CompNet The Competitiveness Research Network

Improving the European 'firm-level' infrastructure: The CompNet dataset

Filippo di Mauro (*)
Founder and Chairperson of CompNet (www.comp-net.org)

(*) We are grateful to Alessandro Zona Mattioli and Marco Matani for their excellent help

Introduction and motivation

The **EU DATA SPACE** really need to improvements

- Situation on firm-level data:

➔ Lots of data sealed at the National Statistical institutes (NSI) sparsely utilized

➔ Two main problems:

1. **Micro data ACTUAL AVAILABILITY** in Europe still very **unsatisfactory**. SHAME,

- **Overall problem** …. **Confidentiality**..
- **But** few EU countries provide very efficient (and secure) data access (FRA, NLD)

➔ many more are just 'planning' better access. We need to speed up

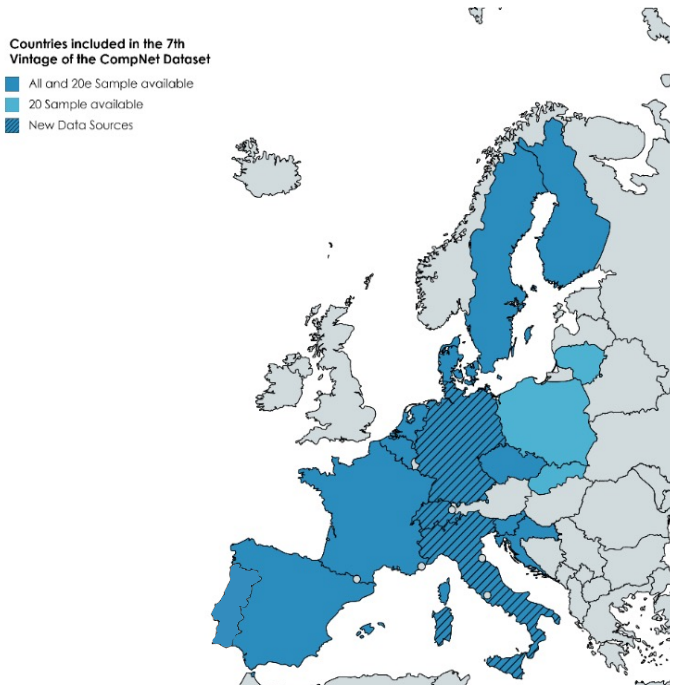
2. **Cross-EU country COMPARABILITY** is still VERY unsatisfactory. **Existing alternatives** are:

- **Orbis** …commercial direct access to raw MicroData which happen to be available…. obviously NOT the best data individual countries could offer
- **CompNet**, best data sources at national level (strictly kept in their vaults) used to run common codes,

  ➔ results are sector aggregated indicators. See below

The CompNet Dataset

The CompNet Dataset



- Started 10 years ago from research departments of the ECB/Eurosystem
- Unbalanced panel of **micro-aggregated** productivity indicators for 23 European countries, 1999-2020/21 (9° vintage)
- Our data providers are **national statistical institutes**, **national central banks**, and **governmental research institutions**
- **They** run our codes on the best existing national datasets (business registers and tax returns), which are cross-country harmonized, with no need to undertake new and costly data collection efforts
- **Full respect of data confidentiality**: the CompNet team is **never directly handling any confidential microdata at the firm-level**, but only the final aggregated output delivered by the data providers

Countries	Aggregation levels
Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland. **New**: Latvia, Malta, Lithuania, UK	Country, Macro-Sector, Macro-Sector-Size-Class, 2-digits NACE Industry, NUTS Region, Technology Class, and Age

The CompNet Dataset – Confidentiality Guarantee

- The CompNet code includes a **specific routine** that checks the eventual output cells to guarantee that **no individual firm can be identified**

- The specific routine bases on **two different thresholds**:
  - **Minimum Number of Observations**: If a cell is based on a limited amount of underlying micro-observations, the cell will be dropped;
  - **Statistical Dominance**: The largest permissible size share a single observation can attain for a given cell. If it is overcome, the cell will be dropped.

- **The data providers freely set both thresholds *a priori*** to satisfy their country or institution-specific requirements. The thresholds apply to each single cell.
- The **comparability** of all data points actually published is **not affected**.

CompNet Dataset 9th Vintage - Variables

Productivity	Financial	Trade	Competition	Labour
Labour productivity	Investment ratio	% permanent exp.	Price-cost margins	% firms that change employment between t and t+3 (t+1)
VA and revenue TFP; various estimation techniques	RoA	% sporadic exp.	Mark Ups –various estimation techniques	Share of high-growth firms
ULC	Cash holdings	Export intensity	Herfindahl index	Job creation and job destruction rates
Firm size	Leverage	Characteristics of top exporters	Concentration of sales in top 10 firms of a sector	Wage premium (proxy for human capital)
Capital Intensity	Financing gap	Productivity premium of exporters		Firm entry and exit
Marginal revenue productivity of inputs	Collateral	Characteristics of firms that export AND import		
Static and dynamic allocative efficiency	Equity to Debt	Exports by destination		
Energy cost	Cash flow			
	Interest coverage ratio			
	Trade Credit/Debt			
	Debt burden			
	Credit constraint index			
	Share of "distressed" firms			

