EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Technology, innovation and economic performances: a microdata integration strategy proposal

De Santis Stefano[1], Di Cecco Davide[2], Di Stefano Roberta[3]

[1]ISTAT, Rome, Italy

[2]UnitelmaSapienza, Rome, Italy

[3]Sapienza, Rome, Italy

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

ICT, Innovation and performance

- The impact of information and communication technologies, and of innovation activities on firm performance is of significant interest and topicality in the economic literature, as well as for policymakers.

- The simultaneous collection of information on these topics at an enterprise level is often limited to small scale specially conducted surveys.

- However, there are two distinct business surveys based on harmonised criteria and methodologies by Eurostat:
  - the Community Innovation Survey (CIS), which focuses on innovation (investments, impact on performance, public funding received,…) in different areas: product, process, organizational and marketing.
  - the Information and Communication Technologies Survey (ICTS) focuses on technology use.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Data Integration

- We propose and compare different integration strategies of the two business surveys to provide a longitudinal dataset with all relevant information.
- Each of these proposals corresponds to advantages, but also imply additional assumptions
- All alternatives represent zero-burden solutions to provide an integrated dataset of microdata.
- We illustrate the analysis on the Istat surveys on Italian enterprises, however, the proposals are potentially extendable to all countries involved in the surveys implementation.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

The surveys

Both surveys target enterprises in industries and services with more than 10 employees. Small difference in domain. All large enterprises with more than 250 employees are included by design.

- ICTS designed for ~ 35k enterprises **each year** (~19k effective size).

  Main source for the Digital Economy & Society Index (DESI).

- CIS designed for ~ 39k enterprises **every three years** (~25k effective size).

  Employed by the European Commission to monitor the level of innovation and competitiveness across the EU, to develop indicators on science and technology utilized into the European Innovation Scoreboard, and in analysing EU countries' research policies and their effect on economy.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Intersection

Previous examples of integration (OECD 2021) analyse the intersection CIS & ICTS over the most recent year of ICTS. The effort focuses on recalibrating the sampling weights.

- The intersection covers only 15% of the joined sample (~5k in 2016)
- Large enterprises are overrepresented
  - all large enterprises (250+) are included in both surveys by design
  - to reduce the burden on small enterprises, they are less likely to be involved in multiple surveys simultaneously
- Large companies are on average more likely to innovate and use technology

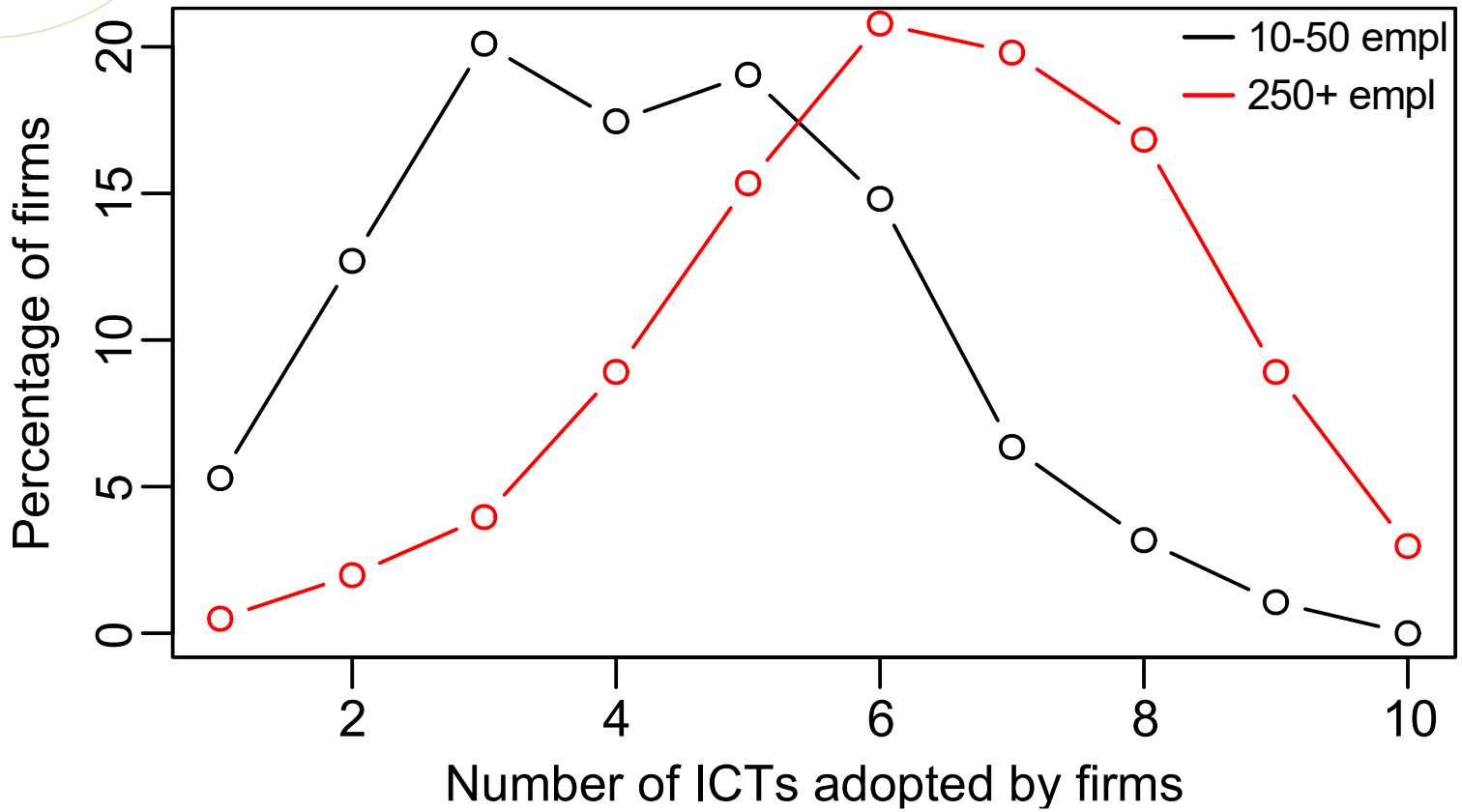
Gierten et al. (2021). Firms going digital: Tapping into the potential of data for innovation. OECD Digital Economy Papers, No. 320, OECD Publishing

