

Methodological note

GUIDANCE ON THE IMPACT OF THE COVID-19 CRISIS ON ANNUAL BUSINESS STATISTICS

EUROSTAT, DIRECTORATE G, BUSINESS AND TRADE STATISTICS

EUROSTAT DIRECTORATE B, UNIT B1 – METHODOLOGY; INNOVATION IN
OFFICIAL STATISTICS

1. Background

As a consequence of COVID 19 outbreak, many countries across the globe have implemented and imposed partial or complete lockdowns of businesses during part of 2020. In this context of lockdowns, a majority of NSIs indicated at some stage concerns regarding the disrupted data collection and production of many key annual business statistics for at least two reasons.

Firstly, surveying businesses about their economic activities and transactions is difficult as many have temporarily shut down or simply do not have the resources to respond to statistical questionnaires. Without timely and reliable data from the enterprises, the process of data production could either be delayed or some of the figures could be estimated based on partial or outdated information.

Secondly, many staff in national statistical offices are now moved to teleworking, which could impose a real challenge to publish relevant statistics of the highest quality, as planned in the Eurostat release calendars.

Eurostat issues this note to provide guidance to NSIs to ensure continuity in production process and dissemination of their annual business statistics, focusing in particular on PRODCOM, structural business statistics, foreign affiliates statistics, R&D statistics, ICT for enterprise statistics and Innovation statistics and on the national statistical business registers that are the backbone for collecting business statistics. The note aims in particular to provide a harmonised approach on how to deal with the lower response rates or delays, as well as to ensure sound imputation techniques also in the case of businesses that were (temporarily) closed down.

2. Businesses closed down due to the COVID-19 crisis

Depending on the extent of the confinement measures taken in the respective EU countries, all or part of the kind of activity units (KAUs) and enterprises (later called “statistical units”) with certain activities have temporarily ceased to function. These statistical units can be partially identified based on their NACE activity code, as sectoral approaches were used to identify those subject to lock down measures. In addition, administrative information on social security could be used, for instance to identify enterprises registering for temporary unemployment of their staff.

It is important to distinguish statistical units that have temporarily ceased their activities from those that have permanently closed down. The national business registers should therefore be kept as much as possible up-to-date with respect to the activity status. In the next weeks, short surveys or direct contact with the statistical units could be used to determine their activity status and update the statistical business registers. Fiscal administrative sources could be less effective, as in many countries the deadlines for VAT and similar fiscal declarations have been postponed by several months.

To increase the identification of the active business population that can still be surveyed because they were only temporarily closed down, additional indicators on the activity status of enterprises or legal units could be added to the national statistical business registers. The inactivity of statistical units as well as the date when they became inactive should be taken into account when estimating or imputing the economic development for the reference years concerned by the crisis.

3. Lower response rates due to the COVID-19 crisis for surveys / administrative data

Despite the difficulties, the best possible efforts should be made to gather survey responses from businesses. In this respect, the communication side toward the observation units (enterprises) is very important. The tone of voice may have to be adapted to make the respondents aware of the importance of the reliability of the statistics and how that would help the policymakers to take the right decisions concerning their sectors of activity. This can be in the form of, for example alerts on the websites of the statistical offices, alerts in the reporting tools for businesses:

- [NL has posted a notice on their website](#)
- [FR took a similar initiative](#)

Reminders to submit reports should not be sent anymore in case an enterprise has shut down or rendered itself in inability to retrieve data.

In the current circumstances, businesses may not be able to provide the (administrative) data they are legally required to provide. In some countries, administrations have allowed additional time for providing administrative data.

In case a full "normal" survey is not possible, simplified questionnaires can be proposed, suitable to the current situation.

4. Methods for estimating and imputing missing data for business statistics

A) ESTIMATION AND IMPUTATION METHODS: OPTIONS WITH AND WITHOUT ADMINISTRATIVE DATA

It is important to differentiate between missing data coming from *unit non-response* or *item non-response*. Although unit non-response could, in principle, also be treated with imputation, weighting methods are more commonly used such as:

- adjusting the weights
- resampling and using secondary sampling unit

It should be taken into account in the current context that the non-response of a statistical unit might more often than in the regular situation be an indication of (temporary) ceased activity. More detailed methods for adjustment for non-response is available in the [Handbook of Theory and Research](#).