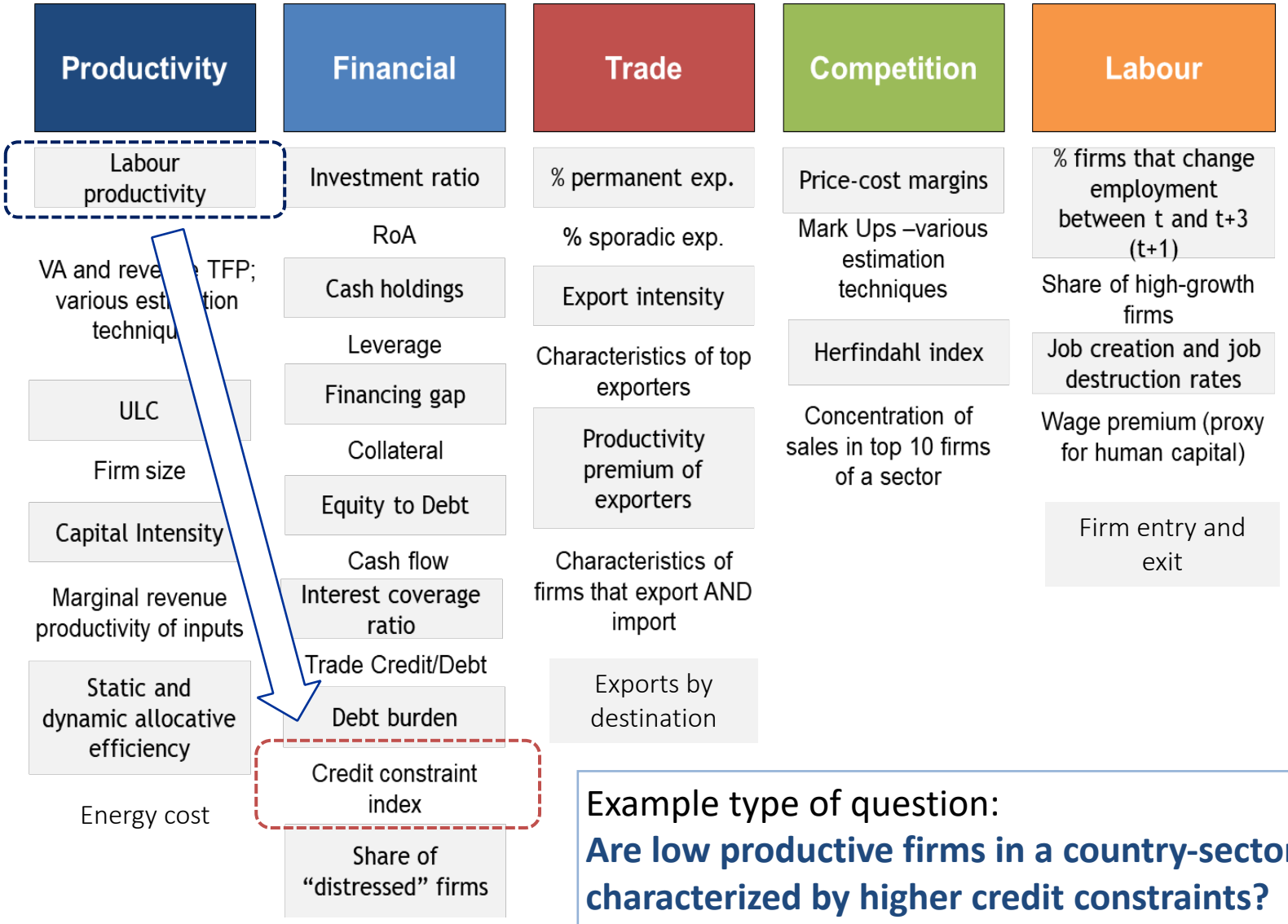
Previously we had added

- Zombie firms
- Regional dimension
- Intangibles proxies

In the current we are adding:

- Business dynamism
- More energy variables

CompNet Dataset 9th Vintage - Joint Distributions



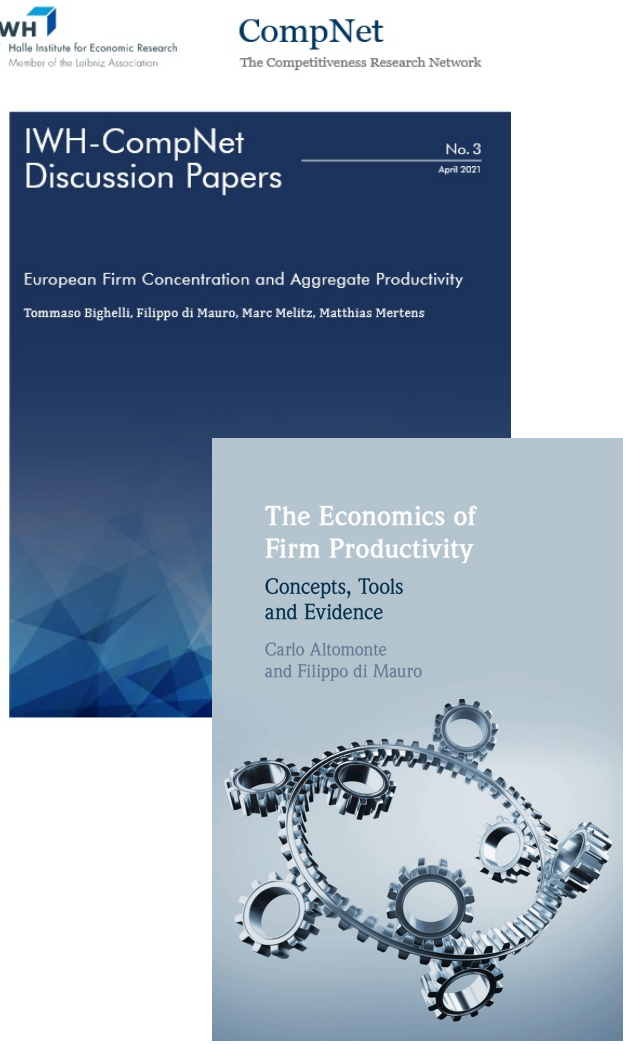
Example type of question:

Are low productive firms in a country-sector characterized by higher credit constraints?

