CompNet The Competitiveness Research Network

Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU comparability

By

The Working group on the "Cross-country comparability" of the CompNet Dataset

chaired by Marc Melitz[1]

[1] The working group consisted of Carlo Altomonte (Bocconi), Eric Bartelsman (VU Amsterdam), Jan-Paul van de Kerke (ECB), Paloma Lopez-Garcia (ECB), Filippo di Mauro (NUS), Marc Melitz (Harvard), Michael Polder (Statistics Netherland) and Sebastien Roux (INSEE). This report was prepared by Filippo di Mauro, Jan-Paul van de Kerke and Michael Polder. Please cite this report as:

CompNet (2018), "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU comparability".

We would like to thank Alessandro Zona-Mattioli, Matteo Sartori, Marta Colombo, Daan Beijersbergen and Daniele Aglio of the ECB for their valuable research assistance. Special thanks to all data providers of CompNet (listed on the next page) for their assistance in providing the essential statistical information on their respective firm-level database and for their patience in responding to our queries.

The opinions expressed are those of the members of the Working Group only and do not necessarily reflect those of the institutions they belong to, nor of the data providers and their respective institutions.

Data providers and their institutions

Country code	Country	Contact person	Institution
BE	Belgium	-	National Bank of Belgium (BACH)
HR	Croatia	Katja Gattin Turkalj, Martin Pintarić	Hrvatska Narodna Banka (Croatian National Bank)
CZ	Czech Republic	Kamil Galuscak, Ivan Sutoris	Česká národní banka (Czech National Bank)
DK	Denmark	Andreas Kuchler	Danmarks Nationalbank (Central Bank of Denmark)
FI	Finland	Satu Nurmi, Juuso Vanhala	Tilastokeskus (Statistics Finland)
FR	France	Sebastien Roux, Raphael Lee	INSEE (Statistics France)
DE	Germany	-	DESTATIS (German Federal Statistics)
HU	Hungary	Judit Rariga, Mihály Szoboszlai	Magyar Nemzeti Bank (Central Bank Hungary)
IT	Italy	Filippo Oropallo	ISTAT (Statistics Italy)
LT	Lithuania	Aurelija Proškutė	Lietuvos Bankas (Central Bank Lithuania)
NL	Netherlands	Michael Polder	CBS (Statistics Netherlands)
PL	Poland	Jan Hagemejer	Narodowy Bank Polski (Central Bank Poland)
PT	Portugal	-	Banco de Portugal (BACH)
RO	Romania	Alexandru Leonte	Banca Națională a României (National Bank Romania)
SK	Slovakia	Tibor Lalinsky	Národná banka Slovenska (National Bank of Slovakia)
SL	Slovenia	Matjaz Koman	Univ. of Ljubljana
ES	Spain	Begoña Gutiérrez del Olmo	Banco de España (BACH)
SE	Sweden	Andreas Poldahl	Statistiska centralbyrån (Statistics Sweden)

Table of contents

Executive summary

In the last few years, CompNet has devoted substantial energy to improve the coverage, representativeness and cross-country comparability of its micro-aggregated dataset. The enlargement of the country coverage (now encompassing 19 EU countries) provides a unique opportunity to check the overall quality of the dataset again. This is particularly relevant because the recent inclusion of several statistical institutes within the network allows to concretely implement measures aimed at improving the quality of the dataset, and most notably its cross-country comparability — a critical requirement for research and policy analysis.

The report analyses the dataset in its current (6th Vintage) version along two dimensions. First, on the input side, a detailed account of the firm-level data sources utilised by the different teams at the country level as well as the methodologies used to collect and treat micro-information is given. Second, on the output side, the report examines methodologies and results obtained when the sector-level indicators are eventually produced.

The ultimate objective of the two parts is to assess whether sources and data collection methodologies are consistent across countries (on the input side and the output side) and whether indicator construction methodologies guarantee a good representativeness of the dataset. When this is not the case, the report indicates, where possible, the extent to which such issues affect overall comparability and provides recommendations on what country teams should do to modify their respective databases and what the network as a whole should aim at to improve representativeness.

The report finds that the dataset has strong fundamentals to rely on. First, the national data sources contain a significant share of sources based on definitions and statistical guidelines set by the EU, which already require substantial harmonisation efforts. Second, most sources are fully or partly under the responsibility of national statistical institutes, which is a guarantee of top methodological standards. Third, when sources do not draw from census information (i.e. are based on surveys), they can be linked to the census population for virtually all the countries; this allows the possibility of *ex-post* reweighing to ensure that the database out of that specific source is adequately representative of the population.