Imputation is used more for item non-response, where there is existing partial information about the units, which can be extended by imputation to improve the quality of the statistics.

The [Memobust handbook](#) lists the following cases and methods for imputation:

- [deductive imputation](#)
- [model-based imputation](#)
- parametric
- non-parametric
- [donor imputation](#)
- cold deck
- random hot deck
- sequential hot deck
- nearest-neighbour
- predictive mean matching
- [longitudinal imputation](#) (panel data)
- last information carry forward
- interpolation
- mean/ratio imputation
- nearest-neighbour/Little-Su method

The [Guidelines on the use of estimation methods for the integration of administrative sources](#) provides recommendations on how to use the above listed methods, if administrative data sources are available.

B) R PACKAGES SUPPORTING IMPUTATION

The statistical production can be assisted by existing **R packages**:

- **VIM: Visualization and Imputation of Missing Values**, package created in the ESS, colleagues from Statistics Austria ensure the maintenance
- **Amelia: A Program for Missing Data**, tool developed for imputing missing data based on bootstrapping based algorithm
- **mice: Multivariate Imputation by Chained Equations**, package for more advanced users

C) OTHER CONSIDERATIONS

To **minimise non-response**, efforts to reduce the burden on the businesses should be made. For instance, CATI/CAWI should be increased, and in case respondents can be contacted, a simplified questionnaire shall be used, if possible, to lower the burden. Best practices in the NSIs for the fieldwork are welcome and will be shared when available, as well as for alternative sourcing such as for webscraping and machine learning.

With respect to the reference years affected by the COVID-19 crisis, some statistical units temporarily may have taken up **alternative activities**, e.g. producing disinfectants, protective masks, offering home delivery services. It is not recommended to reclassify them for the reference year concerned since these will be only temporary activities and might not affect the business structure in the long term. The continuity rules for attributing NACE codes should continue to be taken into account. Although information on temporary conversions of businesses is interesting, ad-hoc experimental statistics rather than official statistics might be preferable for this purpose.

In the period of the crisis **wages and salaries¹ might be partly subsidised by governments**. According to the definitions applied in business statistics, all wages and salaries paid by employers should be included and **such subsidies should not be netted out**. The countries should report as “wages and salaries” only those payments that the statistical units make to employees, and not measures, addressed by government directly or indirectly to the employees.

Other subsidies on production such as subsidies of wages and salaries **should not be reported as turnover** of the statistical units, but they are taken into account in the **value added at factor cost**.

Preference shall be given to the **sectoral approach** for the reference years affected by the COVID-19 crisis whenever possible as some branches are obviously impacted in some positive or negative way. NSIs could be looking at the respective national legislation for lockdown, where the list of activities subject to lockdown are explicitly indicated. This could be used for flagging sectors and the statistical units active in these sectors with potentially temporary inactivity status in the business register. In those cases the absence of data should be treated as non-responses, with the corresponding adjustments in the weights..

¹ As regards to wage subsidies Eurostat published on its [webpage COVID-19 Support to statisticians a Draft note on statistical implications of some policy measures in the context of the COVID-19 as well as Labour cost statistics guidance note on the recording of government schemes related to the covid-19 crisis](#)

5. Dissemination of statistics using estimation and imputation of missing data

Information on the adopted solution should be provided as complementary information. Metadata on reliability of the values should be provided following the current [guidelines for quality reporting](#).

The measures related to the virus may have an impact on the accuracy, on the timeliness or on the completeness of data. This is somehow a trade-off that requires strategic decisions. In any case, there has to be **some communication to the users, preferably coordinated at ESS level**, and at least there should be an exchange of information of the National Statistical Authorities on the communication approach and on the priorities.

When disseminating the data, **proper flags** for imputed or estimated values should be used (see [guidelines how to use flags](#)).

When determining the strategy for business statistics, the **overall coherence with other domains** should be taken into account (i.e. IPI vs. GDP, etc). Relevant in this regard is the consideration that certain sectors, for example the production sector is locally more affected in a given country.

6. Further information and support

- [Measures to address COVID-19 impact on statistical production](#)