CompNet is used widely for policy and research work 1/2

- We are working on several research projects related to *competitiveness*, *productivity*, and *firm performance*:
  - ***"European Business dynamism, market power and technology"*** Javier Miranda, Matthias Mertens, Sergio Inferrera.
  - **"Labor Market power and Wage Inequality"** Matthias Mertens.
  - **"Input Market Power"** Matthias Mertens, Bernardo Mottironi, Alessandro Zona Mattioli, Filippo di Mauro.
  - **"Productivity and pay"** Paolo Mengano, Matthias Mertens, Tommaso Bighelli.
  - ***"Intangible investments and productivity"*** Eric Bartelsman, Mirja Hälbig, Alessandro Zona Mattioli.
- We produce periodically policy output (VOX EU and alike) to market our research

CompNet is used widely for policy and research work 2/2



- Latest is Melitz et al. (2022) on Firm concentration and resource reallocation (JEEA)

- Book on "The Economics of Firm Productivity"

CompNet going forward – additional services and events

Improved and easier to use datasets

- Reduced (CompNet) dataset, carefully 'cleaned', with easy to access interfaces **aimed mostly at policy users**
- Improving the quality of the datasets via continuous discussion with providers and identification of the best ones available country by country

Training and application workshops

- Within the Commission grant we are increasing our offer of **training** on the use of microdata, at different levels in accordance to needs, as well as
- organizing **periodical workshops** on applications of our datasets for research/policy purposes.

Conferences

- **12° CompNet Annual Conference** - hosted by the European Commission, 19-20 October 2023
- **3° Finance and Productivity Conference (FINPRO)** - hosted by Bank of Italy, 8-9 June 2023

The new CompNet MicroData Infrastructure (MDI)

MDI – Conceptual framework

- In recent years there has been some **progress in harmonizing micro-level data**, for example by regulations on Business Registers (Regulation (EC) No 177/2008) and surveys on ICT usage in business (Regulation (EC) NO 808/2004), as well as by Eurostat model questionnaires, e.g. for the Community Innovation Survey (with voluntary participation).
- With the **Business Register as a "backbone"**, the MDI intends to assist NSIs (of a larger number of EU countries as possible) to link information from these datasets and other survey or register-based information at the individual enterprise-level (here loosely referred to as "firm-level").
- The result will be an **incredibly rich set of information** which will allow researchers (in academic and policy institutions) to customize their queries on firm-level data.
- The two **key differences** between MDI and CompNet are:
  - Traditional **CompNet** output consists of **harmonized statistics aggregated from micro-level data** (e.g., average, median, p10, p90, … for firms in industry X of country Y);
  - **MDI** allows researchers to run their codes directly on firm-level data. The code run is managed from the CompNet Team.
- So far, **six NSIs** have been involved in the initiatives (FRA, NLD, SWE, NOR, FIN, DNK) and other three are about to join (SVN, PRT, AUT). In Germany progresses are slow but constant.

MDI – Motivation

To do research, a **Researcher** has to go through:

1. Research proposal
2. Data access procedure
3. NSI technical restrictions
4. Fixed costs of getting to know the data
5. Decisions on data handling (e.g. outliers treatment)
6. Programming the code

