, ε_t, the error. We tested the null hypothesis whereby the reference series is not explained by the CCI past values, that is the coefficients γ_i are equal to zero for each i. The results indicate that both indicators are useful to predict the private consumption: the unweighted CCI at 5% level of significance and the calibrated CCI at 10% level.

In order to test the in sample and out of sample forecasting capability of the CCIs, the following model is run (European Commission, 2018):

$$c_t = \alpha + \beta CCI_t + \varepsilon_t \qquad (1)$$

where c_t is the quarter-on-quarter change in private consumption, CCI_t is the quarterly value of the confidence indicator, α is the constant and ε_t the error. The results of in-sample analysis were assessed by adjusted R^2 values. The out of sample forecasting power was tested according to the following procedure. Firstly, estimation for the period 1998q1 – 2017q1 was made and the 2017q2 was forecasted. Then fixing the beginning of the estimation sample to 1998q1, a second estimation was made at 1998q1-2017q2, being 2017q2 the forecast value, and 2017q3 was forecasted. The out-of-sample performance was evaluated by Root Mean Squared Errors (RMSE). The two CCIs exhibit a very similar forecasting performance both in sample and out of sample. The out-of-sample performance is slightly better for the unweighted CCI (RMSE is lower).

Finally, to measure the short-term volatility of the series, Months for Cyclical Dominance (MCD) were calculated. Also in this case, the weighted CCI shows the same value of the unweighted CCI (equal 3) meaning that the calibration process did not change short-term volatility in the CCI series.

6. Final Remarks

The results show that, at aggregate level, the calibrated estimates are very similar to the unweighted ones, confirming the quality of the underlying sample design. Differences are however present and statistically significant, proving the efficacy of the introduced calibration system. We underline that calibrated estimates are in general increased (more optimistic) to the unweighted ones as a result of relevant changes introduced into age brackets; moreover the weighted series are typified by less variability on the average if compared to the unweighted ones.

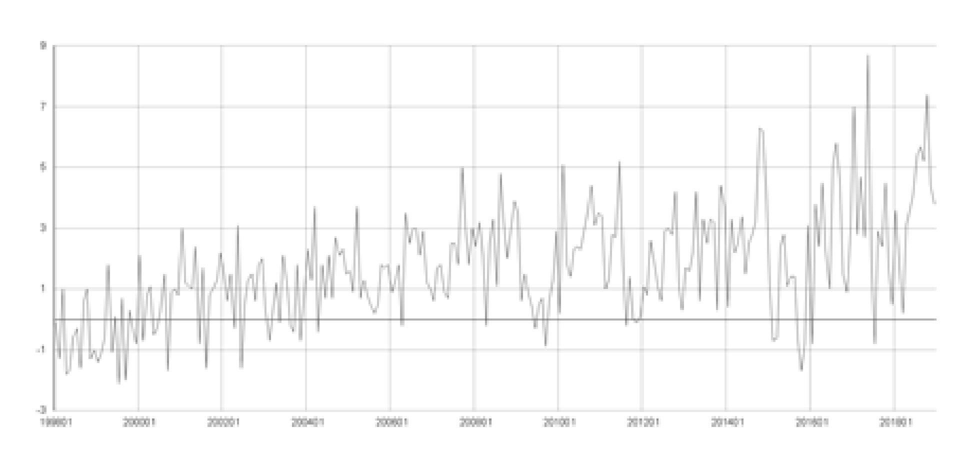
Referring to the various breakdowns, we underline that calibrated estimates of CCI generally show similar features to the aggregate ones (less variability on the average, increase of the estimates and very similar patterns with respect to the unweighted series) for all the sectoral breakdowns exception made for the 30-49 and 50-64 age brackets.

Regarding the relationship between CCI and the reference series, the calibrated and unweighted indicators exhibit equal short-term volatility and similar features in tracking private consumption proving calibration does not modify either the predictive capability of the CCI with respect to the private consumption or the short-term volatility of the series.

All in all, the results are satisfactory and Istat is planning to introduce the calibration estimator in the official data processing.

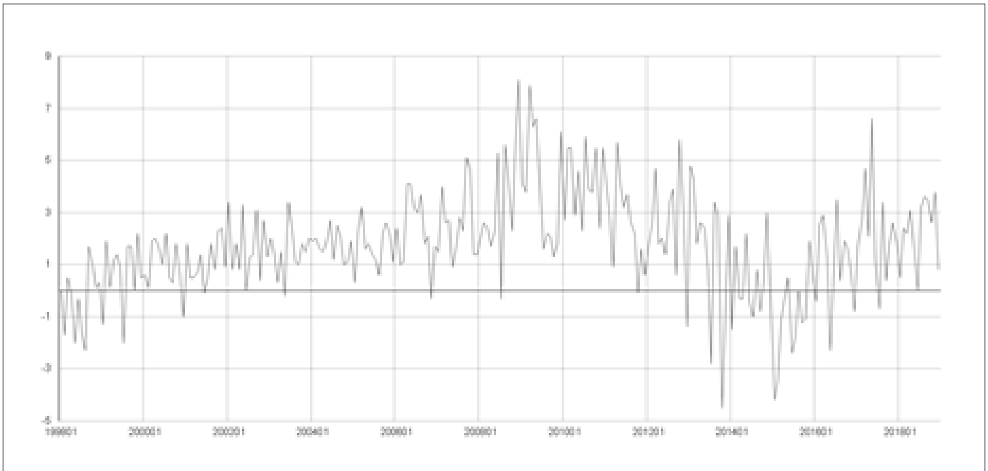
Appendix

Figure A.1 - Assessments on the general economic situation: differences between calibrated and unweighted series



