

Production of European aggregates

Preface

This chapter describes the production by Eurostat of EU aggregates of European business statistics and their dissemination.

The chapter is part of the online European [Business Statistics Manual](#), which provides a comprehensive description of methodologies and background information on how these statistics are produced within the European Statistical System (ESS).

Contents

1. [Introduction](#)
2. [Aggregation under perfect conditions](#)
3. [Aggregation under imperfect conditions](#)
4. [Confidentiality](#)
5. [Domain-specific measures](#)
6. [See also](#)
7. [Further Eurostat information](#)
8. [External links](#)
9. [Contacts](#)

1. Introduction

Based on the national data provided by the national statistical authorities, Eurostat calculates aggregates at EU level. These aggregates are calculated for the EU as a whole (28 countries) and — depending on the domain — for the euro area as well (19 countries).¹

This chapter focuses on aggregation methods for tabular data as detailed in [Data requirements](#).

Section 2 describes the calculation in its most basic form, under perfect conditions: full and timely availability of high-quality national data from all EU countries.

As section 3 explains, these perfect conditions are not always fully met. National data may be missing, incomplete and/or of insufficient quality. There are 2 sorts of possible reasons for this:

¹ Some domains in business statistics also publish other European aggregates, such as EU27 (historical series) and/or EFTA. The statistical disclosure control applies to both new and historical series.

- The new Framework Regulation Integrating Business Statistics (FRIBS) introduces a number of legally embedded simplifications (i.e. relaxations of general data requirements) that take into account the size of the countries/industries concerned so as to reduce the burden on respondents and national statistical authorities. Once the FRIBS Regulation enters into force, EU countries may also be granted a derogation period, during which they are exempted from the mandatory delivery of some specified data.
- There are various possible reasons at purely national level for why national data sent to Eurostat do not meet these perfect conditions.

After calculating the EU-level aggregates and validating them internally, Eurostat decides whether they can be published, taking into account their quality and the confidential nature of the underlying national data. Eurostat generally aims to make statistical aggregates at EU level as widely available as possible, by applying specific statistical disclosure control techniques that do not infringe national rules on dissemination. In doing so, it also focuses on providing EU-level aggregates for the most salient variables.

Section 4 outlines confidentiality measures at EU level and introduces the model for the Confidentiality Charter which describes the specific measures taken in the statistical domains.

The handling of revision of European aggregates is governed by the same ESS principles as revision of national data (see section 4 "Data revision policies and practices" of [Dissemination of business statistics](#)), assuming that the confidentiality pattern is not changed.

Maximisation of the availability of European aggregates as described in sections 3 and 4 is part of what is known as 'European approach to statistics'. This approach, an essential element of the European Statistical System (ESS), is based on [Article 16](#) of the general statistical regulation, 223/2009.

2. Aggregation under perfect conditions

Aggregation under perfect conditions assumes that all EU countries (H) have provided their national data (θ_h) on time and that these data are complete and meet the necessary quality standards.

In its simplest form, the European aggregate (θ_{aggr}) is calculated as

$$\theta_{aggr} = \sum_{h=1}^H \theta_h$$

For *additive* national data in absolute terms, the aggregation is merely the sum of all national data. Examples are the number of persons employed and the turnover expressed in absolute values.

For index-based data such as in short-term business statistics, the EU-level aggregate is calculated as:

$$\theta_{aggr} = \sum_{h=1}^H W_h * \theta_h \quad \text{where } W_h \text{ is the 'weight' of country } h.$$

This weight represents the country's share as a proportion of the total aggregate. Eurostat's weighting system has a dual role: it serves both geographical aggregation and activity aggregation. Each European index has its own number of specific weightings.

There are also different kinds of aggregates:

- product aggregates
- geographical aggregates (EU)
- time aggregates (monthly and annual data).

3. Aggregation under imperfect conditions

The simple aggregation formula set out in section 2 does not hold under imperfect conditions, such as when:

- A. national data are *missing (or incomplete)*: data from one or more countries are missing on account of a temporary delay, for a longer period, or permanently.
- B. national data are *unreliable*: the reason for this could be small sample sizes or low response rates, leading to variances that are too high to permit publication of the national data cell. From an EU perspective, however, these national data cells are still valuable for aggregation to EU-level totals.

Possible reasons for imperfect conditions include:

- There are several simplifications of the data requirements in FRIBS that take into account the size of national economies/business activities. Examples are the 1 % rule and the use of CETO flags (SBS, ProdCom), the reduction of reporting requirements for detailed NACE breakdowns for medium-sized and smaller countries, and the use of an EU sampling frame (STS).

The 1 % rule means there is no need to transmit national data cells representing under 1 % of the EU total (in terms of turnover and employment) to Eurostat. The CETO flags mean that, depending on the size of the country (small, medium-sized or large), a number of national cells for detailed NACE levels may be delivered solely for the purpose of 'Contribution to EU-Totals Only' and will not be published separately.

- Derogations in the context of the implementation of FRIBS: for some data series, some national statistical authorities may be exempted from delivery for a limited number of years.
- Partial non-compliance with the implementation of FRIBS' data requirements.

Where data is missing or incomplete, Eurostat's usual procedure is to estimate the missing values, purely for the purpose of calculating the EU-level aggregates. Such estimates are generally based on a variety of methods, depending on whether past data are available (forecasting methods) or whether there are data from similar countries or adjacent levels of breakdown (imputations). These methods show a close correspondence with processing methods at national level (i.e. [methods for imputing missing values](#) and [methods for calculating aggregated totals](#)).