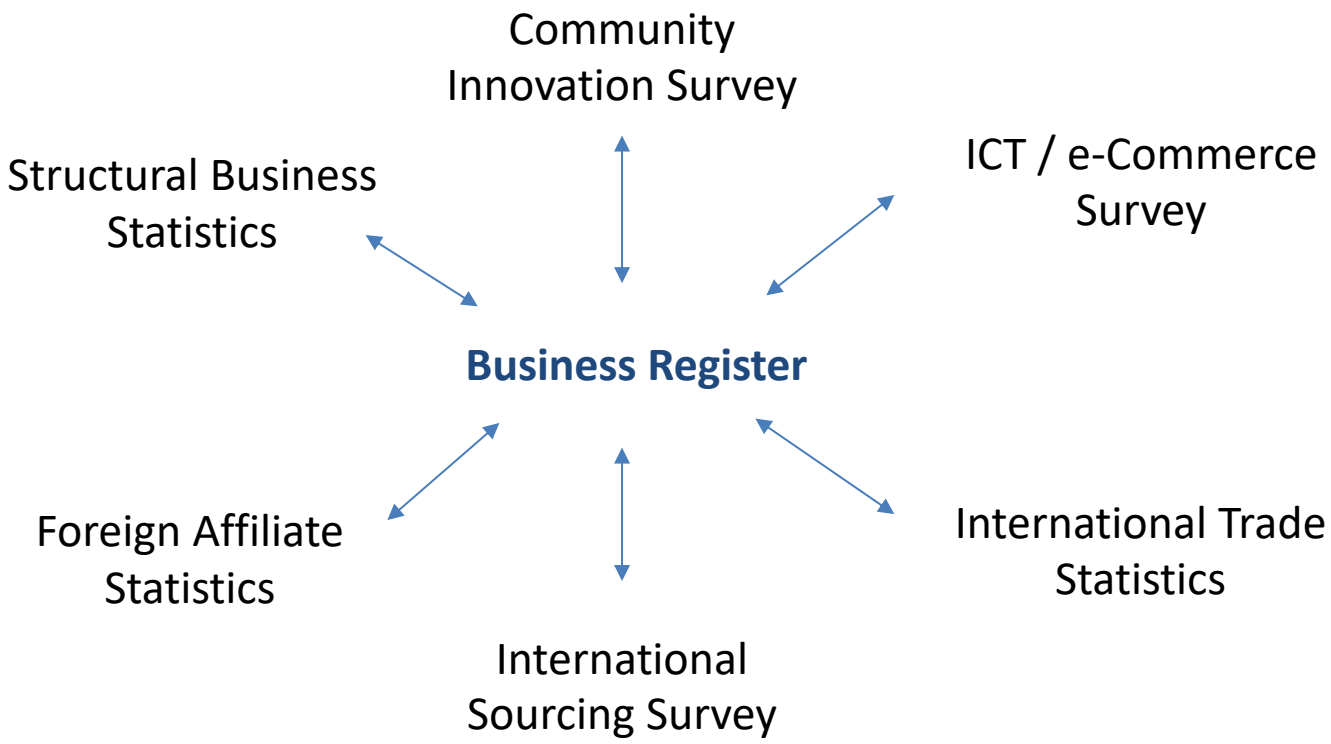
The **NSI**, for each research project has to:

1. Provide access procedures
2. Provide dedicated assistance
3. Provide Microdata research support
4. Do a disclosure analysis
5. Sometimes host researchers

The resulting disclosure-free output will be used in a publication, **virtually impossible to replicate in other countries** after many hurdles!

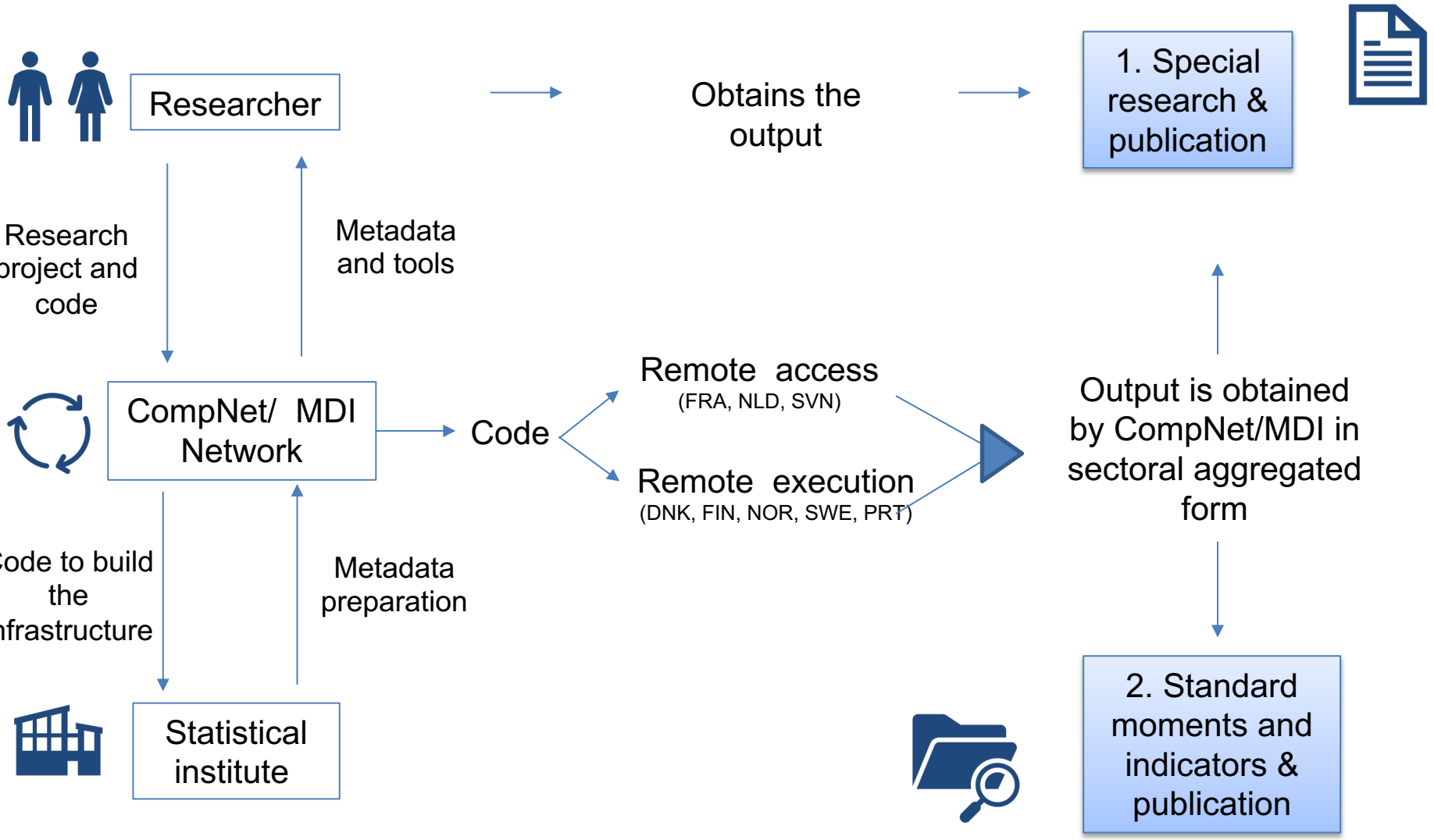
MDI – General Setup

Homogeneized set of **datasets** (sourced from National Statistical Institutes):



- For each dataset, **same list of indicators** with **Firm-level** data.
- **Time coverage**: 2007-2017, 2000-2020 for some countries.
- **Additional Datasets** are about to be added (Prodcom, Energy Consumption Survey, ...)

