EUROPEAN CENTRAL BANK

EUROSYSTEM

ECB Consumer Expectations Survey

Guide to the computation of aggregate statistics

August 2022

Contents

1 Introduction

This guide presents a brief overview of the general methodology used to compute aggregate indicators for the ECB Consumer Expectations Survey (CES). The CES is an online panel survey of consumers that has been carried out on a monthly basis since January 2020. This document reviews the topics, variables and breakdowns covered in aggregate statistics reported in the latest results section of the CES web page on the ECB's website. Moreover, this guide describes how these aggregate statistics are computed as well as the methodology for managing and preparing microdata for the computation of aggregate results.

The microdata for the CES are collected through a survey of a panel of consumers in selected countries of the euro area. The survey is conducted by an external survey company on behalf of the European Central Bank (ECB). During the pilot phase, the survey covered six euro area countries: Belgium, Germany, Spain, France, Italy and the Netherlands. In January 2022 five more euro area countries were added: Ireland, Greece, Austria, Portugal and Finland.

More detailed information on sampling and weighting is available in the CES Occasional Paper, which evaluates the pilot phase of the CES. Further updates and details on the survey methodology are available on the CES page on the ECB's website.

2 Overview of indicators

The aggregate indicators disseminated can be grouped into four main topics: i) inflation, ii) income and consumption, iii) labour markets and economic growth, and iv) housing markets and credit. These are the main areas on which the survey regularly gathers information, either on a monthly or quarterly basis. The survey collects both qualitative and quantitative information. Qualitative variables are based on questions that ask respondents to select an answer from predefined categories or non-numeric response options. Quantitative variables, by contrast, are based on questions that ask respondents to provide a numeric estimate of the relevant economic concept.

Table 1 below summarises the set of variables for each topic included in the latest results section of the CES web page.

Table 1
List of variables reported for each topic

Topic	Label	Variable	Question type	Indicators
Inflation	Inflation perceptions over past 12 months	C1020	Quantitative	Mean Median
	Inflation expectations next 12 months	C1120		
	Inflation expectations 3 years ahead	C1220		
	Inflation expectations next 12 months	C1150	Probabilistic	Implied expectations Implied uncertainty
	Inflation perceptions over past 12 months	C1010	Qualitative	Up Down The same Net percentage
	Inflation expectations next 12 months	C1110		
	Inflation expectations 3 years ahead	C1210		
Labour markets and economic growth	Economic growth expectations next 12 months	C4020	Quantitative	Mean Median
		C4010	Qualitative	Grow Shrink The same Net percentage
	Perceived current unemployment rate	C4030	Quantitative	Mean Median
	Expected unemployment rate next 12 months	C4031	Quantitative	Mean Median
	Probability of losing your job next 3 months	Q2350		
	Probability of finding a job next 3 months	Q2300		
	Probability of looking for a job next 3 months	Q2390		

Topic	Label	Variable	Question type	Indicators
Household income and consumption	Household spending past 12 months	C6020	Quantitative	Mean Median
	Household spending past 12 months	C6010	Qualitative	Up Down The same Net percentage
	Expected household spending next 12 months	C6120	Quantitative	Mean Median
	Expected household spending next 12 months	C6110	Qualitative	Up Down The same Net percentage
	Household income expectations next 12 months	C3220	Quantitative	Mean Median
	Household income expectations next 12 months	C3210	Qualitative	Up Down The same Net percentage
Housing and credit access	Home price expectations next 12 months	C2120	Quantitative	Mean Median
	Home price expectations next 12 moths	C2110	Qualitative	Up Down The same Net percentage
	Mortgage interest rate expectations 12 months ahead	C5111	Quantitative	Mean Median
	Perceived credit access over previous 12 months	C7110	Qualitative	Harder Easier The same Not applicable Net percentage
	Expected credit access 12 months ahead	C7120	Qualitative	Harder Easier The same Not applicable Net percentage
	Credit application rate over previous 3 months	Q4010	Qualitative	Yes

3 General procedure

The computation of the CES aggregate indicators is based on the same microdata set that will be made available via the CES web page. This document guides the user through the steps needed to replicate the results disseminated by the ECB, starting with the underlying microdata.

