



RESEARCH DATA CENTER

REX - Remote Execution / Sistema di elaborazione a distanza

Business Outlook Survey of Industrial and Service Firms (SONDTEL)

Dataset handbook

Reference period for the data: 1993-2023

Updated in: November 2023

Keywords: Business Outlook Survey, sample microdata, industry and services

Compiled by: Sample Surveys Division of the Statistical Analysis Directorate



BANCA D'ITALIA
EUROSISTEMA

November 2023

Table of contents

General information.....	3
Reference population.....	3
Availability period and frequency of the data.....	3
Data collection and regulatory references.....	3
Aggregate statistics and reference publications.....	4
Databases structure and characteristics.....	4
Methodology and imputation of missing data.....	5
Variables description and characteristics.....	6
How to cite the database.....	6
Appendix: Variables in the archives but not in the questionnaires.....	7

General information

The Business Outlook Survey of Industrial and Service Firms (the Business Outlook Survey – ‘Sondtel’) is conducted annually on a sample of Italian firms comprising approximately 3,000 firms in industry excluding construction, 1,000 private non-financial service providers (including trade, hotels and restaurants, transport and communication, real estate, IT and other private services) and 500 construction firms. Firms are asked to provide mainly qualitative information on trends in key economic variables for the current year.

Reference population

The reference population is the same as for the Survey of Industrial and Service Firms (‘Invind’) and is made up of firms with administrative offices in Italy and with at least 20 employees in industry excluding construction and in private non-financial services, and with at least 10 employees in construction.

Availability period and frequency of the data

The databases are updated at the end of each run of the survey using the responses provided by the firms during the interviews.

The dataset containing the figures updated to the new reference year is made available to the users as soon as the Survey report is published on the Bank of Italy’s website (see the section ‘Aggregate statistics and reference publications’).

Data collection and regulatory references

The survey is conducted annually between the last ten days of September and the first twenty days of October, with reference to economic activity in the current year. The interviews are led by the Bank of Italy's branches.

Aggregate statistics and reference publications

The main aggregate statistics are collected in [tables](#) available on the [survey's webpage](#) on the Bank of Italy's website.

The key findings of the survey are also published in a specific [report](#) of the '[Statistics Series](#)' of the Bank of Italy and, alongside other information, in the [Economic Bulletin](#) and the publications of the '[Regional Economies](#)' series.

Databases structure and characteristics

The data collected through the Business Outlook Survey cover all editions of the survey since 1993 for firms in industry excluding construction and in services and since 2007 for construction firms.

Since a large portion of the questionnaire changes from year to year, only for firms in industry excluding construction and in services, two different databases (in csv format) have been made available to users:

- Sondstor, which contains only the variables observed continuously over the years. The various years are identified by the variable **annoril**. Please note that in the Business Outlook Survey, the variable annoril indicates the year in which the survey is conducted (e.g. 2007 for the survey conducted between September and October 2007), while in the Survey of Industrial and Service Firms ('Invind') it indicates the year surveyed, i.e. the year prior to the one in which the data were actually collected (e.g. 2007 for the survey conducted in the early months of 2008);
- Sondxxxx, which contains all the variables observed in the year xxxx.

In the Sondxxxx databases some of the variables used to record data on the same phenomenon may have been given different names and a partially different response scale over the years. In the sondstor dataset, variables with the same meaning have been attributed a unique code over time, and their scale of values has been harmonized, to ensure comparability over time. For example, the name of the variable 'Investments planned for the following year' is s3 in the sond1993 database and x3 in the sond1994 database, while in the sondstor database, the name of the variable is stg3 for all the years in which the variable was recorded. For construction firms, sondcost is the only

database and it contains all the variables measured in the various years, with names that may change over time.

Each firm is assigned an identification code (the variable **ident**), which makes it possible to retrieve information only from the sondstor database for firms that were included in the survey in multiple years. This code, generated randomly, is used exclusively for longitudinal analysis, cannot be tabulated and is entirely unrelated to the variables that identify the firms.

In the sondstor and sondcost databases, the **annoril/ident** pair identifies data on a given firm in a given year. If the same firm participates in the Survey of Industrial and Service Firms ('Invind') and in the Business Outlook Survey, it will have the same **ident** variable in all the databases.

Please note that the Business Outlook Survey is considered valid only if fully compiled. To this end, all the variables include among the response options '9 = do not know, do not wish to answer' and, where necessary, '8 = not applicable'. The missing values should thus not be interpreted as a firm refusing to provide a given response, but as questions asked only to a subset of firms; in some cases, for the group of firms that does not receive these questions, the relevant field can be populated with the figure '99'.

Methodology and imputation of missing data

The sampling design of the Business Outlook Survey essentially mirrors the design of the Survey of Industrial and Service Firms ('Invind') conducted in the spring of the same year. In principle, the firms contacted are the same for both surveys and any differences in the composition of the sample are due exclusively to sample attrition and to some firms possibly leaving the survey reference population in the summer.

The survey follows a stratified weighting procedure similar to that used for Invind, which produces two weighting schemes:

- *Weighting by employees*, which uses a variable of scale (the number of employees); the estimates made with this weighting scheme take account of the different sizes of firms and provide data on the macroeconomic findings;
- *Weighting by firms*, which is better suited to summarizing the behaviour of individual firms.

Before the complete dataset is created, the responses undergo quality checks. The estimates published in the survey report, in the Statistics Series, refer to the databases as at the moment in which the calculations were performed and to the most recent data available regarding the population of firms. Small differences between the database used to prepare the estimates presented in the report and the one made available to prepare the calculations remotely can be the result of both subsequent revisions of the responses provided by the firms and of revisions in the weighting scheme due to a change in the reference population. The weighting scheme is revised systematically, while any changes made to the responses provided in previous years are to be considered occasional.

