EUROPEAN CENTRAL BANK

EUROSYSTEM

ECB Survey of Monetary Analysts (SMA): description of SMA data[1]

1 Introduction

This document serves as a manual for the SMA Aggregate Results, and is available for each survey round. It describes the structure and the technical characteristics of the dataset.

The data are stored in comma-separated files (.csv format), a standard which is supported by most analytical software across various system platforms.

For each question, the identifiers or series keys for the results can be found in the notes below the respective tables in the Aggregate Results document published on the SMA webpage.

Time series structured data can be constructed by appending each survey round's individual data files.

2 Data structure and metadata

The longitudinal data are structured in 17 columns, where each dimension holds a specific attribute of the data. The data types are based on SQL data formats. Empty entries in the .csv format are denoted by "NULL" or "\N".

When a spreadsheet programme is used to open the .csv files some of the data may not appear as they are defined below. This may be caused by the default settings of the spreadsheet programme forcing some data into a particular format. The data defined in timestamp format are a common example of this.

Data structure and data types

Column No	Field	Variable type
1	survey_prov	string
2	vintage	string
3	vintage_date	timestamp
4	measure	string
5	question_no	string
6	series_key	string
7	item	string
8	area	string
9	frequency	string
10	horizon	int

[1] Version from 13 June 2022.

11	units	string
12	time_stamp	timestamp
13	date_str	string
14	date_descr	string
15	category	string
16	bin_descr	string
17	value	string

Column 1: survey provider

Value	Description
SMA	Identifier

Column 2: vintage

Value	Description
MMMYY	Example: JUN21

Column 3: vintage_date

Value	Description
YYYY-MM-DD hh:mm:ss	Timestamp according to ISO 8601

Column 4: measure

Value	Description
MEDIAN	Median (50th percentile)
P25	25th percentile
P75	75th percentile
MEAN	Mean (average)
PC	Percentage of respondents providing this answer

Column 5: question_no

Value	Description
string	Question number in current questionnaire

Column 6: series_key

Unique identifier which is concatenated from selected dimensions of the data. For example: SMA.APR22.MEDIAN.Q.Y.U2._Z.HIC

Dimension	d1	d2	d3	d4	d5	d6	d7	d8
Column	survey_prov	vintage	measure	frequency	units	area	category	item
Example	SMA	APR22	MEDIAN	Q	Y	U2	_Z	HIC

Column 7: item

Value	Description
String	Identifier for the item

Column 8: area

Value	Description
U2	Euro area changing composition

Column 9: frequency

Value	Description
D	Daily
M	Monthly
GC	Governing Council
Q	Quarterly
H	Semi/half annual
A	Annual
_Z	Not applicable

The frequency “GC” denotes the eight monetary policy meetings of the Governing Council held per year. When the survey was launched, these regular meetings took place in January, March, April, June, July, September, October and December. However, the meetings do not necessarily have to take place in a particular month – for organisational reasons the date of a meeting may fall in the subsequent month. This was the case for the first Governing Council meeting in 2022, when the meeting took place in February instead of January. The csv file, however, still refers to January as the month is predefined.

Column 10: horizon

Value	Description

Integer (int)	Horizon indicates the period ahead.

Column 11: units

Value	Description
Q	Quarter-on-quarter
Y	Year-on-year
PC	Percent
EUR	EUR billion
I	Interest rate (%)
BP	Basis points
PR	Probability out of 100
_Z	Not applicable

Column 12: time_stamp

Value	Description
YYYY-MM-DD hh:mm:ss	Denotes the period of the value; timestamp according to ISO 8601.

Column 13: date_str

Value	Description
string	Denotes a date as string; not used at the moment.

Column 14: date_descr

Value	Description
“<” or “>”	The bin description for extreme date bins containing “>” or “<” where the mass is assigned to the following bins
survey date	Survey date for scalars or collection-type questions
YYYY	The year in a time series-type of question
long run	Long run for annual time series where “long run” is understood as the ten-year forecast

Column 15: category

Value	Description
string	Identifier for the category

Column 16: bin_descr

value	Description
“<” or “>”	The bin description for extreme bins containing “>” or “<” where the mass is assigned to the following bins

Column 17: value

Value	Description
string	Result of question

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