These fundamentals result in a 6th CompNet vintage covering on average about 40% of the corresponding population of firms (drawn from Eurostat), which is high in comparison with other datasets. Admittedly, the firm coverage of the CompNet dataset varies across countries, but this does not appear to represent a problem, since overall the country samples are representative. In particular,

the CompNet dataset captures the various segments of the firm-size population well and in line with Eurostat, albeit with some exceptions.[2]

On a number of issues, work is ongoing. For instance, more investigation is needed on (i) the ability to use national business registries for decentralising the weighting procedure (now 7 countries indicated that business registers are available for this transition), (ii) the impact different units of observation (e.g. enterprises versus legal entities) have in, among others, Germany, Poland and Spain and (iii) different procedures and solutions being used at the country level to follow business dynamics (e.g. entry, exit, mergers) on comparability and what the common denominator is.

Going forward, the report sets up an ambitious, manageable agenda to further improve the comparability of the dataset over time. Besides the above-mentioned issues requiring more analysis, some of these improvements can already be achieved — at least partly — along the following two dimensions:

(1) By simply revising the common code. In this context, plans are already in place to (i) implement correction routines for cross-country variation in variables such as value added, employment and labour taxation as well as (ii) adjust the common code to better match the cleaning and weighting procedures at the network level, with the ones already adopted at the country level and (iii) start pilot runs of the common code to deepen the common denominator.

(2) At the network level by providing precise instructions and definitions to all current and future data providers to have clarity on the content of the variables and to share best practices among data providers better.

The priority now should be to implement the recommendations in this report for the next vintage, thus solidifying the accuracy of the CompNet micro-aggregated dataset.

[2] In addition, the cross-country report on 6th vintage CompNet data provides a comparison of CompNet output with other published data sources as a general validation of the indicators. Overall, risks of incomparability under the dimensions of coverage, representativeness and validation appear to be limited.

1. Introduction

The use of micro-information is becoming a recurrent feature in all fields of economics. Data are increasingly available, computing power has tremendously increased and sophisticated techniques have been developed for the use of such information. It is by now also evident to researchers and policymakers that firm-level information can provide value added, i.e. by allowing to disentangle the heterogeneous responses of firms, usually hidden when only the macroeconomic figures are used. Everybody is now aware that looking at averages only, instead of distributions, can be misleading when trying to assess impacts of shocks or macro-policies, since macro-responses will vary depending on the underlying microstructure. A prominent example of this can be found in recent literature investigating the increases in market concentrations in the US and Europe. A subset of firms is gaining market shares over the rest. In traditional aggregate information, which average across all firms, these increases in market share would not be observed.

Actual use of firm-level information is, however, still hampered by a number of hurdles, especially in a cross-country setting. First, because ensuring confidentiality is differently interpreted and applied across data providers, actual data availability is uneven across countries and sectors/industries. Second, and partly related to the above, information gathered from firm-level data is not directly comparable across countries if no serious attempts are made to harmonise the sources and computation methodologies.

The CompNet project has the explicit objective to make this international comparison possible. This report aims at analysing the extent to which this objective has been achieved in terms of input variable harmonisation and adequate coverage and representativeness of the dataset. Its ultimate goal is to indicate how the dataset can be further improved leveraging on the skills and data sources available within the network. The underlying assumption is that increasing the reliability of a firm-level-based dataset — and most notably its cross-country comparability — is a process. There is therefore a need to periodically assess — with a very transparent analysis of the status quo — the overall quality of the dataset to promote its correct use and further enhance its quality.

This report is structured as follows. Section 2 introduces the use of firm-level data and their adequacy for cross-country analysis, providing also a discussion of comparability issues. Section 3 analyses in detail such issues in regards specifically to the 6th Vintage of the CompNet dataset, focussing on the input side. Section 4 extends the comparability analysis to the output side. Section 5 discusses improvements and recommendations to the database and offers guidelines to data providers within the CompNet network. Section 6 concludes the report.

2. Using firm-level data for cross-country analysis and research

Firm-level analysis has a number of critical advantages but is hampered by a number of issues. Subsection 2.1 discusses data availability. Subsection 2.2 briefly introduces the CompNet micro-aggregated database. Subsection 2.3 will then sketch the issue of "cross-country comparability". The approach we take to investigate cross-country comparability is highlighted in Subsection 2.4. Subsection 2.5 presents the so-called Comparability Tool.