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Intersection

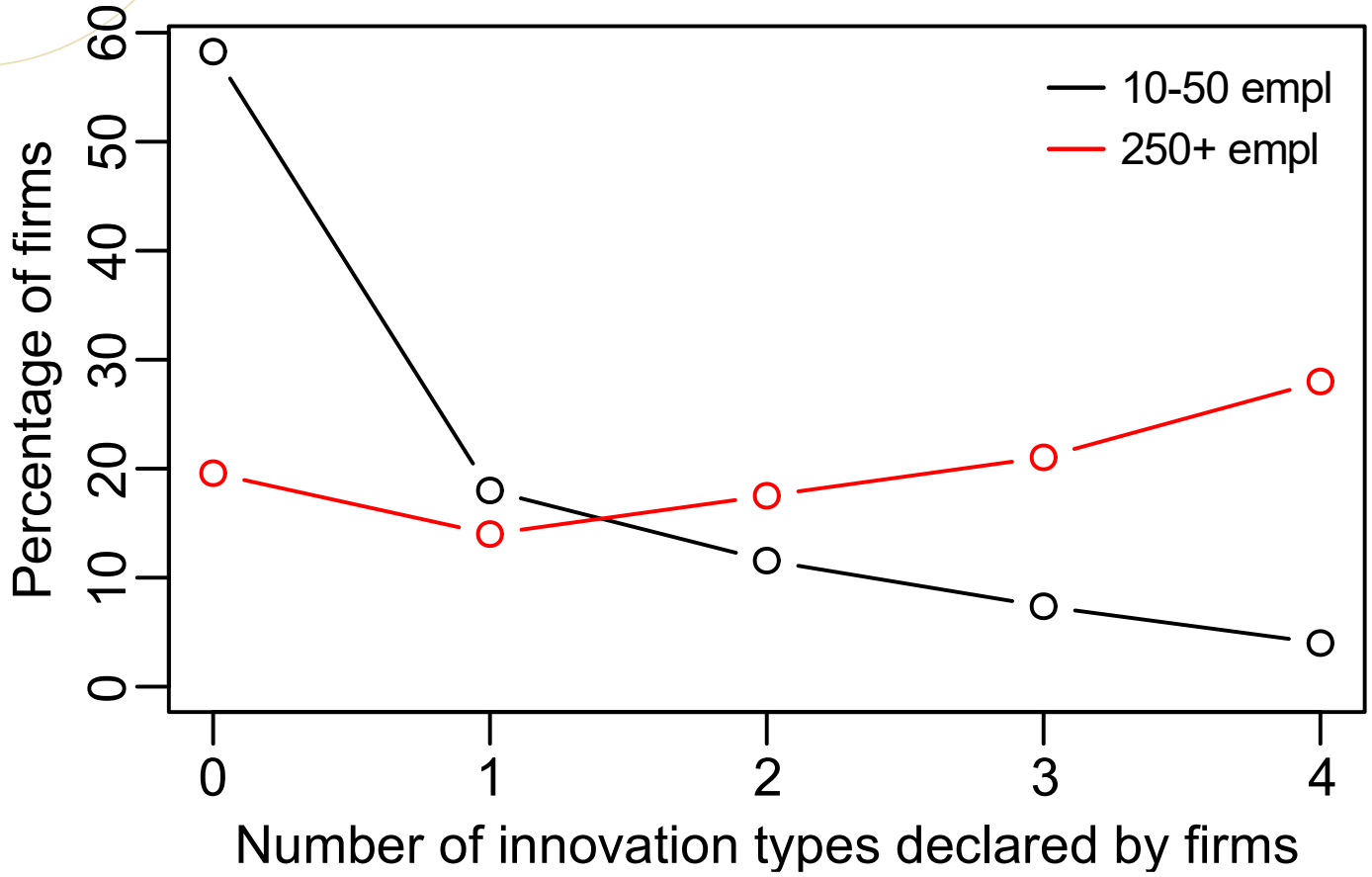


INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Intersection



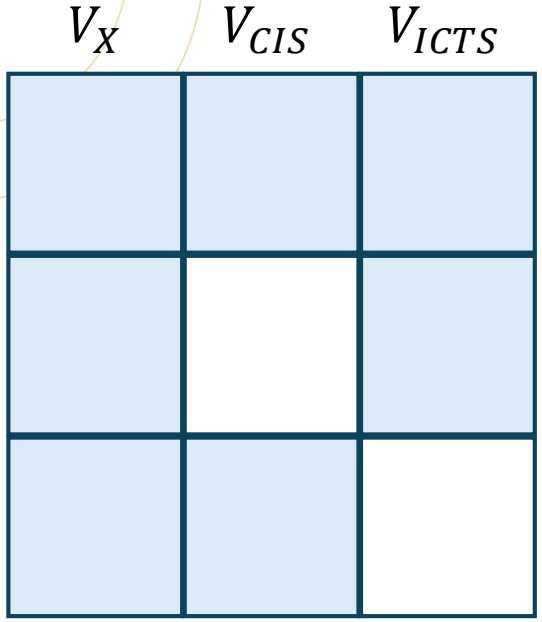
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed
by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Statistical Matching



5,335 enterprises in common

13,543 only in the ICTS

20,092 only in the CIS

V_X = structural variables from the Business Register

V_{ICTS} = 12 variables for the construction of the Digital Intensity Index

V_{CIS} = innovation/not product, process, organizational and marketing

We impute the missing data of the units covered by just one survey with SM methods.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Statistical Matching

We selected NACE, number of employees and revenue as matching variables V_X. A set of donors is selected for each recipient and one is selected at random according to weights.

We compared two non-parametric methodologies with similar results:

- Hot deck Nearest Neighbour Donor - Mahalanobis distance
- Predictive Mean Matching

Both based on conditional independence assumption (CIA):

$$P(V_{CIS}, V_{ICTS}|V_X) = P(V_{ICTS}|V_X)P(V_{CIS}|V_X)$$

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Linking multiple years

We integrate the CIS with the corresponding three ICTS waves.

- The most recent data will be given priority in cases where multiple ICTS interviews have been conducted with the same responding unit.
- Good increase in the intersection coverage (~10k)
- Increase in the number of small enterprises included
