Confidentiality in Integrated Farm Statistics with τ-Argus

Denisa Florescu, Wladimir Raymond and Rudolf Seljak

SDC-tools User Group Meeting 2025

27 February 2025

Summary

- Background

- Primary confidentiality

  - SAS

- IFS tables

- Secondary confidentiality

  - τ-Argus

- Challenges and further research

- Questions

Background

Background (1/2)

- Main aim of Integrated Farm Statistics (IFS) data collection

  - provide comparable data on agricultural holdings of EU

- Variables pertain for instance to

  - types of land, livestock species, animal housing, labour force activities etc.

- Data on farm structure collected at regular intervals since 1966

  - censuses every 10 years (e.g., 2020 and 2030)

  - sample surveys every 3 or 4 years (e.g., 2023 and 2026)

- Data are used to

  - evaluate the income of agricultural holdings

  - evaluate the impacts of the Common Agricultural Policy

  - tackle agri-environmental issues

European Commission

Background (2/2)

- Eurostat receive microdata on holdings from EU countries collected by

  - National Statistical Authorities (NSA)

- Eurostat build tabular data from microdata

- Eurostat publish aggregated data at regional, national and EU level

  - disseminated data must be useful to data users (e.g., DG AGRI)

  - data on individual respondents must not be disclosed (confidentiality)

- Primary confidentiality has been applied to tabular data (e.g., 2016 data)

- Secondary confidentiality is implemented since <u>July 2023</u>

  - WG 2022 recommendations (e.g., DE)

European Commission

Primary confidentiality

European Commission

Primary confidentiality

- Primary confidentiality is implemented in SAS

  - Build tabular data from country microdata

- **Old** confidentiality rules (up to IFS 2020)

  - threshold rule with fixed parameter $th = 4$ for all countries

  - (2,85)-dominance rule for all countries

- **New** confidentiality rules (IFS 2023)

  - threshold rule with country-specific parameter $th = \{3, 4, 5\}$

  - threshold rule is implemented for **all** countries (3 being the default)

  - $p\%$-rule with country-specific parameter $p \in [5, 25]$

  - $p\%$-rule is implemented for selected countries

European Commission

IFS tables

IFS tables

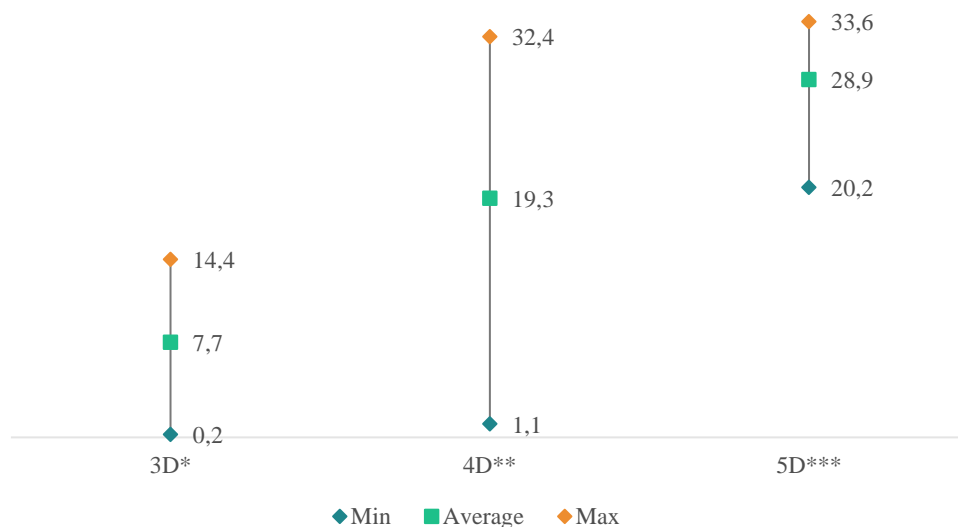| Name of table	Title	Explanatory variables	Modalities of spanning variables	Size
LSK_PIGS	Pig species	NUTS2, UAA, LSU	287*10*12	34,440
LSK_POULTRY	Poultry species	NUTS2, UAA, LSU	287*10*12	34,440
LSK_BOVINE	Bovine species	NUTS2, UAA, LSU, FODDER	287*10*12*7	241,080
LSK_SHEEP	Sheep species	NUTS2, UAA, LSU, FODDER	287*10*12*7	241,080
LSK_GOATS	Goat species	NUTS2, UAA, LSU, FODDER	287*10*12*7	241,080
LUS_ALLCROPS	Types of crops	NUTS2, UAA, SO_EUR	287*10*12	34,440
LUS_PEGRASS	Permanent grassland	NUTS2, CROPAREA, UAA, SO_EUR	287*11*10*12	378,840
LUS_MAIN	Main farmland use	NUTS2, FARMTYPE, UAA, SO_EUR	287*23*10*12	792,120
M_ORG	Production methods	NUTS2, FARMTYPE, UAA, SO_EUR	287*23*10*12	792,120
MP_MANORG	Manager characteristics	NUTS2, AGE, SEX, TRAINING, UAA	287*8*3*4*10	275,520