2.1 Firm-level data availability

Firm-level data exist for almost all European countries, in some cases even containing very fine details (most notably, and depending on the topic, in Belgium, Denmark and France). However, the possibility of acquiring access to and actually using this micro-information varies. Some statistical institutes facilitate access to this information for research purposes to eligible parties, while in other instances access to this firm-level information can be a complicated procedure. As a result, the firm-level-based literature in the EU is concentrated on a few countries only and often does not go beyond single-country studies. There is an increasing recognition that for benchmarking analysis and for analysing the impact of policies that vary across countries and industries, having the data available for multiple countries is essential. Progress has been made on multi-country data and several initiatives have taken off. On the commercial side, cross-country firm-level datasets such as ORBIS and CompuStat are available. International organisations have also initiated projects by using the micro-aggregated approach (Bartelsman, 2004). Rather well-established datasets in this category include EFIGE (Barba Navaretti et al, 2011), ESSNet MMD (Bartelsman et al. 2018), CompNet (Lopez-Garcia and di Mauro, 2015) and Multiprod (Berlingieri et al. 2017). The access to these datasets varies as well. The MMD can be accessed through Eurostat. CompNet is available for member institutions and upon request. Multiprod is not accessible for researchers external to the OECD.

2.2 CompNet as micro-aggregated database

The Competitiveness Research Network (CompNet) was created in March 2012 as an initiative of a group of research departments of European Central Banks. Since then, CompNet has developed as a standalone network with a broader membership, with a distinctive focus on competitiveness-related research and policy work taking a firm-level perspective. The network aims at providing a robust theoretical and empirical link between drivers of competitiveness and macroeconomic performance for the purpose of research and policy analysis. This is done primarily by systematically updating the CompNet micro-based dataset. The three main advantages of the approach followed by the network in constructing the database are that (i) the dataset uses existing information, with no need to undertake

new and costly new data collection efforts, (ii) confidentiality of the micro-information is fully protected and (iii) member institutions participate actively in improving and using the database. The eventual product consists of a broad set of micro-aggregated measures based on firm-level data, nationally available in each of the institutes participating in the network as data provider, i.e. national statistical agencies, national central banks and research institutes. The indicators produced (about 70) encompass means, totals and variances at different levels of aggregation, namely the sectoral level, macro-sector level, size-class level and at the country level.[3] The database provides additional indicators capturing the characteristics of the underlying distribution of firm characteristics such as the percentiles, skewness and the dispersion, as well as parameters of joint distributions. Recent improvements to the CompNet database entailed a higher coverage across European countries, new indicators related to productivity estimation and zombie firms, increases in cross-country indicators comparability and improvements in the quality and efficiency of the computing routines.

2.3 Outline of comparability issues

There are two main ways to ensure that firm-level data are comparable across countries. The most effective way is to generate datasets/surveys that are *ex-ante* built with similar characteristics (see for instance the EFIGE project in Altomonte and Aquilante, 2012); this, however, is very costly and difficult to sustain over time for a large number of firms. As a result, it is not possible to gather historic as well as comprehensive data in this manner, or to provide indicators on a large variety of topics.

Alternatively, one can try to ensure better comparability *ex-post* by starting from different datasets — constructed for different purposes and with possibly varying sampling criteria. This can be achieved through the use of harmonised procedures to generate the relevant indicators and other means, which, as we will see in detail in this report, is the way chosen by the CompNet project.

In general, the issue of cross-country comparability of firm-level datasets is paradoxically researched very little (see, for instance, Airaksinen et al. 2013, Hagsten et al. 2012). Among the few existing papers, Kalemli-Ozcan et al. (2015) are, however, more concerned about how to ensure representativeness of the firm-level sample at the national level using different vintages of ORBIS. Commercial datasets — such as ORBIS-AMADEUS — completely lack focus on comparability and leave the issue entirely to the users. The result is that, typically, the user will possibly mention some caveats at first, but will overlook the issue entirely thereafter when arriving at the discussion and interpretation of the results. This is not satisfactory because it undermines the credibility of the firm-level analysis. This is the reason why

[3] The CompNet database uses the NACE rev.2 classification as sectoral identifier for both the sector as well as the macro sector dimensions.

CompNet has invested substantial effort into this issue across all 6 vintages of its dataset — on the one hand, to make sure that the dataset is increasingly reliable and can be used also for policy purposes, and on the other hand, to make all caveats and provisions very clear at the outset, to not undermine the kind of analysis and/or specific indicators that are being used.

When assessing cross-country comparability, there are at least three caveats that should be stressed at the outset.