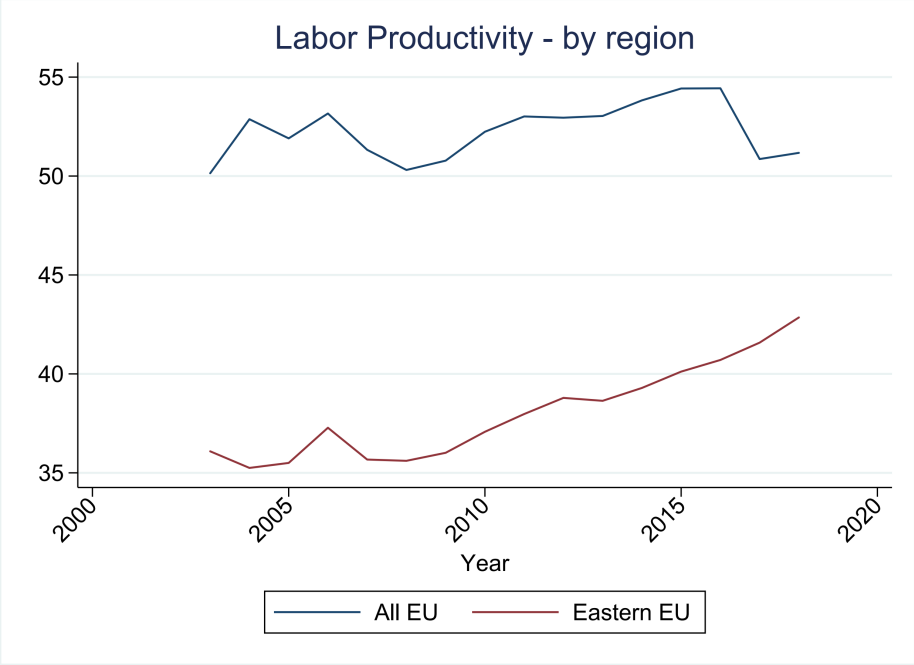
MDI – Architecture



Examples of Applications

Application 1: Labor productivity – by region

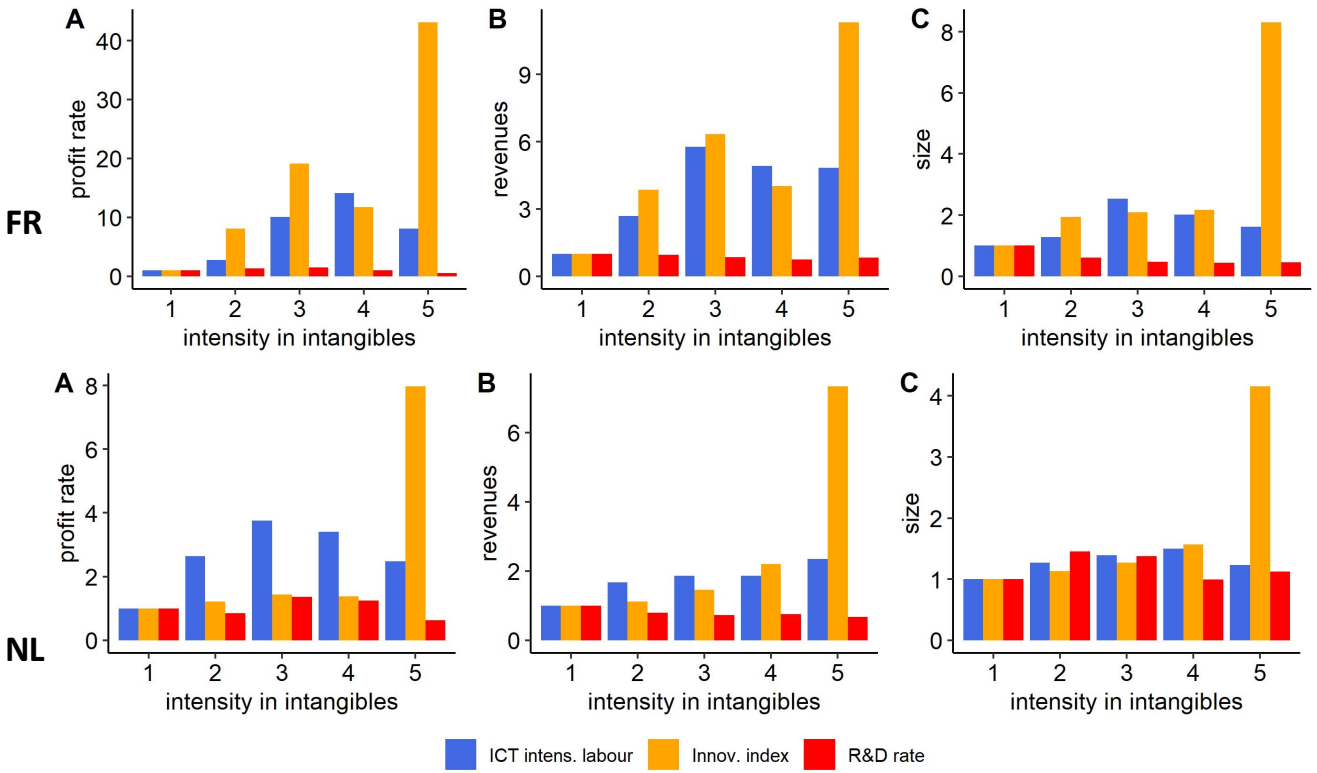
1) Productivity convergence in Eastern Europe



