

Farm structure - estimation of missing data

Statistics Explained

Introduction This article examines how Eurostat calculates estimates for missing data in integrated farm statistics (IFS).

It forms part of an [online methodological series](#) on farm statistics.

Missing data in integrated farm statistics

With the entry into force of [Regulation \(EU\) 2018/1091](#), the implementation of [integrated farm statistics](#) (IFS) introduced important methodological changes in the collection and dissemination of agricultural data. It established a distinction between 'core data', which are collected regularly in each survey implementation and cover essential structural variables, and 'module data', which are collected less frequently and focus on specific thematic areas such as labour force or rural development. In parallel, IFS introduced a division of the target population into 2 components: the main frame includes agricultural holdings above one of the physical thresholds defined in the legislation, while the frame extension covers smaller holdings below these thresholds only when the main frame does not cover 98% of the total [utilised agricultural area](#) (excluding [kitchen gardens](#)) or 98% of the total [livestock units](#) of the country.

The legislative framework includes different data collection requirements across survey years. In census years (e.g. 2020), core variables are collected for both the main frame and the frame extension, while module variables can be collected from main frame only. In sample years (e.g. 2023), there is no obligation for data collection on frame extension, and countries differ in the extent to which they provide non-mandatory data. As a result, data availability for below-threshold holdings varies across countries and over time.

This variability leads to gaps in data availability for the entire population of agricultural holdings and poses challenges for the comparability of results. To address this issue, an estimation procedure has been developed to produce consistent estimates for below-threshold units where data are missing.

Estimation procedure for below-threshold units

The estimation procedure is based on imputation techniques applied at the level of microdata.

Two main scenarios are distinguished, reflecting the different situations in census and sample years:

- **Estimation with available core data**

This scenario typically occurs in census years (e.g., 2020), where core data for the frame extension are available, but module data for this part of the population are missing. The available core variables (e.g., region, farm type) are used as auxiliary information to define homogeneous groups and identify similar units. Missing values are imputed using a donor-based approach, whereby information is transferred from similar units.

Donors are primarily selected from previous survey years of the same country, ensuring consistency with national structures. Where necessary, additional information from other countries may be used. To enhance robustness, the procedure includes multiple imputation runs, generating alternative versions of the completed dataset. The final dataset is selected based on its ability to preserve key aggregate characteristics. This approach improves the

stability and plausibility of the results.

The completed microdata are subsequently aggregated to produce the required statistical outputs. As all outputs are derived from a single dataset, consistency across different breakdowns is ensured. The use of multiple imputations also allows for an assessment of the reliability of the estimates.

- **Estimation in the absence of core and module data**

In cases where no data are available for below-threshold holdings, the procedure first reconstructs the underlying population. The number of holdings is estimated based on trends observed in other countries and historical data for the country concerned. A synthetic dataset is then created, representing the below-threshold population. Initially, this dataset contains only basic identifiers and classification variables. Key characteristics are subsequently assigned using information from previous periods, with adjustments introduced to reflect recent developments and cross-country differences.

Once the basic structure has been established, the estimation proceeds in the same way as in the previous scenario. Missing variables are imputed using donor-based methods, and the completed dataset is used to produce consistent statistical outputs.

Dissemination of estimated farm structure data

The implementation of this estimation procedure for below-threshold units has allowed the production and dissemination of comparable country-level farm structure datasets, under [Farm Structural indicators](#). To maintain consistency across domains, the procedure is implemented within groups of related tables, referred to as clusters, which share the same target population and conceptual structure. All estimates within a cluster are derived from a common underlying dataset.

In the interest of clarity and transparency, all figures that result from the implementation of this estimation procedure are duly flagged (i): value imputed by Eurostat or other receiving agencies. Table 1 lists the countries for which estimation of below-threshold units was necessary to guarantee comparability across countries and survey years.

Countries with estimation of below-threshold units, by survey year

2020 (module data)	2023 (core and module data)
Spain	Bulgaria
Lithuania	Greece
Slovenia	Spain
	Italy
	Lithuania
	Poland
	Slovenia

Source: Eurostat



Table 1: Countries with estimation of below-threshold units, by survey year Source: Eurostat

For more detailed information on the estimation procedure for below-threshold units, please consult the document: [Estimation of “below threshold” and “totals” \(EUROSTAT, 2025\)](#).

Explore further

Other articles

- [Agriculture - Statistics Explained](#)

Thematic section

- [Agriculture](#)
- [Agricultural census 2020](#)

Publications

- [Key figures on the European food chain - 2025 edition](#)

Methodology

- [Principal quality characteristics of EU countries' farm data collections for 2023 – 2026 edition](#)
- Farm structure: ([ESMS metadata file — ef_esms](#))
- [National quality reports](#)
- [Integrated Farm Statistics Manual, 2023 survey - 2026 edition](#)
- [Principal quality characteristics of EU countries' Agricultural Censuses for 2020 - 2024 edition](#)

Legislation

- [Regulation \(EU\) 2018/1091](#) of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011
- [Commission Implementing Regulation \(EU\) 2021/2286](#) of 16 December 2021 on the data to be provided for the reference year 2023 pursuant to Regulation (EU) 2018/1091 of the European Parliament and of the Council on integrated farm statistics as regards the list of variables and their description and repealing Commission Regulation (EC) No 1200/2009
- [Commission Implementing Regulation \(EU\) 2018/1874](#) of 29 November 2018 on the data to be provided for 2020 under Regulation (EU) 2018/1091 of the European Parliament and of the Council on integrated farm statistics and repealing Regulations (EC) No 1166/2008 and (EU) No 1337/2011, as regards the list of variables and their description

External links

European Commission

- [Agriculture and rural development](#)
- DG Agriculture and Rural Development
 - [Common Agricultural Policy \(CAP\)](#)