1) First, it is a fact that data sourcing, legal settings and tax policies vary across jurisdictions.[4] As a consequence, any international data exercise — such as CompNet — will face the challenge of full data comparability. The clue, however, is to make sure that at the stage of generating the micro-aggregated data, known issues are taken into account as much as possible in order to preclude avoidable misalignments. Moreover, when areas of caution have been identified, issues of comparability may be dealt with or mitigated in the analysis, for example through weighting or econometric modelling. The usage of micro-aggregation is already an example of a mitigating factor to comparability issues. In fact, these country differences do not alter the shape or the distribution of the measured variables while these distributions are used for analysis. So the topic of comparability, while very important, impacts micro-aggregated data less. Section 5.3 describes such econometric procedures in more detail.
2) Second, regarding the comparison of CompNet indicators (i.e. the "output" as we will define it later) with official statistics, one should be cautious about (i) choosing the right benchmark as well as (ii) interpreting the results of the comparison. For instance, while it is obvious to start by comparing the outcomes to familiar aggregates, such as the National Accounts, one should also acknowledge that these are an "integration framework of data". They originate from a variety of sources, not only on businesses, but also on households, workers, consumers and government administrative sources and others. The result of integrating these sources is that the original micro-data sources will not add up to the macro-totals.
3) Related to the previous point and given the complex nature of the compilation process of the National Accounts, it should therefore not be a goal to fully mimic or reproduce these macroeconomic totals.[5] Rather, one should view micro-aggregation initiatives such as CompNet as a way to complement the official statistics with policy-relevant information that is otherwise

[4] For a large share of input sources within CompNet, harmonisation is achieved under the Eurostat regulations. CompNet also draws upon sources that do not fall under these regulations and are thus to a lesser extent harmonised.

[5] However, the trends and observations done from both macro- and micro-information sets should correspond or coalesce to some degree.

unavailable, in particular information on business dynamics and firm heterogeneity, including univariate or multivariate distributions and cross-correlations.

2.4 Our approach to assess comparability

Despite the efforts made, a challenge that is inherent in the compilation of cross-country micro-aggregated dataset such as CompNet is that individual country data are derived from a variety of sources, each with an individual strategy for reporting unit, inclusion of thresholds, and timing and definition of variables, among other things. These differences obviously affect cross-country comparability. With the enlargement of the CompNet membership, both in countries providing data and in the user community, testing data source heterogeneity has become more urgent. Indeed, the involvement of more experts from NSIs in the network can aid in the ambition for higher standards of comparability across countries.

In this report, we provide a description and assessment of the comparability issues still faced by the latest vintage of the CompNet database and the way the network is planning to overcome such remaining issues. Specifically, we will analyse the comparability issues, distinguishing the input side (i.e. the firm-level data sources utilised at the country level) from the output side (i.e. the indicators which are constructed using such individual sources). More specifically, we will look at the following dimensions:

1) On the input (firm-level) data sources (Section 3):
    1. Harmonisation of input data
        1. Type of sources and data collection methods
        2. Time period covered
        3. Representativeness of sources
        4. Statistical treatment of the data
        5. Documenting the unit of observation
    2. Linking of input sources
        1. Linking different sources within countries
        2. Longitudinal linkages
        3. Other known quality issues
    3. Assessment of comparability of input variables
        1. Currency and units
        2. Employment
        3. Value added
        4. Definition and valuation of production variables

5. Timing of variables
6. Variable-specific information

2) On the output side (Section 4):
    1. Industry coverage of the indicators
    2. Representativeness and weighting
    3. The within-cell firm size bias

The following sections examine the newly constructed CompNet 6th vintage along the lines outlined above.

2.5 The "Comparability Tool"

In Section 3, to assess the comparability on the input side, the information is primarily obtained from an extensive survey carried out among the data providers. To structure the analysis of the information from that survey, a "Comparability Tool" was constructed, which traces — for each input variable and/or output indicator — all the metadata information related to sources and potential cross-country comparability issues. With the tool, researchers can assess the comparability across countries for the specific variables and indicators used in their analysis.

To construct the tool, we proceeded as follows, see Figure 1 for reference:

1) We have coded all the information and summarised them in tables (matrix B1–B3), which can be broken down into three categories:[6]

    B1. (Within country) Input variable-specific data features (m_{var_v,cou_c}, e.g. employment in country Y concerns headcount).

    B2. Country-specific data features (m_{cou_c}, e.g. country Y provides data for period T)