Aggregate indicators are computed from April 2020 (Wave 4 of the survey) to the most recent survey wave available.[1] The countries currently included in the aggregate results are Belgium, Germany, Spain, France, Italy and the Netherlands. There are plans to add further countries to the aggregate statistics in the future.

The CES is structured in different modules. This includes:

- the recruitment and background modules, comprising questions asked once, when respondents first join the survey;
- the core module, comprising questions asked on a monthly basis;
- the quarterly module, comprising questions asked every three months.

Due to the different frequencies of the CES modules, the number of disseminated data points can vary across indicators. Only households that have completed the background questionnaire are kept in the panel; those that have not are removed (filtering applied with *survey_status_b*).

Weights are applied to all calculations in order to ensure representativity. For the computation of these indicators, we apply core weights (*wgt_bld*) when the variable is collected on a monthly basis, and quarterly weights (*wgt_bld_q*) when the variable is collected on a quarterly basis. Only population weights are used to compute the indicators.[2] More details on the weighting strategy applied in the CES can be found in the methodological guide available on the CES web page.

Where no additional breakdown is specified, aggregate indicators are computed by pooling information across all six CES countries. By contrast, where calculations are computed to show trends across the different breakdowns, the aggregate indicators represent the weighted result pooled across countries by breakdown group.

Following the methodology of other household surveys, indicators that are based on a sample of less than 20 observations are not published.[3] Indicators computed from a sample size of between 20 and 49 observations are published with the flag “u”, for

1 During the first three waves the CES sample was in a build-up phase. The targeted sample size was reached in April 2020.

2 In other words, we do not apply any further adjustment to account for euro area countries that are not part of the survey, and weighting is not question-specific.

3 See, for example, the methodological note on the EU Statistics on Income and Living Conditions (EU-SILC).

"unreliable". Indicators are reported without a flag if they are based on samples of 50 observations or more.

3.1 Qualitative indicators

Net percentages

Qualitative questions ask respondents to indicate a negative or a positive change. Net percentages are obtained by subtracting the percentage of individuals whose answer is a negative change from the percentage of individuals indicating a positive change in the variable of interest.

3.2 Quantitative indicators

Outlier correction

To correct for outliers, observations for all quantitative indicators are winsorised at the 2nd and 98th percentiles of the weighted distribution of responses for each survey round and country.[4]

Interpolated median

Medians are calculated using the symmetric linear interpolation approach suggested by Cox (2009) to account for response behaviour (bunching at integer values) and questionnaire design (only one decimal allowed). We assume, first, that the underlying true distribution is continuous and, second, that the true distribution is uniform between the discrete responses (required for *linear* interpolation). In applying linear interpolation, we adjust the cumulative distribution function to account for the aforementioned bias towards integer values.[5] The *symmetrically* linearly interpolated median[6] is then:

$$G^{-1}(0.5) = x_1 + (x_2 - x_1) * \frac{0.5 - G(x_1)}{G(x_2) - G(x_1)}$$

where $x_1 < x_2$ without loss of generality and the adjusted cumulative distribution function (CDF) $G(x)$ based on CDF $F(x)$ is the so-called mid-distribution function:

---

4 Robustness checks for winsorising at the 1st and 99th percentiles or at the pooled distribution of responses across all countries did not alter the aggregate results substantially. For inflation-related variables, alternative outlier correction strategies such as cutting off observations at specific values (e.g. below -12 and above 12) do affect the level of aggregate series. However, overall dynamics from month to month remain comparable.