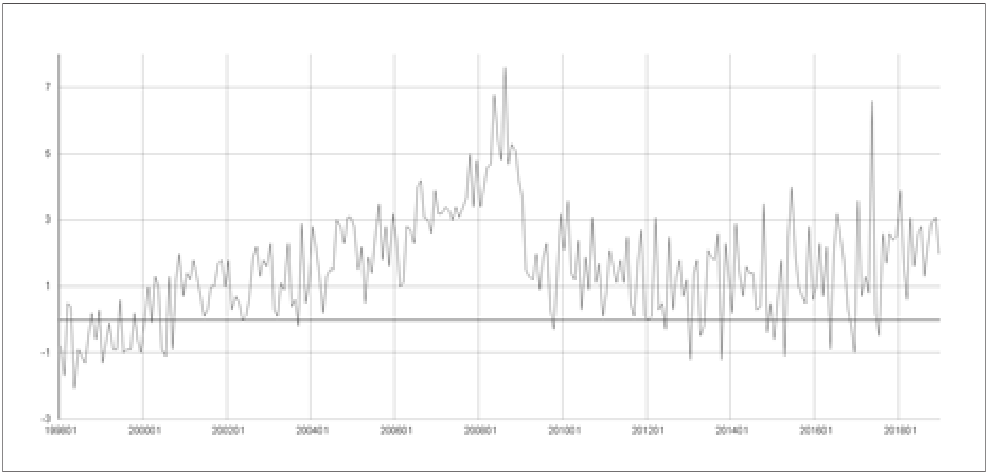
Source: Istat

Figure A.2 - Expectations on the general economic situation: differences between calibrated and unweighted series



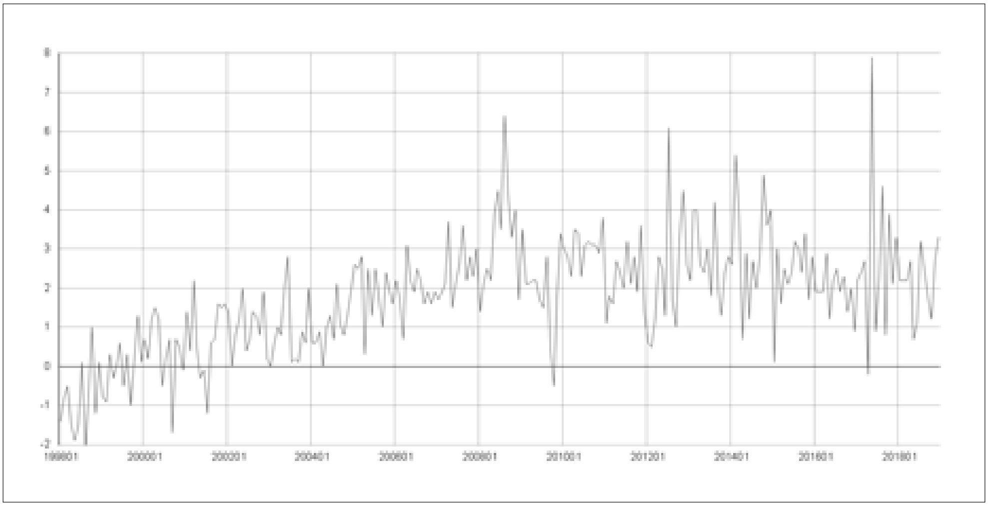
Source: Istat

Figure A.3 - Assessments on households' financial situation: differences between calibrated and unweighted series