4. Confidentiality

Eurostat is in a specific position: it can work only on European aggregates, so as to protect national figures. It does not normally remove countries' data so as to protect confidential national figures. This is because the other data may already have been published at national level and could be used to recalculate confidential figures (see below).

The legal rules on confidentiality in general are set out in Articles 20-26 of Regulation 223/2009 and further detailed by ['Statistical Disclosure Control in business statistics'](#).

The practical rules on confidentiality when publishing European aggregates composed of confidential national figures may be laid down in what are known as Confidentiality Charters. These charters are applicable to tabular data based on quantitative variables and may be adapted to domain-specific needs. The standard model for Confidentiality Charters was discussed by the Working Group on Methodology in April 2016 (see [Working Group document](#); the standard model for Confidentiality Charters is included in the annex). It was then improved further.

In general, the more detailed the data received by Eurostat, the more efficient the treatment of the statistical confidentiality of European aggregates.

Over 2017-2018, the confidentiality charters are expected to be discussed, adapted and populated at the level of the statistical domains. The specific confidentiality charters will describe methods and parameters for confidentiality treatment in a particular statistical domain of business statistics, ensuring proper documentation and transparency with regard to the methods used. Once established and approved, the domain-specific confidentiality charters will be published in 2 versions:

- For *data compilers* - full charter
- For *end-users* - same charter, but without the domain-specific exact confidentiality parameters (see explanation below).

In the field of business statistics, the confidentiality charters will be developed at domain level. It is considered impossible to apply the charter in the field of the international trade in goods and services (ITGS). This is because of the sheer amount of data and the impossibility of collecting the necessary meta-data. Moreover, there are far fewer problems with restricted dissemination in the case of ITGS, given the application of passive confidentiality and the fact that EU countries are legally obliged to publish data at least at the Combined Nomenclature's chapter level.

In general, the national statistical authorities (NSAs) apply the following common rules for primary confidentiality:

- **Threshold rule:** a data cell is confidential if the number of contributors is under a specified threshold. This threshold can vary across countries and across domains.
- **Dominance rule (n,k):** a data cell is confidential if the n largest units contribute more than $k\%$ to the cell total.
- **P-per cent rule:** a data cell is confidential if a respondent can estimate another respondent within $p\%$ of the respondent's true value.

To make these rules protective, the abovementioned qualifiers (n,k,p) are generally kept confidential for end users. In 2016 the Expert Group on SDC approved a set of non-mandatory recommendations for the confidentiality parameters (only available to data compilers).

If a confidential cell contributes to an EU-level aggregate, the aggregate has to be dealt with in such a way as to prevent disclosure. The EU rules that apply are based on the same approaches as national rules (threshold, dominance, P-per cent) and will be specified in the confidentiality charters. The following observations are specific to Eurostat:

- If no detailed information about individual contributors (statistical units) is available for national confidential figures, these figures are treated as covering one contributor (statistical unit).
- Normally only aggregates are protected, not figures inside the tables provided by national statistical authorities.

The EU aggregate is **unsafe** if:

- (a) only one national total is confidential, or
- (b) 2 national totals are confidential, and at least one has only one contributor, or
- (c) 3 national totals are confidential, and one contributor dominates the confidentiality cluster (i.e. the sum of the 3 national totals).

Cells missing from national data (as clarified in section 3, so not for reasons of confidentiality) can be used to protect confidential cells that contribute to the same aggregate.

If an **unsafe** European aggregate is failing to abide by the confidentiality rules, the aggregate itself will be suppressed in the publication (secondary confidentiality treatment). In doing so, Eurostat tries to maximise the number of highest-level EU aggregates. When transmitting national data to Eurostat, the national statistical authorities are asked to provide additional meta-information on confidentiality. This enables Eurostat to determine the confidentiality of European aggregates more efficiently.

If the national figures are revised without any changes to the confidentiality pattern, the European aggregates can be updated in line with the confidentiality approach taken in the previous release. If, however, the revision changes the national confidentiality pattern, the confidentiality of the European aggregates has to be reassessed, bearing in mind that potential intruders have access to both the original and the revised release of the data.

An alternative to keeping unsafe EU-level aggregates out of publications is to publish them in such a form (e.g. interval, highly rounded) that it is impossible to determine the real figure within a given range. This method is applied in SBS and ProdCom and a special case of Controlled Tabular Adjustment (CTA) described in [Handbook on SDC](#) (page 159).

5. Domain-specific characteristics

The domain-specific rules and practices for calculating European aggregates will be described in more detail by the **EBS-domain sections** to be developed in the course of 2017-2018. See for example the [STS prototype guide](#) especially its section on calculating EU-level aggregates. References

to these domain specific sections on EU-level aggregation will be included under this section once they have been developed.

6. See also

- Overview of methodologies of European business statistics: [EBS manual](#)
- Legal aspects related to this chapter can be found in the following [overview](#).
- [Data requirements](#)
- [Dissemination of business statistics](#)
- [Statistical Disclosure Control in business statistics](#)
- [STS prototype guide](#)

7. Further Eurostat information

- Legal base of the European approach to statistics: [Article 16](#) of the general statistical regulation, 223/2009
- [methods for imputing missing values](#)
- [methods for calculating aggregated totals](#)
- [Handbook on SDC](#)
- [Model for Confidentiality Charters](#), Eurostat April 2016 (in the annex)

8. External links

9. Contacts

For questions or comments on this chapter, please contact ESTAT-EBS-MANUAL@ec.europa.eu.