European Commission

Secondary confidentiality

Strategy for secondary suppression in т-Argus

- Build tabular data from microdata in SAS

- Apply primary confidentiality rules in SAS

- Load the resulting tabular input data in т-Argus (version 4.2.4 build 2)

- Define the safety status of the cells in metadata

- Use given status in т-Argus

- Run secondary suppression with the **hypercube** method

European Commission

Suppression results (1/2)

Graph 1: Percentage of secondary confidential cells (%)
Hypercube method, IFS 2020

Graph 2: Percentage of secondary confidential cells (%), excluding the 4D tables having dimension FARMTYPE
Hypercube method, IFS 2020
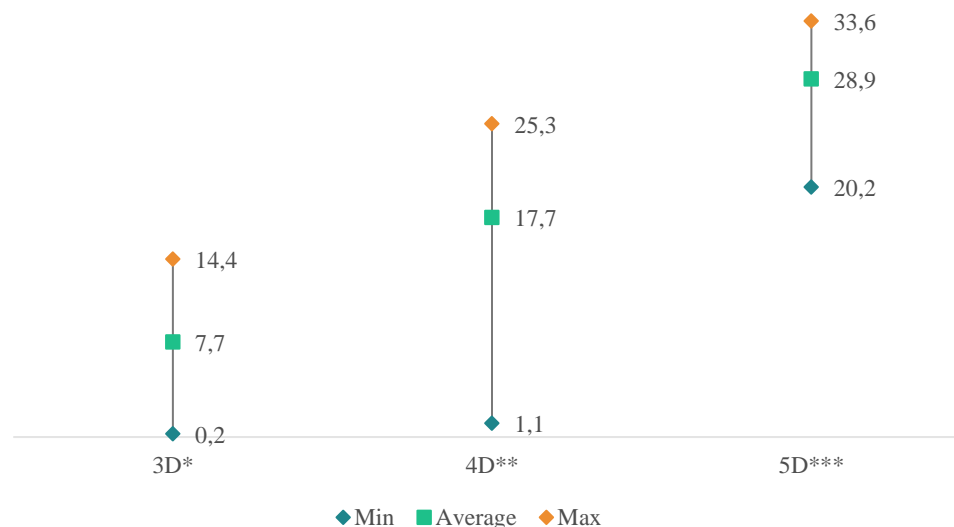

Min   Average   Max

* LSK_PIGS, LSK_POULTRY, LUS_ALLCROPS
** LSK_BOVINE, LSK_SHEEP, LSK_GOATS, LUS_MAIN, M_ORG, LUS_PEGRASS
*** The statistics are based on a single 5D table EF_MP_MANORG, as only that 5D table could be successfully treated for secondary confidentiality using the hypercube method.

Min   Average   Max

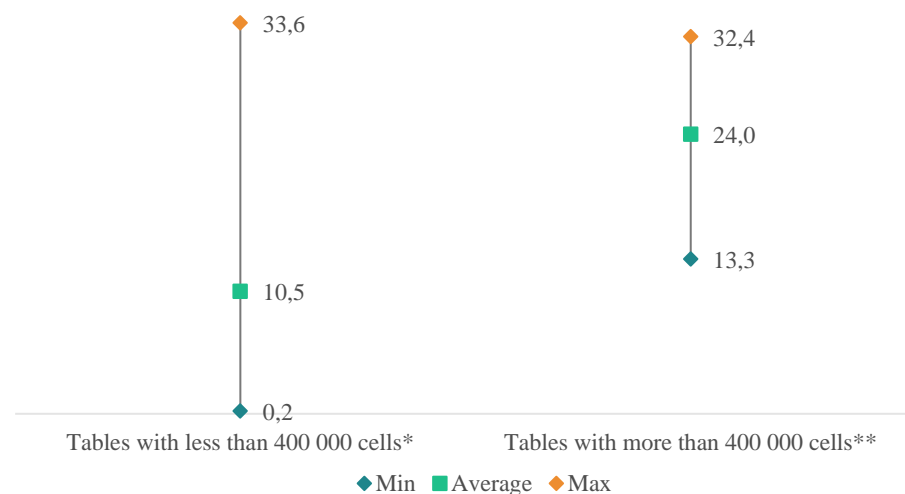* LSK_PIGS, LSK_POULTRY, LUS_ALLCROPS
** LSK_BOVINE, LSK_SHEEP, LSK_GOATS, LUS_PEGRASS
*** The statistics are based on a single 5D table EF_MP_MANORG, as only that 5D table could be successfully treated for secondary confidentiality, using the hypercube method.