- Using the **CompNet dataset**, labor productivity shows a *convergence* between **Eastern EU** and the **EU average**, while maintaining similar growth in GDP per capita.
- This is happening at both the average **regional level** and at the **individual country levels**.

Application 2: More granular data (MDI) give a more nuanced picture

2) Intangibles and productivity



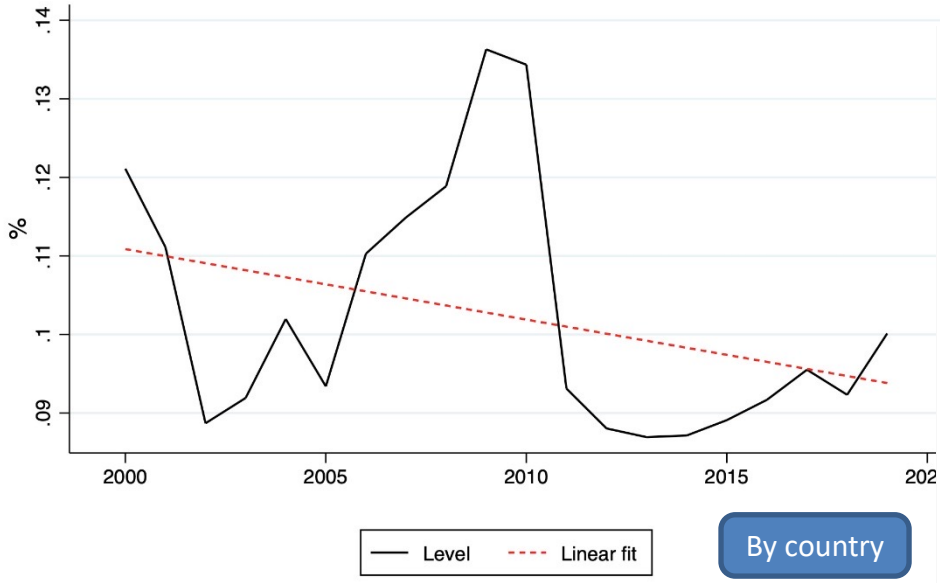
Firms results **(profit, revenues)** are highly skewed when using **innovation index** as a proxy for intangibles

Alternative measures of intangible assets are available (software and IPR investments, R&D, innovation strategies, ICT usage…)

Application 3: Business dynamism is declining

3) Business dynamism in the EU

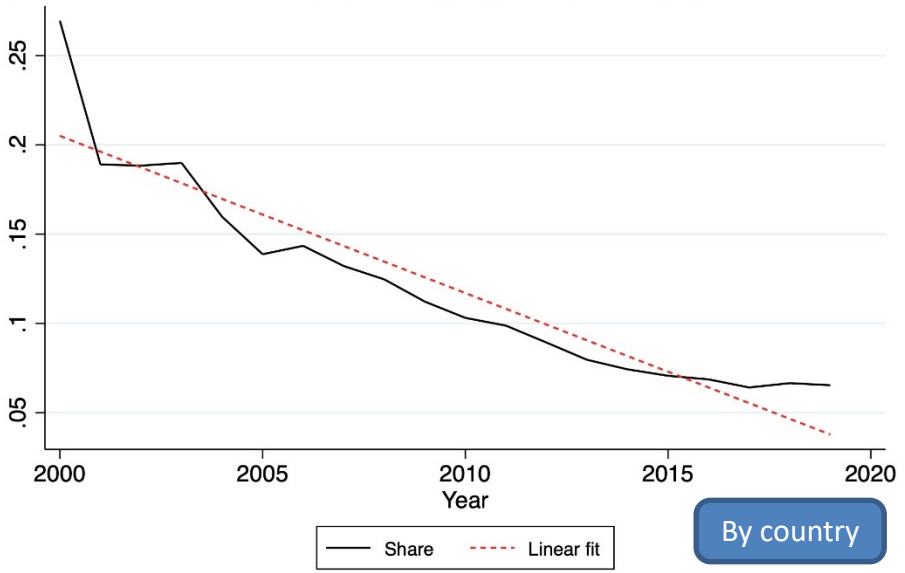
Job Dynamism in Europe – 20e sample



By country

Average weighted by total number of employees in a sector. Secotrs considered:
Manufacturing, Construction, Wholesale and retail trade, Transportation and storage, Accommodation and food service activities, Information and communication, Professional, scientific and technical activities, Administrative and support service activities