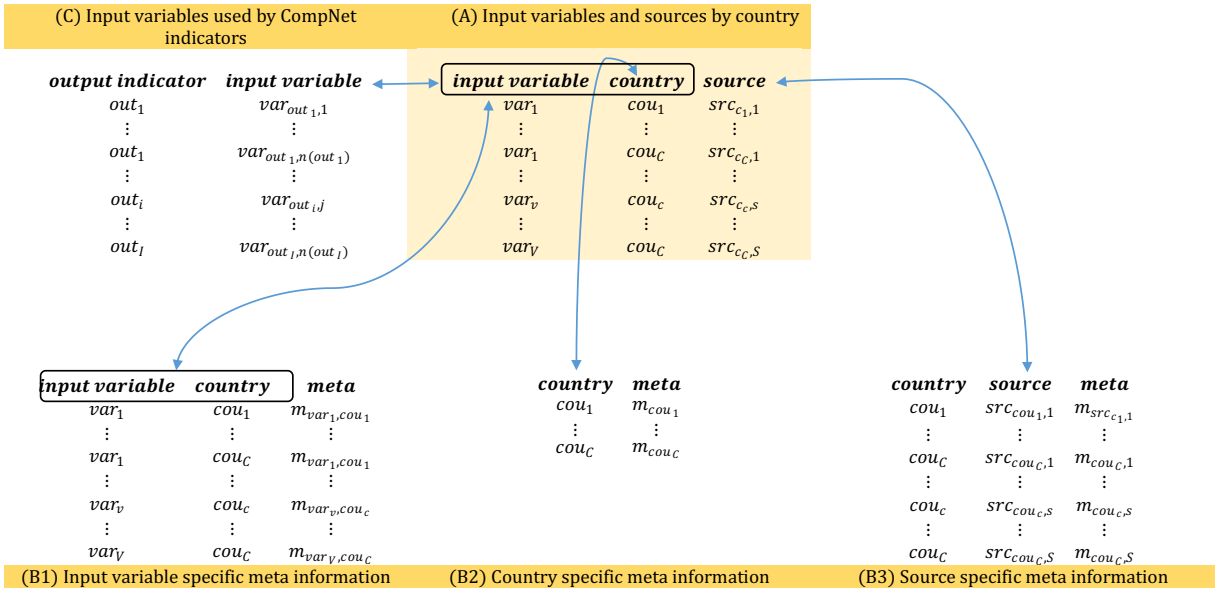
    B3. (Within country) Source-specific data features ($m_{cou_c,s}$, e.g. source X in country Y is a survey)

2) We have linked Tables B1–B3 to a correspondence table A to get metadata information by input variable and country.
3) Finally, matrix C provides the correspondence between CompNet indicators to input variables (e.g. labour productivity is composed of value added and employment), which allows linking the meta-information by input variable to the output indicators.

[6] More detailed information is posted on our website; www.comp-net.org.

As an illustration, suppose we are using the output indicator "Value-added based labour productivity", which matrix C indicates is calculated from two input variables: value added and employment. Matrix A in Figure 1 below gives the sources used in each country for both input variables. In country Y, they may be called SBS (value added) and BR (employment). The matrices B1–B3 then gives us all the available metadata by country for these two variables. Matrix B1 may tell us that "Value added" in country Y is at market prices, and "Employment" refers to headcount. Matrix B2 could tell us that for all data in country Y (thus, including employment and value added) the reporting unit is the enterprise (see Subsection 3.1.4 for more information on this issue). Finally, matrix B3 may tell us that SBS is a sample-based source, and that BR is a census, so that our labour productivity measure can be calculated only for the SBS sample in country Y.

Figure 1. Metadata mapping



Note: output indicators: out_i, $i = (1, \dots, I)$; input variables: var_v, $v = (1, \dots, V)$; countries: cou_c, $c = (1, \dots, C)$; source: sou_s, $s = (1, \dots, S)$.

3. CompNet 6th vintage: the input side

This section assesses the comparability of national databases that are used by participant countries as the input into CompNet calculation routines, by documenting cross-country differences with respect to the relevant issues outlined in Subsection 2.4. Furthermore, it (i) identifies whether any differences affect specific parts of the CompNet dataset, (ii) assesses the magnitude of the issue where possible, and (iii) provides recommendations for usage of its current version as well as for improvements in future rounds. The current section is divided in three subsections, dealing with, respectively, the type of input sources and data collection methods; the linking and comparability of input sources over time; and the assessment of comparability of input variables. Section 4 will turn to the output side, which is the resulting cross-country micro-aggregated dataset that is available to the research community.

3.1 Type of input sources and data collection methods

Most data providers in CompNet use multiple sources to construct the firm-level database used to run the common codes. Generally, providers combine different kinds of information, e.g. business registry and customs data, but in some instances also sources related to the same type of information are combined to improve the coverage of the national database. For instance, as regards financial information, the French team at INSEE combines a source covering larger firms with turnover exceeding €788,000, with another covering firms with turnover below €788,000.