European Commission

Suppression results (2/2)

Graph 3: Percentage of secondary confidential cells
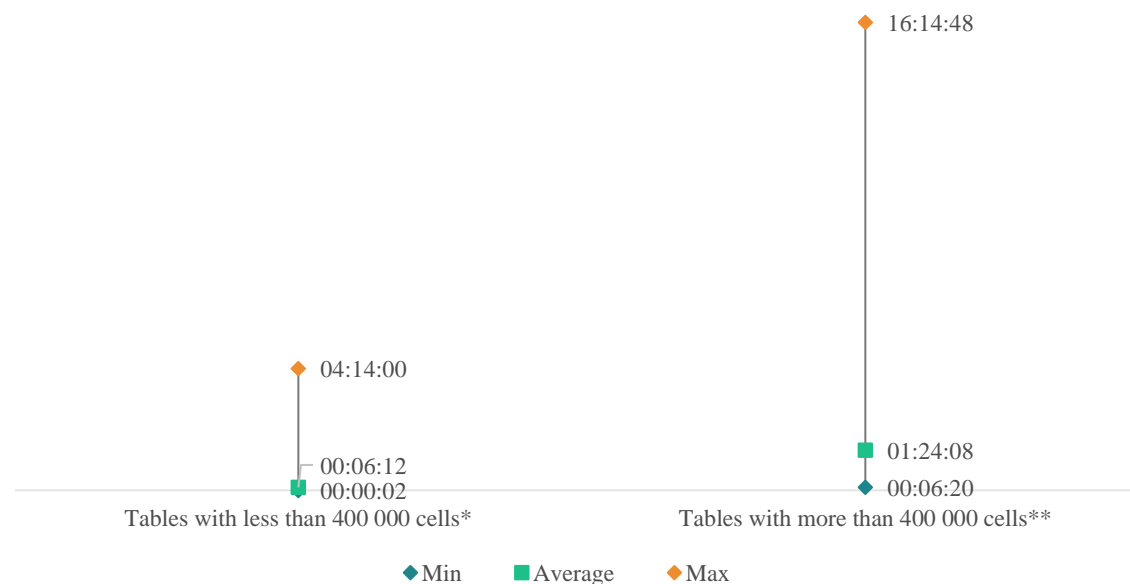Hypercube method, IFS 2020

* LSK_PIGS, LSK_POULTRY, LUS_ALLCROPS (3D)
LUS_PEGRASS, LSK_BOVINE, LSK_SHEEP, LSK_GOATS (4D)
MP_MANORG (5D)

** LUS_MAIN, M_ORG (4D)
5D tables in this category have between 800 000 and 10 000 000 cells.
Secondary confidentiality is almost impossible to achieve, considering all
30 countries. Therefore, these tables are excluded from the graph.

Graph 4: Secondary suppression time
Hypercube method, IFS 2020

* LSK_PIGS, LSK_POULTRY, LUS_ALLCROPS (3D)
LUS_PEGRASS, LSK_BOVINE, LSK_SHEEP, LSK_GOATS (4D)
MP_MANORG (5D)

** LUS_MAIN, M_ORG (4D)
5D tables in this category have between 800 000 and 10 000 000 cells. Secondary confidentiality
is almost impossible to achieve, considering all 30 countries. Therefore, these tables are excluded
from the graph.

Challenges and further research

European Commission

Challenges and further research

1. Secondary confidentiality treatment almost impossible to achieve for very granular tables (>400K cells)

2. Suppression time of large tables deemed too long by Eurostat

3. Suppression rate is too high in very granular tables

4. Secondary suppression of linked tables is a formidable task

5. Secondary suppression must account for relation among response variables

6. Table redesign (discussion ongoing with DG Agri)

European Commission

Thank you

European Commission

Questions?

European Commission