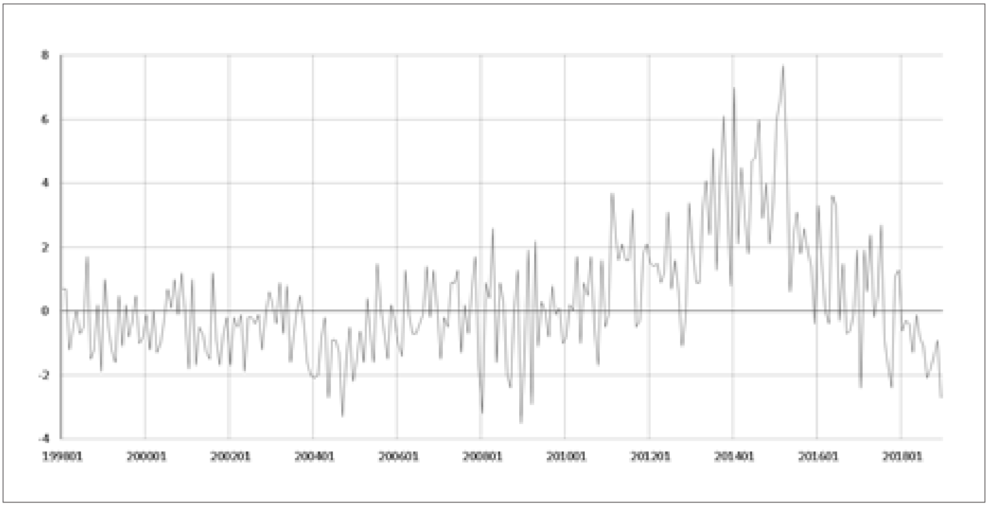
Source: Istat

Figure A.4 - Expectations on households' financial situation: differences between calibrated and unweighted series



Source: Istat

Figure A.5 - Expectations on unemployment: differences between calibrated and unweighted series

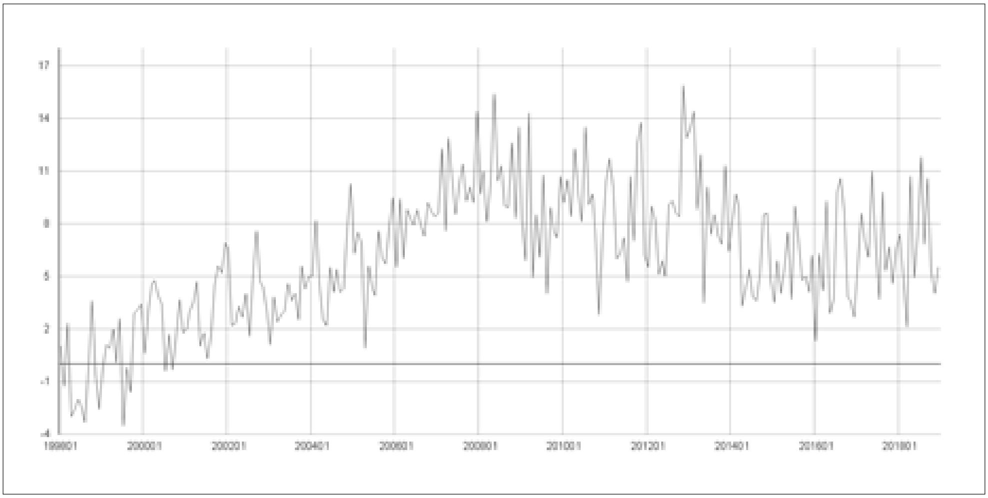


Source: Istat

Figure A.6 - Current households' financial situation: differences between calibrated and unweighted series

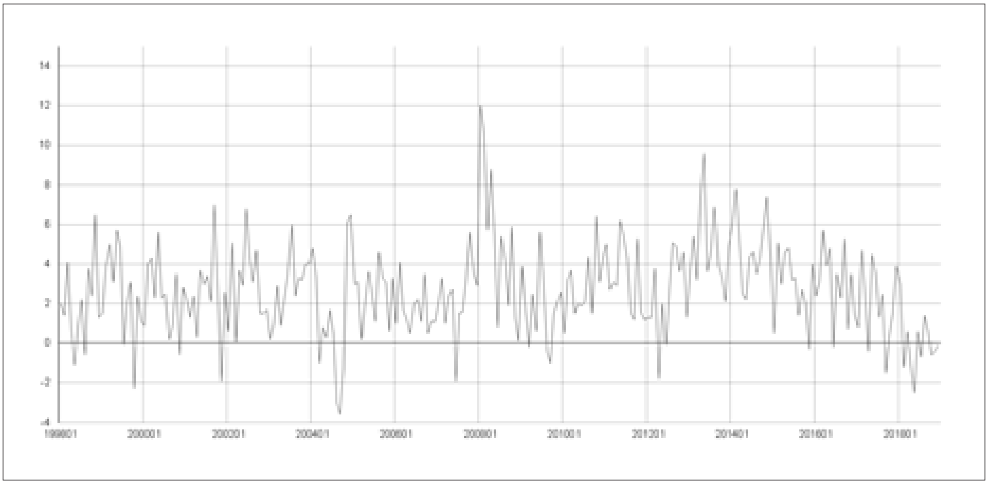
Source: Istat

Figure A.7 - Expectations on saving: differences between calibrated and unweighted series