Table 1 gives an overview by country of all data sources used in the 6th vintage of CompNet and the national institutes responsible. Various types of sources are used including administrative, financial and balance sheet information as well as customs data. Among the data sources there are business registers (BR) and structural business statistics (SBS) surveys in Croatia, Finland, Hungary, Czech Republic, Italy, Lithuania, the Netherlands and Sweden. These sources are based on EU regulations[7] and are harmonised across countries. The majority (27 out of 36) of the databases are maintained (at least partly) by national statistical institutes (NSI), which confirms the suitability of these sources for statistical

[7] Business register regulation was adopted by the European Parliament on 25 October 2007 and by the Council of Ministers of the European Union on 21 January 2008 and came into force on 25 March 2008. It replaces the previous Regulation (Council Regulation (EEC) No 2186/93 of 22 July 1993 on Community coordination in drawing up business registers for statistical purposes (OJ L 196, 5.8.93)) and is part of a series of regulations intended to harmonise the European business statistics infrastructure.

Structural Business Statistics regulation: EC Regulation 58/1997 (complemented by EC 1618/1999). Commission Regulation (EC) No 250/2009 of 11 March 2009 implementing Regulation (EC) No 295/2008 of the European Parliament and of the Council as regards the definition of characteristics, the technical format for the transmission of data, the double reporting requirements for NACE Rev.1.1 and NACE Rev.2 and derogations to be granted for structural business statistics – Articles 2 and 3, Annexes I and II.

purposes. Although 9 out of 36 sources are maintained by institutions other than the NSI, 13 national statistical offices out of 15 countries included in the table are represented by at least one data source. Subsections 3.1.1–3.1.4 provide further details on the utilised sources.

Table 1: Brief description of input data sources across countries

Country	Data source name	Acronym	Institution	Time span	Sample/ census	Link to business registry
Croatia	Yearly financial statements of firms	FINA	Financial Agency Croatia	2002–2016	Census	BR*
Czech Republic	Annual report of economic units in selected production industries P5-01	P501	Statistics Czech Republic	2003–2015	Sample*	Yes
Czech Republic	Extrastat/Intrastat foreign trade transaction data	TRADE	Statistics Czech Republic	2005–2015	Census	Yes
Czech Republic	Business Register	RES	Statistics Czech Republic	2003–2015	Census	BR
Denmark	Accounts statistics — non-agricultural industries	Acc. Stat.	Statistics Denmark	2004–2015	Sample	Yes*
Denmark	General enterprise statistics	Gen. Stat.	Statistics Denmark	2004–2015	Census	Yes
Finland	Structural business and financial statement statistics data	SBS	Statistics Finland	1999–2015	Census*	Yes
Finland	International trade statistics data	ITS	Finnish Customs	1999-2015	Census	Yes
France	Regime of Normal Real Profits	BRN	Statistics France	2008–2014	Census	Yes
France	Simplified Regime for the Self-Employed	RSI	Statistics France	2008–2014	Census	Yes
Germany	Administrative firm-level data	AFiD	Statistics Germany	2001–2014	Census	Yes
Hungary	Tax registry database of National Tax and Customs Administration	NAV	National Tax and Customs Administration	2000–2015	Census	Yes
Hungary	Business Registry	VR	Statistics Hungary and Central Bank of Hungary	2000–2015	Census	BR
Hungary	Export–Import data of Hungarian Enterprises	Külker	Statistics Hungary	1975-2015	Census	Yes
Italy	Statistical Business Register	ASIA	Statistics Italy	2001–2015	Census	BR
Italy	Balance Sheets of non-financial companies	BIL	Statistics Italy	2001–2015	Census	Yes
Italy	Large enterprise survey	SCI	Statistics Italy	2001–2015	Census	Yes
Italy	Foreign Trade Statistics based on custom data	COE	Statistics Italy	2001–2015	Census	Yes
Lithuania	Statistical Survey on the Business Structure (Annual questionnaire F-01)	F01	Statistics Lithuania	1995–2015	Census	Yes
Lithuania	Business Register	BR	Centre of Registers	1995–2015	Census	BR
Lithuania	Customs, Customs declarations	CU	Customs of the Republic of Lithuania	1995–2015	Census	Yes
Netherlands	Statistics finances of non-financial enterprises	SFO	Statistics Netherlands	2000–2014	Census*	Yes
Netherlands	Business register	ABR	Statistics Netherlands	2000–2014	Census	BR
Poland	Reports on revenues, costs, profit and outlays on fixed assets	F01	Statistics Poland	2005–2015	Census	*
Poland	Stat. financial report	F02	Statistics Poland	2005–2015	Sample	
Portugal	Central balance sheet database, annual survey	CBSD	Central Bank of Portugal	2000–2005*	Sample	Yes
Portugal	Simplified corporate information	IES	Statistics Portugal and Central Bank of Portugal	2006–2016*	Sample	Yes
Romania	Balance sheet information on non-financial enterprises	Bal. Sheet	Ministry of Public finances	2005–2016	Census	
Romania	Exports and imports of goods, firm-level data	TRADE	Statistics Romania	2005–2016	Census	
Slovakia	Annual report on production industries	Reports	Statistics Slovakia	2000–2016	Sample	Yes
Slovakia	Statistical register of organisations	Register	Statistics Slovakia	2000–2016	Census	BR
Slovakia	Foreign trade statistics	Customs	Statistics Slovakia	2004–2016	Census	Yes