Share of employment in young firms over total employment – 20e sample

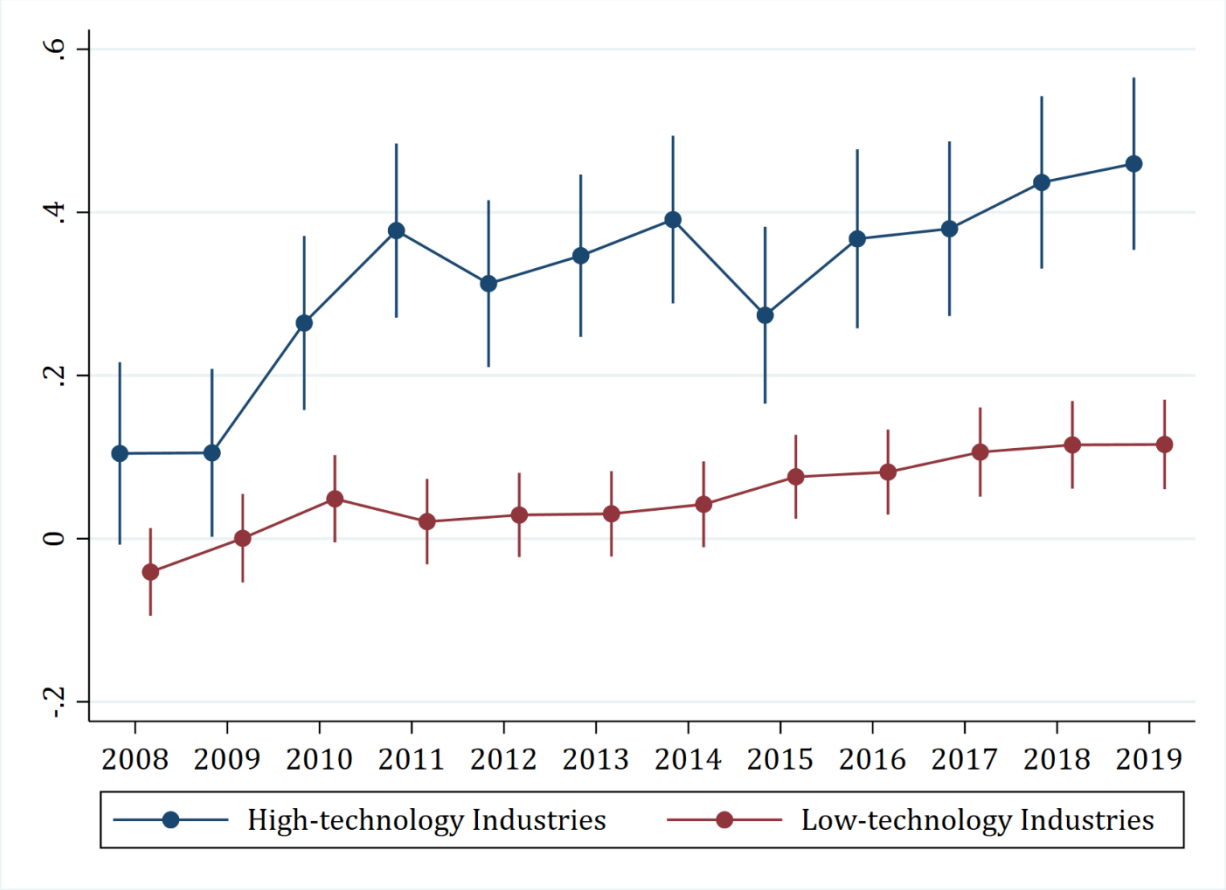


By country

- **Job dynamism** (sum of job creation and job destruction) is declining across the board
- **The importance of young firms** in terms of employment is falling, too

Application 4: Technology and productivity

4) High- vs Low-technology industries



- In **high-technology industries** productivity growth is more sustained

Application 5: Technology and labor

5) High- vs Low-technology industries

TECNOLOGY INTENSITY AND LABOR-DIGIT SECTOR ANALYSIS

	Log Avg Wage		*Log ICR*	
Technology class	0.117***	0.122***	0.099**	0.103***
	(0.00)	(0.00)	(0.01)	(0.03)
log(*Avg. Firm Size*)	0.038	-0.021	-0.23**	-0.21*
	(0.03)	(0.02)	(0.08)	(0.07)
Capital Intensity		0.153***		-0.084
		(0.031)		(0.07)
log(*Avg. Firm Age*)			-0.79***	-0.70***
			(0.013)	(0.16)
Time FE	YES	YES	YES	YES
Country- FE	YES	YES	YES	YES
Observations	6,491	6,460	4,772	4,770
# of sectors	23	23	23	23
R-squared	0.953	0.946	0.941	0.944

- In **high-technology industries both** wages and job creation rates are higher

Conclusions

- Granular datasets are essential to **tackle issues** such as **market power** and **productivity prospects**.
- In Europe however access to such info is *de facto* limited only **at the national level**
- CompNet provides **Micro aggregated information** which uses the best national data sources and ensures **cross country comparability** (we talk al lot with our NSIs)
- Our MDI now available for 6 countries has enormous potential

➔Periodically we can run dedicated codes on previously **harmonised datasets**.