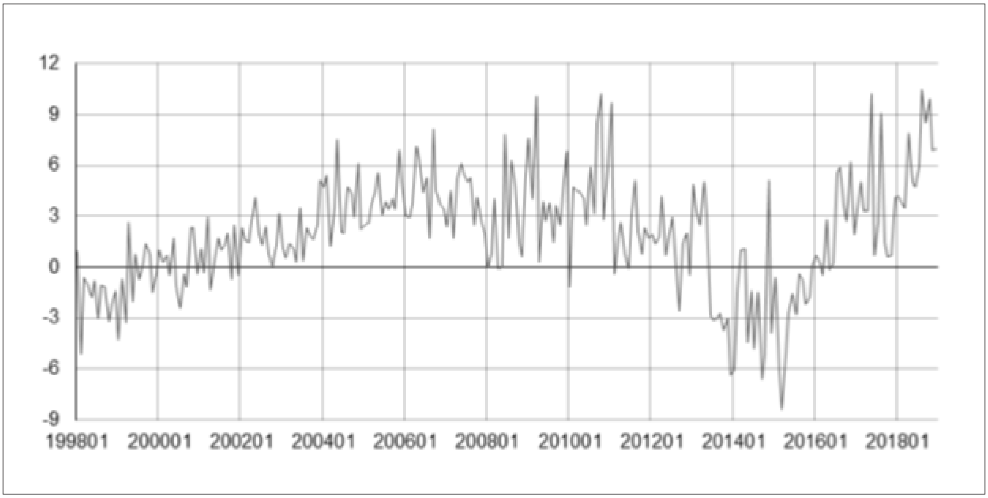
Source: Istat

Figure A.8 - Assessments on saving: differences between calibrated and unweighted series



Source: Istat

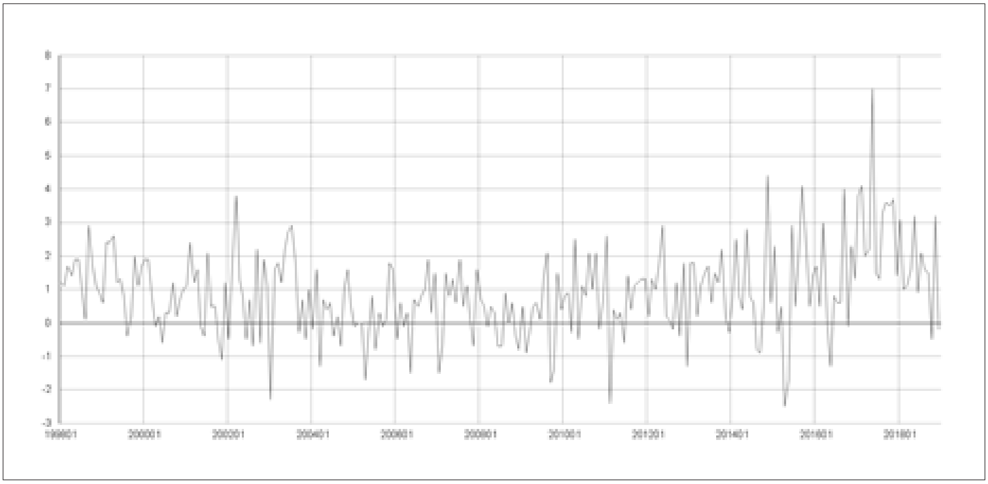
Figure A.9 - Assessments on purchase: differences between calibrated and unweighted series



Source: Istat

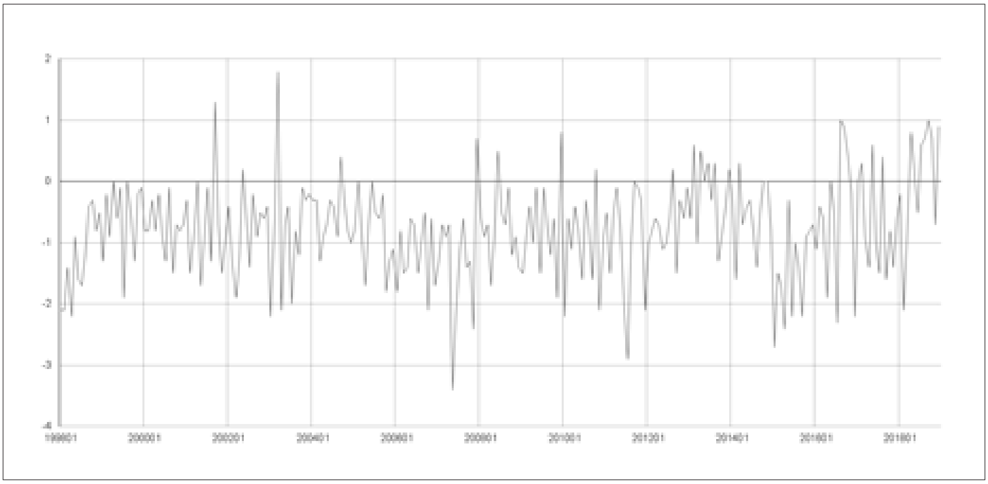
Breakdown by Age Brackets

Figure A.10 - 18-29 Age Bracket – Consumer Confidence Indicator – Differences between calibrated and unweighted series