Table continued

Slovenia	Slovenia Public and Legal Records and Related Services	AJPES	Agency for Public Legal Records and Related Services	2005–2016	Census	Yes
Spain	CBSO voluntary survey	CBA	Central Bank of Spain	2009–2016	Sample	Yes
Spain	Spanish mercantile register	CBB	Mercantile registry	2009–2016	Census	Yes
Sweden	Structural business statistics	SBS	Statistics Sweden	2003–2015	Census*	Yes
Sweden	International trade in goods	ITG	Statistics Sweden	2003–2015	Census	Yes
Sweden	Business register	BR	Statistics Sweden	2003–2015	Census	BR
Belgium						

Source: Survey carried out among data providers.
* Notes: Croatia: FINA source acts essentially as a business register.
Czech Republic; P501 covers the Census for firms >50 employees, a sample for firms <50 employees.
Denmark; Acc. Stat. information is derived from source documentation.
Finland; preliminary, SBS and ITS information derived indirectly from source documentation. SBS covers for small firms a sample but within SBS framework refers to census information.
Hungary: data derived from administrative sources.
Netherlands: census of corporate tax paying firms; small firms from tax register large firms from survey.
Poland: Information on the ability to link to business registry is pending.
Portugal: CBSD sample covers 5% of enterprises and 40% of employees in population; therefore, only years post 2006 are included where coverage of IES sample is 98%.
Sweden: SBS framework of surveys and administrative data covering the census.

3.1.1 Time period covered in the data sources

Column 5 of Table 1 reports the time span available in the data sources. All sources have data up to 2014, and most up to 2015 or 2016. Most data sources start somewhere in the early 2000s, with some sources going further back, such as in Lithuania and trade information in Hungary. Thus, the large majority of countries cover the period from early 2000s until 2014/2015, with the exception of Spain (2009–2015) and Portugal (2006–2015).[8] For specific analysis where a longer time dimension is required, the researcher is therefore faced with a trade-off between either including more countries or more years, or to take for granted that some countries do not cover the entire period. This decision is research specific and does not affect directly the comparability of the CompNet database, although the user should take care to select comparable time periods when the research question requires this.

3.1.2 Representativeness of sources

The extent to which the sources are representative, perhaps to varying degrees, for the universe of firms is important not only for assessing comparability but also for the overall strength of our dataset. This subsection looks at the representativeness of the input sources in terms of industry and size-class coverage, as well as whether the source is a census or a sample.

[8] In Portugal, this is due to a significant improvement in data quality, attributable to switching from the CBSD to the IES. In Spain, the data used covers only the period 2009–2015.

- **Reported industry coverage**

Table 2 gives the industry coverage for each country by firm-level source. Except for Denmark, Spain and Poland, the data covers the total economy, whereas all countries cover the non-financial business sector (Non-Financial Corporations (NFC), i.e. NACE Rev 2 Chapters B–J and L–N and division 95).[9] The industry coverage by the various sources does therefore not cause any concern about comparability, although in macroeconomic comparisons researchers should be aware of whether firms outside the business economy are included or not.

Table 2: Reported industry coverage

Country	Source	Industry coverage	Excluded industries
BE			
CZ	P501	Total economy	Division 01
CZ	RES	Total economy	
CZ	TRADE	Total economy	
DE	AFiD*	Manufacturing	All chapters other than C
DK	Acc. Stat.	NFC	
DK	Ent. Stat.	Total economy	
ES	CBA	NFC	
ES	CBB	NFC	
FI	ITS	Total economy	
FI	SBS	Total economy	Chapter A excludes firms with no employees; Chapter K, groups 851-854, 856; 9101, Division 94
FR	BRN	Total economy	Chapter A; K
FR	RSI	Total economy	Chapter A; K
HR	FINA	Total economy	Chapter K
HU	BR	Total economy	
HU	NAV	Total economy	
HU	Külker	Total economy	
IT	ASIA	Total economy	Chapter A; division 64-65; 84; 94
IT	BIL	Total economy	Chapter A; division 64-65; 84; 94
IT	COE	Total economy	Chapter A; division 64-65; 84; 94
IT	SCI	Total economy	Chapter A; division 64-65; 84; 94

[9] Please note that CompNet *input* includes all industries in the NFC. The CompNet output, however, slightly deviates from this classification, excluding NACE rev.2 chapters B, D, E and division 95 and excluding self-employed. Please refer to Table 24 in the Annex for a detailed overview of the NFC sectors included in CompNet.