5 All underlying distribution functions are weighted according to the appropriate survey weights.

6 Note that the standard linear interpolation uses the same formula as below, where $G(x)$ is simply replaced by the standard unadjusted CDF $F(x)$.

$$G(x) = \Pr(X < x) + 0.5 \Pr(X = x) = F(x) - 0.5 \Pr(X = x).$$

Note that the assumption of uniformly distributed expectations necessary for linear interpolation also justifies such an adjustment.[7]

Household spending

The variable C6020 is an open-ended question on past household spending in which respondents can choose a "Don't know" option. By answering "Don't know", respondents are routed to question C6030, which provides different spending values in brackets as answer options. The quantitative measure of household spending combines information from variables C6020 and C6030. To interpolate the bracketed responses, we use the median value of responses to C6020 that fall within each specific bracket range of C6030 and combine it with the median answers to C6020.[8]

3.3 Indicators from probabilistic bin questions

The probabilistic questions on expectations of inflation (C1150), home price changes (C2150), and household net income developments (C3250) in the CES enable us to elicit time-varying and consumer-specific subjective probability distributions. In particular, respondents allocate 100 percentage points across a set of predefined intervals of percentage changes ("bins"). The allocated number of points represents the likelihood that the expectations for inflation, house price inflation, and changes in household net income will be within each of the intervals with a specific probability. Based on this, statistics such as first and second moments from the implied distribution function can be computed for each respondent. For example, the implied interquartile range (IQR) is frequently used as a measure for respondents' uncertainty.

To estimate a respondent's subjective distribution function, we follow the approach developed in Engelberg, Manski, and Williams (2009)[9]. That is, if the respondent allocates non-zero probability mass to one or two bins, the subjective probability distribution is assumed to have the shape of an isosceles triangular distribution. If the respondent uses three or more bins in their forecast, a generalised (four parameter) beta distribution is fitted to estimate the implied CDF. Further details on

---

7 The same methodology is used also in other similar surveys. For example, the New York Fed Survey of Consumer Expectations applies the same methodology for the computation of interpolated medians. The overall dynamics are robust to using different approaches of median computations. In particular, the standard calculation of the median, the standard linear interpolation, the interpolation approach proposed by Harrel and Davis (1982) and a kernel-based median computation all yield similar results in terms of overall dynamics over time.

8 Alternative methods of combining these two variables other than conditional medians do not alter the results substantially. Such robustness exercises include (i) drawing a random value from an assumed uniform distribution for each bracket range (ii) using the simple mid-point of the respective bracket range and (iii) hot-deck imputation based on household characteristics.

9 Estimating subjective distribution functions from probabilistic bin questions remains an area of active research with various alternative methods (e.g. most recently Krüger and Pavlova, 2020). Users of the data might choose to implement alternative measures based on the probabilities assigned to the intervals.

the computations related to these subjective distributions can be found in the technical appendix to this document.

Moreover, when fitting subjective probability distributions, we exclude from these computations any cases where the sum total of assigned probabilities does not equal 100. Furthermore, we do not provide estimates for cases in which respondents only selected non-adjacent bins. More information and a thorough discussion on the use of non-adjacent bins can be found in the Occasional Paper that provides a first overview and evaluation of the CES.

The reported aggregate series represent the sample median of the implied average expectation and the sample median of the implied IQR (uncertainty).[10] The 25th and 75th percentile of the implied expectations and implied uncertainty allow us to quantify the sample "disagreement" about these two concepts.

[10] Their overall dynamics are robust to alternative specifications (i.e. alternative initial guesses, alternative constraints on location parameters, absolute difference as objective function).

4 Breakdowns

Each aggregate indicator is also broken down by the following household characteristics: country, age and income quintile.

Country

The country dimension indicates the country of residence of the respondent; Belgium, Germany, Spain, France, Italy and the Netherlands are the countries currently covered by the CES aggregate indicators. There are plans to add further countries to the aggregate statistics in the future.