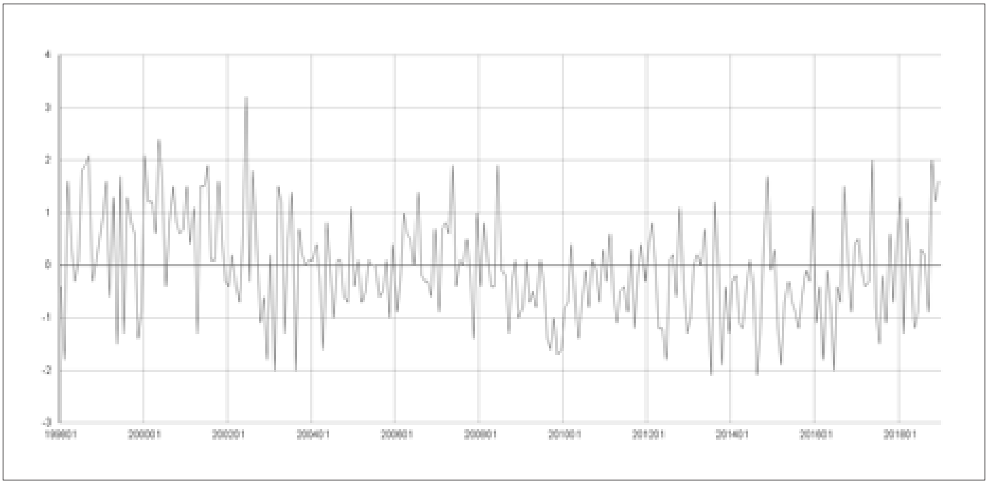
Source: Istat

Figure A.11 - 30-49 Age Bracket – Consumer Confidence Indicator – Differences between calibrated and unweighted series



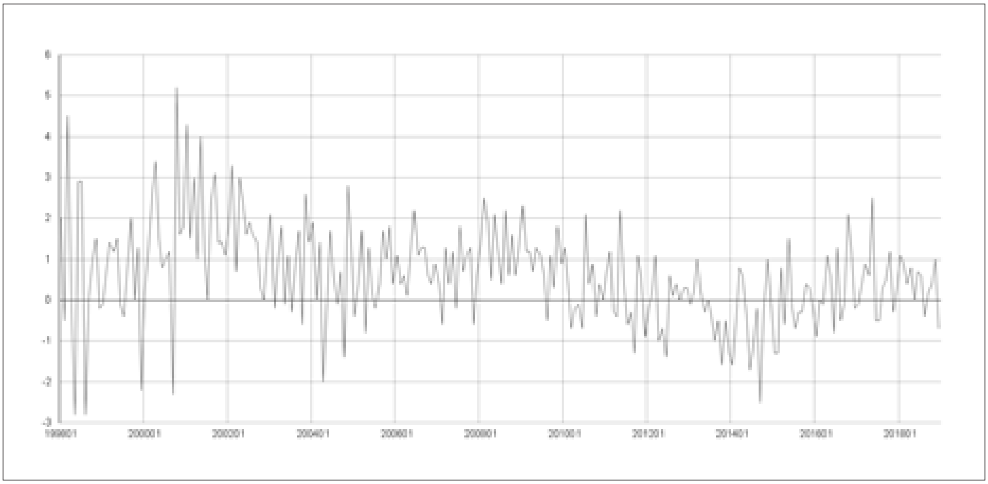
Source: Istat

Figure A.12 - 50-64 Age Bracket – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

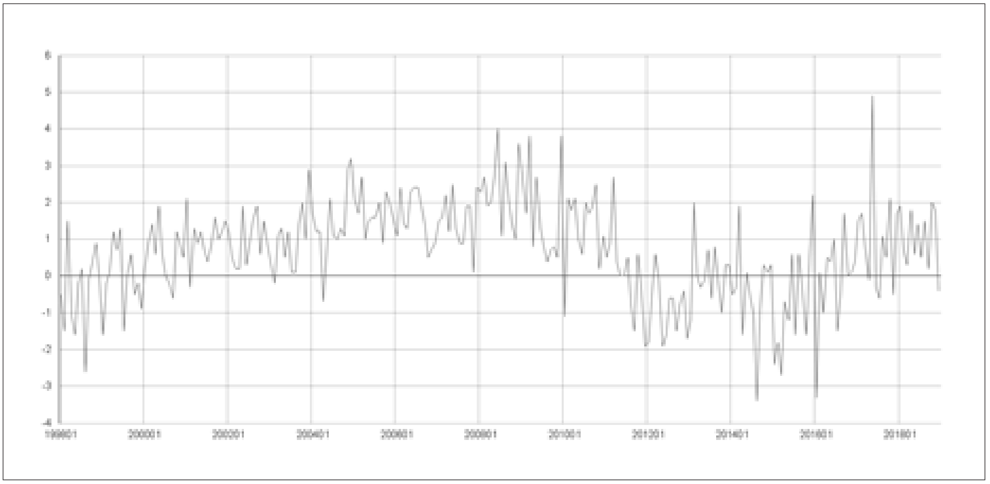
Figure A.13 - 65+ Age Bracket – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