- We implicitly assume that ICT investments are not reversible. (Reasonable, as an investment done in a given year is unlikely to be dismantled in relation to the sunk costs already incurred)

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Linking multiple years

- An analysis conducted to assess the CIA, revealed that V_{ICTS} and V_{CIS} remain statistically associated even after accounting for NACE classification, number of employees and revenue, particularly for small enterprises (10-49).
- The increased number of common units deriving from linking three ICTSs allows us to relax the CIA.
- We adopted the approach presented in Singh et al. (1993) to exploit this information.
- The matching phase is followed by a recalibration of the CIS weights.

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Singh et al. (1993). Statistical matching: use of auxiliary information as an alternative to the conditional independence assumption. Survey Methodology

	N	*Product innovation* Mean	*Process innovation* Mean	*Organisational innovation* Mean	*Marketing innovation* Mean	*Digital intensity index* Mean
			Year 2012			
Sample	18,697	24.47	25.58	31.25	29.56	34.24
Single year linkage	3,375	24.48	25.85	29.43	29.95	36.33
3 years linkage	6,294	22.37	25.14	31.92	28.93	35.64
Stat Matching	6,294	24.47	25.58	31.25	29.56	34.45
			Year 2014			
Sample	17,532	20.67	21.08	22.49	22.45	33.03
Single year linkage	4,707	18.85	21.46	20.25	24.2	34.42
3 years linkage	9,592	21.79	22.65	22.97	22.45	35.96
Stat Matching	9,592	20.67	21.08	22.49	22.45	33.43
			Year 2016			
Sample	21,127	26.69	26.96	27.11	23.78	33.44
Single year linkage	5,335	25.73	24.5	26.95	22.4	35.14
3 years linkage	10,408	27.96	27.43	26.59	23.14	35.01
Stat Matching	10,408	26.69	26.96	27.11	23.78	33.54