Variables description and characteristics

The description of the variables and the domain of the relevant attributes are contained in this excel file. [Variables description and characteristics \(formats of the variables and codes/domains\)](#)

For confidentiality reasons, the databases are not searchable for questionnaire variables that would make it possible to identify the respondent firm. These variables are usually found in the first pages of the questionnaires and include: tax identification number, company name, subsidiary surveyed and group affiliation. Open text replies to questions of the type 'Other, specify' are also not searchable.

Conversely, the databases contain some variables that are not included in the questionnaires (e.g. survey reference year, variable **annoril**) but which are useful for calculation purposes (see the full list in the Appendix).

How to cite the database

Banca d'Italia (2023): 'Business Outlook Survey of Industrial and Service Firms', November 2023 (1993-2023).

Appendix: Variables in the archives but not in the questionnaires

a) Classification variables with respect to the sector of economic activity¹

Variable	Values	Description	ATECO 2002	ATECO 2007
settor11	SS1	Food industries, beverages and tobacco products	DA	10, 11, 12
	SS2	Textiles, clothing, and hide, leather and footwear products	DB, DC	13, 14, 15
	SS3	Coke manufacturing, chemical industry, rubber and plastics	DF, DG, DH	19, 20, 21, 22
	SS4	Processing of non-metallic minerals	DI	23
	SS5	Metal engineering industry	DJ, DK, DL, DM	24, 25, 26, 27, 28, 29, 30, 33
	SS6	Other manufacturing industries	DD, DE, DN	16, 17, 18, 31, 32
	SS7	Other industries excluding construction	CA, CB, CE	05, 06, 07, 08, 09, 35, 36, 37, 38, 39
	SS8	Wholesale and retail commerce	G	45, 46, 47
	SS9	Hotels and restaurants	H	55, 56
	SS10	Transport and communications	I	49, 50, 51, 52, 53, 58, 59, 60, 61, 62, 63
	SS11	Real estate activities, IT, etc.	K	68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82
indag3	1	Manufacturing industry	D	C
	2	Extractive Industries – Energy	C, E	B, D, E
	3	Services	G, I, H, K	G, I, H, J, L, M, N
indagine	1	Industry excluding construction	C, D, E	C, B, D, E
	2	Services	G, I, H, K	G, I, H, J, L, M, N

¹ Until 2009 the variables were obtained by aggregating two-letter groups (sub-sections) of the ISTAT ATECO 2002 classification; Starting from 2010, the classification variables by sector of economic activity are derived by aggregating two-digit groups (divisions) of the ISTAT ATECO 2007 classification. The variables in this table are contained only in the databases of firms in industry excluding construction and in services; all the firms in the construction database have a single ATECO code (2002, 2007).

b) *Classification variables with respect to the size class²*

Variable	Values	Description
cldimet	0	20 - 49 employees
	1	50 - 99 employees
	2	100 - 199 employees
	3	200 - 499 employees
	4	500 - 999 employees
	5	≥ 1.000 employees
cc2	1	20 - 49 employees
	2	≥ 50 employees

c) *Classification variables with respect to the geographical area³*

Variable	Values	Description
areag4	1	North- West
	2	North-East
	3	Centre
	4	South and Islands
areag2	1	North, Centre
	2	South and Islands

d) *Variables concerning the sample design and the weighting system⁴*

strato: Consisting of 66 combinations of settor11 and cldimet. It should be noted that firms with at least 5,000 workers have a weight of one and can be considered, individually for industry and services, as being part of two separate strata (i.e. 67 and 68).

poststrato: Consisting of 48 combinations of areag4, cc2 and a re-aggregation of the sectors of economic activity into 6 groups: 1) indag3=1; 2) indag3=2; 3) settor11=ss8; 4) settor11=ss9; 5) settor11=ss10; 6) settor11=ss11.

pesorisc: Sample (expansion) weight: at the stratum and poststratum level, the sum of the weights reflects the distribution of firms in the survey reference population and it sums to the number of firms in the sample. A weight equal to 1 is assigned to firms with more than 5,000 employees ('self-representatives') and to a limited number of firms that are not considered representative of the stratum to which they belong.

² Up to the 2003 reference year, the size class refers to the number of workers at the end of the year; from 2004 onwards to the average number of workers during the year.

³ For confidentiality reasons the classifications by region and province are not available, only those by macro-region.

⁴ For further details on the sampling design, on the construction of the weights and deflators and on all other methodological aspects, see the 'Methodological Notes' published on the Bank of Italy's website in conjunction with each of the [survey reports](#).

pesoadd: Sample (expansion) weight: at the stratum and poststratum level, the sum of the weights reflects the distribution of employees in the survey reference population and it sums to the number of firms in the sample. The variable pesoadd is recommended if one wants to produce a frequency distribution of the different types of response for a variable that takes into account the firms' size and therefore attaches greater importance to the responses of larger firms by number of employees.

popstr: Size of the population at the stratum level.

poppost: Size of the population at the post-stratum level.

e) *Classification variables of the share of turnover exported*

Variable	Values	Description
a6	0	non-exporting firm
	1	less than 1/3 of turnover exported
	2	between 1/3 and 2/3 of turnover exported
	3	more than 2/3 of turnover exported
qexp	1	less than 1/3 of turnover exported or non-exporting firm
	2	between 1/3 and 2/3 of turnover exported
	3	more than 2/3 of turnover exported