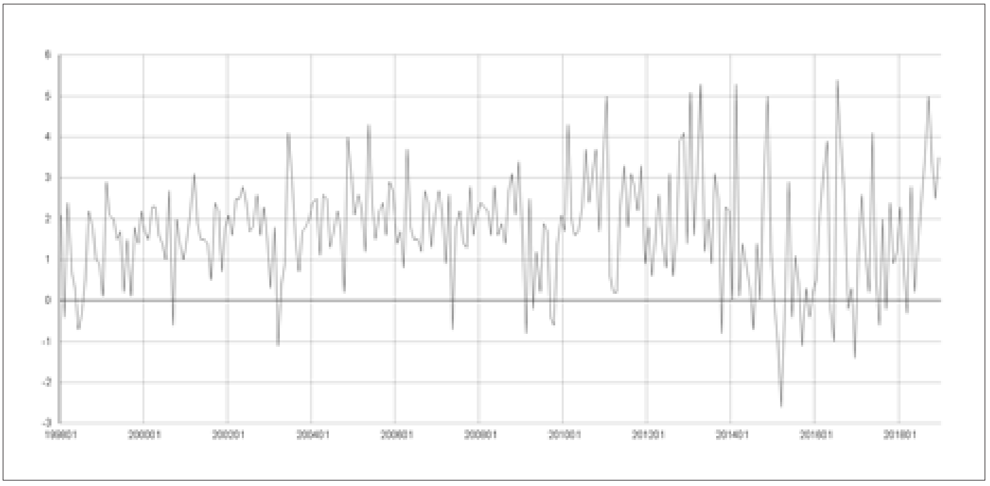
Breakdown by Education

Figure A.14 - Primary Education – Consumer Confidence Indicator – Differences between calibrated and unweighted series



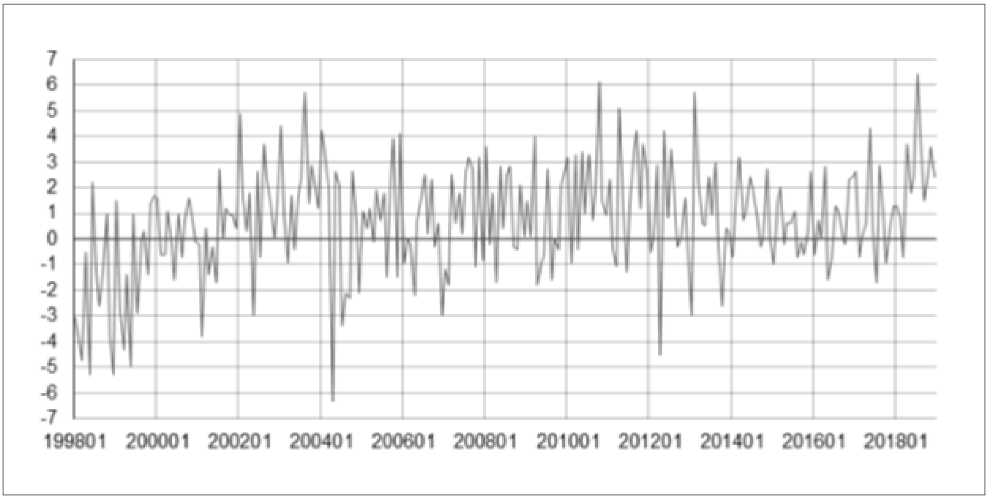
Source: Istat

Figure A.15 - Secondary Education – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

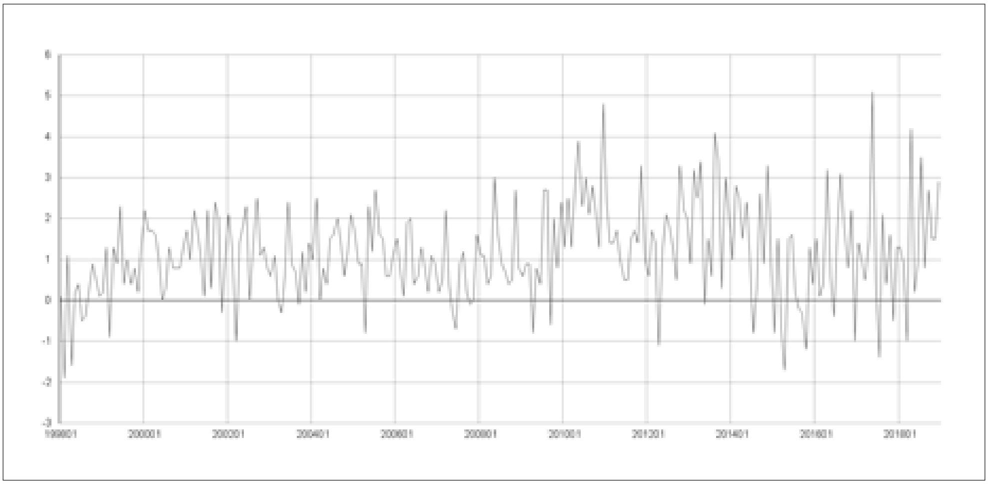
Figure A.16 - Further Education – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