Age

Respondents are classified into the following age groups: 18 to 34 years old, 35 to 54 years old and 55 to 70 years old.[11] Respondents' age is updated by the survey provider at the start of each new year.

Household net income

Income quintiles indicate the part of the income distribution that each household's disposable income belongs to. Thresholds are computed over weighted income distributions by country, wave and module. For this purpose, continuous imputed income variables are used.

[11] Respondents aged above 70 are excluded from the age breakdowns owing to their lower coverage in the CES. For all other aggregate series, respondents over 70 are included in the calculations.

5 Access to aggregate indicators

The aggregate indicators can be accessed through the Excel file made available on the data and methodological information section of the CES web page.

The list of main elements below may help users navigate through the dataset:

- *Wave:* indicates the month in which the observations were collected.
- *Topic:* indicates the subject to which the variable relates. There are four topics: inflation, income and consumption, labour market and economic outlook, housing and credit access.
- *Var:* refers to the coding of the variables in the questionnaire as presented in column three of Table 1, as well as in the documentation files that will accompany the disseminated micro dataset.
- *Var_label:* briefly describes the concept behind each variable as presented in column two of Table 1.
- *Breakdown:* indicates the household characteristics that are used to calculate the aggregate series. If breakdown is equal to *wave*, then the indicator is obtained by pooling data for all countries in each survey round. If breakdown is equal to *country, age or income*, then the aggregate series represent statistics for each of the sub-groups starting from April 2020 (Wave 4 of the CES).
- *Breakdown label:* indicates, for the lines where breakdown is equal to *country, age* or *income,* the sub-group in question. As an example, for breakdown equal to country, the breakdown label will be equal to *"BE", "DE", "ES", "FR", "IT" or "NL",* each one indicating the country to which the observations belong. If breakdown is equal to *wave*, the breakdown label will not contain any information because the data for all countries are pooled.
- *N:* indicates the number of observations (sampled units) for each variable and breakdown.
- *Weighted_N:* indicates the weighted number of observations (i.e. the population estimates corresponding to sampled units N).
- *Flag_N:* takes value "u", for "unreliable", if the indicator is based on a sample size of between 20 and 50 observations. It is empty for indicators based on samples of 50 observations or more.

References

Armantier, O., Topa, G., Van der Klaauw, W., and Zafar, B. (2017), "An overview of the Survey of Consumer Expectations," *Economic Policy Review*, Federal Reserve Bank of New York, Vol. 23, No 2, pp 51-72.

Becker, C., Duersch, P., Eife, T., and Glas, A. (2021), "Extending the Procedure of Engelberg et al. (2009) to Surveys with Varying Interval-Widths," *Discussion Paper Series,* University of Heidelberg, Department of Economics, No 707.

ECB (2021), "ECB Consumer Expectations Survey: An Overview and First Evaluation," *ECB Occasional Paper*, No. 2021/287, European Central Bank.

Cox, N.J. (2009), "IQUANTILE: Stata module to calculate interpolated quantiles," *Statistical Software Components*, Boston College Department of Economics.

Engelberg, J., Manski C. F., and Williams J. (2009), "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters," *Journal of Business & Economic Statistics*, Vol. 27,No 1, pp 30-41.

Harrell, F. E., and Davis, C. E. (1982), "A New Distribution-Free Quantile Estimator," *Biometrika*, Vol. 69, No 3, pp 635–640.

Georgarakos, D. and Kenny, G. (2022), "Household spending and fiscal support during the COVID-19 pandemic: Insights from a new consumer survey", *Journal of Monetary Economics*, Vol. 129, pp S1-S14.

Krüger, F., and Pavlova, L., (2020), "Quantifying subjective uncertainty in survey expectations," Working Papers of the Priority Programme 1859 ''Experience and Expectation. Historical Foundations of Economic Behaviour'', No 14.

