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

- Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland.

**New:** Latvia, Malta, Lithuania, UK

### Aggregation levels

- 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

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

**Previously we had added**
- Zombie firms
- Regional dimension
- Intangibles proxies

**In the current we are adding:**
- Business dynamism
- More energy variables
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.
  ▪ “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)
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.
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!
Homogeneized set of **datasets** (sourced from National Statistical Institutes):

- **Community Innovation Survey**
- **ICT / e-Commerce Survey**
- **Business Register**
- **Structural Business Statistics**
- **Foreign Affiliate Statistics**
- **International Trade Statistics**
- **International Sourcing Survey**

- For each dataset, **same list of indicators** with **Firm-level** data.
- **Additional Datasets** are about to be added (Prodcom, Energy Consumption Survey, ...)

---

**CompNet** The Competitiveness Research Network

www.comp-net.org
1. Special research & publication

Output is obtained by CompNet/MDI in sectoral aggregated form

2. Standard moments and indicators & publication

Researcher

Obtains the output

CompNet/ MDI Network

Code

Metadata and tools

Research project and code

Remote access (FRA, NLD, SVN)

Remote execution (DNK, FIN, NOR, SWE, PRT)

Code to build the infrastructure

Metadata preparation

Statistical institute

15

www.comp-net.org
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...)

CompNet The Competitiveness Research Network www.comp-net.org
Application 3: Business dynamism is declining

3) Business dynamism in the EU

- **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
5) High- vs Low-technology industries

<table>
<thead>
<tr>
<th>Technology Intensity and Labor-Digit Sector Analysis</th>
<th>Log Avg Wage</th>
<th>Log ICR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology class</td>
<td>0.117***</td>
<td>0.122***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>log(Avg. Firm Size)</td>
<td>0.038</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Capital Intensity</td>
<td>0.153***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>log(Avg. Firm Age)</td>
<td>-0.79***</td>
<td>-0.70***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Time FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Country- FE</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Observations</td>
<td>6,491</td>
<td>6,460</td>
</tr>
<tr>
<td># of sectors</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.953</td>
<td>0.946</td>
</tr>
</tbody>
</table>

- 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 a lot 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

<table>
<thead>
<tr>
<th>Country</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>2000 - 2020</td>
</tr>
<tr>
<td>Croatia</td>
<td>2002 - 2021</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>2005 - 2020</td>
</tr>
<tr>
<td>Denmark</td>
<td>2001 - 2020</td>
</tr>
<tr>
<td>Finland</td>
<td>1999 - 2020</td>
</tr>
<tr>
<td>France</td>
<td>2003 - 2020</td>
</tr>
<tr>
<td>Germany</td>
<td>2001 - 2020</td>
</tr>
<tr>
<td>Hungary</td>
<td>2003 - 2020</td>
</tr>
<tr>
<td>Italy</td>
<td>2006 - 2020</td>
</tr>
<tr>
<td>Latvia</td>
<td>2007 - 2017</td>
</tr>
<tr>
<td>Lithuania</td>
<td>2000 - 2020</td>
</tr>
<tr>
<td><strong>Malta</strong></td>
<td>2010 - 2020</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2007 - 2019</td>
</tr>
<tr>
<td>Poland</td>
<td>2002 - 2020</td>
</tr>
<tr>
<td>Portugal</td>
<td>2010 - 2020</td>
</tr>
<tr>
<td>Romania</td>
<td>2005 - 2020</td>
</tr>
<tr>
<td>Slovakia</td>
<td>2000 - 2020</td>
</tr>
<tr>
<td>Slovenia</td>
<td>2002 - 2021</td>
</tr>
<tr>
<td>Spain</td>
<td>2008 - 2020</td>
</tr>
<tr>
<td>Sweden</td>
<td>2003 - 2020</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2009 - 2020</td>
</tr>
<tr>
<td><strong>United Kingdom</strong></td>
<td>2005 - 2020</td>
</tr>
</tbody>
</table>

Source: CompNet 9\textsuperscript{th} Vintage, 20e weighted dataset

Note: Countries highlighted in red are new additions in the 9\textsuperscript{th} Vintage.
APPENDIX 2

Data Sources
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.
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.
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.
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.
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.
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.
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.
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.