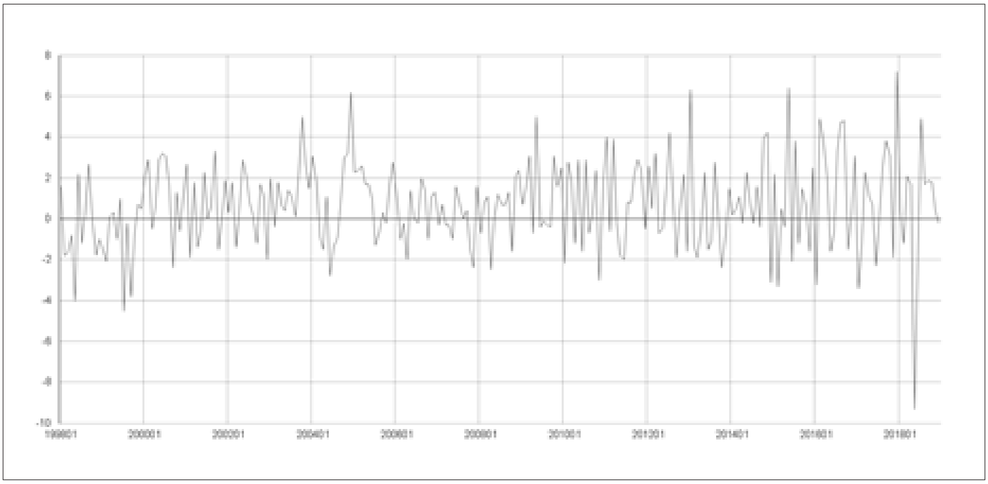
Breakdown by Occupation

Figure A.17 - Employed – Consumer Confidence Indicator – Differences between calibrated and unweighted series



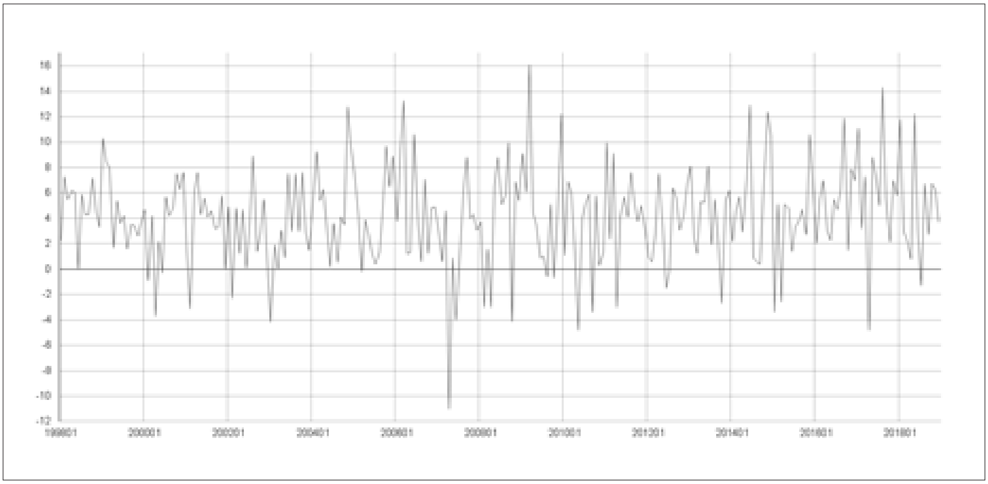
Source: Istat

Figure A.18 - Self-employed – Consumer Confidence Indicator – Differences between calibrated and unweighted series



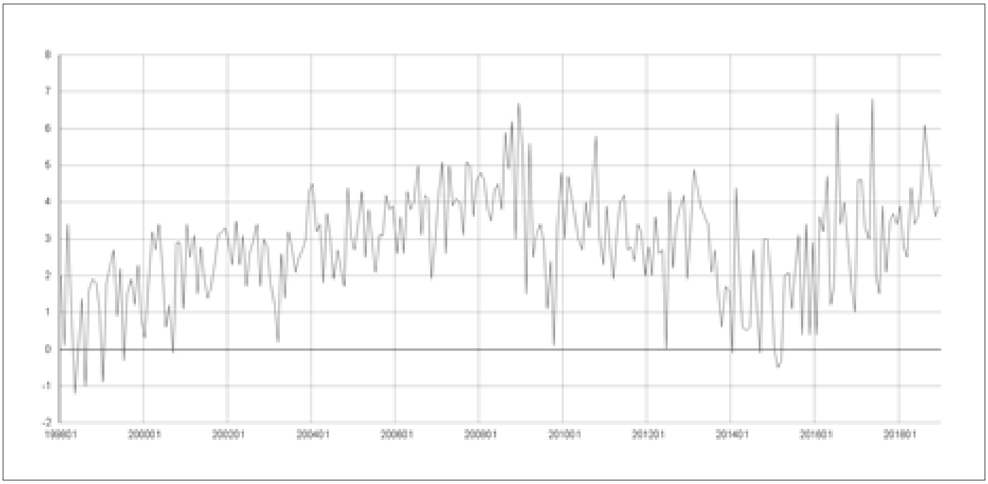
Source: Istat

Figure A.19 - Unemployed – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

Figure A.20 - Inactive – Consumer Confidence Indicator – Differences between calibrated and unweighted series