Technical Appendix

As outlined above, we closely follow a procedure outlined by Engelberg, Manski, and Williams (2009) for fitting subjective distribution functions to probabilistic forecasts. The specific procedure applied in the CES aggregate indicators in cases where the selected bins are adjacent is outlined below.

Case 1 – One or two adjacent bins selected by respondents

If one bin is selected, the support of the triangular distribution lies between the lower and upper bound of the selected bin. In the case of two bins with the ranges $[l; m)$ and $[m; r)$, respectively, and probabilities assigned equally to each of the two bins, the base of the fitted triangular distribution ranges from the left lower bound (l) of the first bin to the right upper bound (r) of the second bin.

Where the assigned probabilities differ, the base includes the bin with the highest assigned probability in its entirety. Only part of the lower-probability bin is contained in the triangle, since the outer bound is shifted inwards to ensure that a *symmetric* triangular distribution is fitted.[12]

That is, the fitted triangle is fixed at the outer bound of the bin with an assigned probability of more than 50%, either l or r. Defining α as the probability assigned to the first of the two selected bins, the adjusted outer bound l' or r' can be obtained from one of the following shifting formulas:[13]

$$l' = m - (r - m)\frac{\sqrt{\frac{\alpha}{2}}}{1 - \sqrt{\frac{\alpha}{2}}}, \qquad for\ \alpha < 0.5$$

$$\mathrm{r}' = m + (m - r)\frac{\sqrt{\frac{1-\alpha}{2}}}{1 - \sqrt{\frac{1-\alpha}{2}}}, \qquad for\ \alpha > 0.5$$

That is, if a higher probability lies in the second bin ($\alpha < 0.5$), we are shifting the lower bound l inwards, replacing it by l', while we shift the upper bound inwards from r to r' when the larger probability mass lies in the first bin ($\alpha > 0.5$).

Assuming an isosceles triangular distribution with limits l and r, the general triangular distribution formulas then collapse to the following special cases:

---

12 We only follow this approach in the case of two bins that are not open-ended. If one of the two outer bins is selected, we anchor the triangular distribution at the innermost bound of the two bins, irrespective of whether the higher probability is assigned to the outer or inner bin.

13 More detailed explanations on shifted upper or lower bounds can be found in Krüger and Pavlova (2020) and Becker et al. (2021).

$$Mean: \mu = \frac{l + r}{2}$$

$$Standard\ deviation: \sigma^2 = \frac{(r - l)^2}{24}$$

$$Interquartile\ Range: IQR = (r - l)\left(1 - \frac{1}{\sqrt{2}}\right)$$

Case 2 – Three or more adjacent bins selected by respondents

If the respondent uses three or more bins in their forecast, a generalised (four parameter) beta distribution is fitted. There are closed form solutions for the mean, variance, and interquartile range.

The parameters of the generalised beta distribution are estimated by solving the following constrained minimisation problem:

$$\min_{a>1,b>1,l>L,r>R} \sum_{i=1}^{8} [B(x_i; a, b, l, r) - F(x_i)]^2$$

where a, b are shape parameters restricted to be greater than 1 and l, r are location parameters of the generalised beta distribution. Moreover, L, R are the constraints on the location parameters if respondents allocate positive probability to bins with unbounded intervals. $F(x_i)$ denotes the cumulative distribution function of the respondent's allocation of probability mass across the eight bins.

If neither the lowest nor the highest bin is used, the minimisation is over the shape parameters a, b only, with parameters l and r constrained to the upper and lower endpoints of the respective outer bins. If either the lowest or highest bin is used, the minimisation is over three parameters (two shape ones and the respective location parameter l, r). Consequently, if both outer bins are used, the minimisation is over four parameters (two shape ones and both location parameters l, r).

Once the generalised beta distribution is fitted, the mean, variance, and IQR are obtained. The initial guesses are $\{a, b, l, r\} = \{2,2, -12,12\}$. The constraints L and R are set to -38 and +38, respectively.

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

For specific terminology please refer to the ECB glossary (available in English only).