- **CompNet is thriving** …thanks for your support and great contribution to the common work
- **Thank you for your attention!**

APPENDIX 1

Country Coverage

Country coverage

COUNTRIES AND YEARS	
Country	Years
Belgium	2000 - 2020
Croatia	2002 - 2021
Czech Republic	2005 - 2020
Denmark	2001 - 2020
Finland	1999 - 2020
France	2003 - 2020
Germany	2001 - 2020
Hungary	2003 - 2020
Italy	2006 - 2020
Latvia	2007 - 2017
Lithuania	2000 - 2020
Malta	2010 - 2020
Netherlands	2007 - 2019
Poland	2002 - 2020
Portugal	2010 - 2020
Romania	2005 - 2020
Slovakia	2000 - 2020
Slovenia	2002 - 2021
Spain	2008 - 2020
Sweden	2003 - 2020
Switzerland	2009 - 2020
United Kingdom	2005 - 2020

Source: CompNet 9th Vintage, 20e weighted dataset
Note: Countries highlighted in red are new additions in the 9th Vintage.

APPENDIX 2
Data Sources

Business register

Statistical Business Register (BR) The statistical business register (BR) plays a central role in the production of business statistics and is the starting point for establishing statistical survey frames. The BR contains information on identifying characteristics such as ID numbers, names and addresses, demographic characteristics, economic activity, legal form and institutional sector code as well as information on control and ownership relations for enterprises, their local and legal units and enterprise groups. In MDI, the BR serves as a 'backbone' or connection between various surveys and administrative datasets.

← BACK

Structural Business Statistics

Structural Business Statistics (SBS) The Structural Business Statistics (SBS) describe the economic activities within the business economy, including industry, construction, distributive trade and services. SBS indicators at the detailed sector level are transmitted to Eurostat and published by all European Statistical System (ESS) members (EU Member States, Norway and Switzerland, some candidate and potential candidate countries). Harmonization of the SBS has taken place regarding the detail and coverage of the sectors (now NACE 2.1) and the statistical definition of the transmitted indicators (Commission Regulation (EC) No 250/2009). Generally, the SBS indicators in each country are collected at the level of individual enterprises engaged in economic activity.

← BACK

Community Innovation Survey

Community Innovation Survey (CIS) The Community Innovation Survey (CIS) is part of the EU science and technology statistics and provides mostly qualitative information on firm innovative activity. Surveys are carried out every two years by EU member states and a number of ESS member countries on a voluntary basis. The harmonized survey contains information on the types of innovation and various aspects of the development of an innovation, such as the type of funding and innovation expenditures. The CIS covers both innovation outputs and the innovative process and inputs (type of funding, R&D expenditure) and distinguishes four innovation types: process, product, organizational, marketing, thus covering both innovative property as well as capabilities and organizational capital. Additionally, the CIS asks about the novelty of the innovation, i.e. whether it is new for the market, new to the country, developed by the firm or was adopted, and thus provides information about the innovative value.

← BACK

ICT usage/ E-commerce Survey

ICT usage/ E-Commerce Survey (ICTEC) The Community survey on ICT usage and e-commerce in enterprises is an annual survey conducted since 2002, which collects information on the use of information and communication technology, the internet, e-government, e-business and e-commerce in enterprises. Like the CIS, the EC survey contains mostly qualitative data. The ICT use survey measures various dimensions of firm technology use. Besides software and databases being considered as an integral part of intangibles, the adoption of certain technologies also provides information about firms' organizational capital and ICT capabilities both in the firms' internal operations and regarding the firms' supplier and buyer relationships. The qualitative information in the survey can be used to construct an ICT intensity index which allows for variation in the underlying source variables, thereby overcoming the issue with changing survey questions and the saturation of certain variables over time

← BACK

International Trade Statistics

International Trade Statistics Firm-level statistics concerning exports and imports are the International Trade in Goods Statistics (ITGS) and International Trade in Services Statistics (ITSS). International trade in goods statistics (ITGS) measure the value and quantity of goods traded between EU Member States (intra-EU trade) and goods traded by EU Member States with non-EU countries (extra-EU trade) broken down by types of goods (Combined Nomenclature) and by partner countries. The providers of statistical information differ between intra and extra EU-trade. In the first case, it corresponds to all taxable persons reporting transactions exceeding a certain threshold fixed by member states; in the second one, it corresponds to administrative data from the customs declarations lodged by natural or legal persons in the customs administration. International Trade in Services Statistics (ITSS) typically cover trade in services, i.e. transactions paid for the services that have taken place between the residents and non-residents.

← BACK

Foreign Affiliate Statistics

The Foreign Affiliate Statistics (FATS) is distinguished into inward FATS, i.e. the activity of foreign affiliates resident in the compiling country, and the outward FATS, that is, the activity of foreign affiliates abroad but controlled by the compiling country. The FATS allows to qualitatively assess the degree of economic activity of a domestic enterprise abroad and identify foreign-controlled firms.

← BACK

International Sourcing Survey

The International Sourcing Survey (ISS) gathers data on international organisations and sourcing of business functions in 16 European countries, covering the period 2014-2016 or 2015-2017, depending on the country. The survey results cover nearly 60,000 businesses each with more than 50 persons employed. However, since the survey is still in pilot stage, the survey design varies across countries.

← BACK

Balance Sheet and Income Statement data

Balance Sheet and Income Statement data (BS) are generally sourced from the tax authority, as it leverages on compulsory tax declarations firms submit before paying their taxes. They normally cover the universe of companies, unless some small firms rely on personal taxes instead of corporate taxation, which is possible in most European countries. Data harmonization is provided by a standard reclassification of the most common items of the Balance Sheet/Income Statement developed by the CompNet team in its almost 10 years of work, in strong collaboration with the National Statistical Offices.

← BACK