Source: Istat

Tables

Table A.1 - Breakdown by Age Brackets – Correlation, Paired t-test and S-signed Rank test for the differences of CCIs

Series	18-29 years					30-49 years				
	Differences (Calibrated - Unweighted)					Differences (Calibrated - Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	11.09	<.0001	10910	<.0001	0.99	-15.43	<.0001	-12412.5	<.0001
	50-64 years					At least 65 years				
	Differences (Calibrated - Unweighted)					Differences (Calibrated - Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	-1.00	0.3165	-1622.5	0.1369	0.99	8.11	<.0001	8333	<.0001

Source: Istat
Period: 1998-2018.

Table A.2 - Breakdown by Education – Correlation, Paired t-test and S-signed Rank test for the differences of CCIs

Series	Primary				
	Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	8.44	<.0001	8670	<.0001
	Secondary				
	Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	20.33	<.0001	14606.5	<.0001
	Further				
	Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	5.43	<.0001	6311.5	<.0001

Source: Istat
Period: 1998-2018.

Table A.3 - Breakdown by Occupation – Correlation, Paired t-test and S-signed Rank test for the differences of CCIs

Series	Employed					Self employed				
	Differences (Calibrated – Unweighted)					Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	16.00	<.0001	13221.5	<.0001	0.99	4.97	<.0001	5552	<.0001
	Unemployed					Inactive				
	Differences (Calibrated – Unweighted)					Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.97	17.51	<.0001	13824.5	<.0001	0.99	30.62	<.0001	15564.5	<.0001

Source: Istat
Period: 1998-2018.

Table A.4 - Breakdown by Gender – Correlation, Paired t-test and S-signed Rank test for the differences of CCIs

Series	Female					Male				
	Differences (Calibrated – Unweighted)					Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	24.65	<.0001	15108	<.0001	0.99	20.31	<.0001	14323.5	<.0001

Source: Istat
Period: 1998-2018.

Table A.5 - Breakdown by Geographical Area – Correlation, Paired t-test and S-signed Rank test for the differences of CCIs

Series	North West					North East				
	Differences (Calibrated – Unweighted)					Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	18.14	<.0001	13823	<.0001	0.99	19.20	<.0001	14716.5	<.0001
	Centre					South				
	Differences (Calibrated – Unweighted)					Differences (Calibrated – Unweighted)				
	R^2	Paired t-test	p-value	S-signed Rank test	p-value	R^2	Paired t-test	p-value	S-signed Rank test	p-value
CCI	0.99	16.81	<.0001	13608	<.0001	0.99	23.28	<.0001	15095.5	<.0001

Source: Istat
Period: 1998-2018.

References

Cochran, W.G. 1977. *Sampling Techniques*. Hoboken, NJ, U.S.: John Wiley & Sons.

Deville, J.C., and C.E. Särndal. 1992. "Calibration estimators in survey sampling". *Journal of the American statistical Association*, Volume 87, N. 418: 376 – 382.

De Vitiis, C., A. Guandalini, F. Inglese, e M.D. Terribili. 2015. "Fiducia dei consumatori – Confronto tra i sistemi di calibrazione". *Mimeo*. Roma: Istituto Nazionale di Statistica - Istat.

European Commission, Directorate-General for Economic and Financial Affairs. 2018. *A revised consumer confidence indicator*. Brussels, Belgium: European Commission.

European Commission, Directorate-General for Economic and Financial Affairs. 2016. *The joint harmonised EU programme of business and consumer surveys. User guide*. Brussels, Belgium: European Commission.

Guandalini, A. 2019. "Inference for the consumer confidence index". Mimeo. Roma: Istituto Nazionale di Statistica - Istat.

Martelli, B.M. 1998. "Le inchieste congiunturali dell'ISCO: aspetti metodologici". *Rassegna di Lavori dell'ISCO*, Year XV, N. 3: 13-67.

Martelli, B.M., e F. Fullone. 2008. "Re-thinking the ISAE Consumer survey processing procedure". *ISAE working paper series*. *Working Paper* N. 92/2008.

Särndal, C.-E. 2007. "The calibration approach in survey theory and practice". *Survey Methodology*, Volume 33, N. 2: 99-119.

Zardetto, D. 2015. "ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys". *Journal of Official Statistics*, Volume 31, N. 2: 177-203.