Table continued

LT	BR	Total economy	
LT	CU	Total economy	
LT	F01	Total economy	Chapter K; division 01; 84; 94
NL	BR	Total economy	
NL	SFO	Total economy	Chapter K
PL	F01	NFC	
PL	F02	NFC	
PT	CBSD	NFC	
PT	IES	Total economy	Chapter K
RO	bal. sheet	Total economy	
RO	Ex&Imp	Total economy	
SE	SBS	Total economy	Chapter K, O
SE	ITG	Total economy	
SE	BR	Total economy	
SI	AJPES	Total economy	Chapter K
SK	customs	Total economy	
SK	register	Total economy	
SK	reports	Total economy	Chapter K

Source: Survey carried out among data providers.
Total economy: NACE Rev 2 Chapters A–S NFC = non-financial corporations: NACE Rev 2 Chapters B–J, L–N; Division 95. Chapter K is the financial sector (Division 64–66).
* Germany, while the AFiD captures the total economy only the manufacturing sector was provided for the CompNet sample.

- Transformation to NACE Rev.2

An aspect strongly linked to the industry coverage is the classification system used to assign these industries. Specifically, how industries were assigned to the NACE Rev.2 sectors used in CompNet in data before 2009, when the NACE Rev 1.2 was operational. From one survey to another, we can deduce that all samples are NACE rev.2 compliant but the methodology to do so differs across countries. Data providers from Denmark, Croatia, Poland, Hungary and Lithuania indicated that the transformation was done by the statistical office. Other data providers give more information on the applied reclassification methodology. The Netherlands and Romania apply a three-tier approach; Concordance tables are used for NACE rev.1 and NACE rev.2, information of 2009 (2008 in Romania) is used to assign earlier years and an additional check is applied. The latter consists of a derivation algorithm of the Dutch NSI used for financial data and an employment check when a firm fits into 2 sectors. Other data providers apply similar methods as the Dutch and Romanian data providers such as the correspondence tables and the classification keys in the cases of Finland, Italy, Spain and Slovakia. In the Czech Republic, data from 2005 to 2009 were already NACE rev.2 compliant, data before 2005 were matched separately at the 2-

digit level. While there is no apparent risk for comparability, it is beneficial to share the common methodologies with new data providers so that input data remains NACE rev.2 compliant.

- **Reported size-class coverage**

Table 3 gives an overview of any size classes excluded in the various sources. Overall, the exclusion of smaller firms does not appear to be a pressing issue in CompNet, it is limited to Poland, Germany and Slovakia — for firms of less than 10 and 20 employees, respectively. Although sources in France and Italy target specific size classes, they are complemented by additional sources to cover all size classes and thus do not face binding exclusions.

To help researchers overcome differences due to the exclusion of smaller firms in Poland and Slovakia, CompNet offers the opportunity to make use of an additional dataset (so called 20E) containing the indicators based on firms with 20 employees or more.

Table 3: Exclusion of size classes by data source

Country	Source	Excluded size classes
BE		
CZ	P501	Full coverage for firms with >50 employees, survey for smaller firms
CZ	RES	
CZ	TRADE	
DE	AFiD	< 20 employees
DK	Acc. Stat.	
DK	Ent. Stat.	
ES	CBA	
ES	CBB	
FI	ITS	
FI	SBS	
FR	BRN	All size classes (turnover > €788k)
FR	RSI	All size classes (turnover < €788k)
HR	FINA	
HU	BR	
HU	NAV	
HU	TRADE	
IT	ASIA	
IT	BIL	
IT	COE	
IT	SCI	< 100 annual average employees
LT	BR	
LT	CU	
LT	F01	
NL	BR	

Table continued

NL	SFO	
PL	F01	< 10 employees
PL	F02	< 10 employees
PT	CBSD	
PT	IES	
RO	Bal. sheet	
RO	Ex&imp	
SE	SBS	
SE	ITG	
SE	BR	
SI	AJPES	
SK	Customs	
SK	Register	
SK	Reports	< 20 employees

Source: Survey carried out among data providers

- **Census and non-census sources and sample characteristics**

Data sources can cover the entire firm population, i.e. they form a census, or they can cover only a part of the population, in which case they are a sample. Census data is representative of the population by definition, whereas sample-based data usually require weighting to obtain representative aggregates.

Comparability in this respect does not appear to be an issue. First, drawing from column 6 in Table 1, out of the 38 sources which are being used, the large majority (31) refer to census information.

Second, drawing from census information rules out possible selection bias which can affect comparability adversely.

Third, when sources concern samples, they are always able to link these sources to information on the population, such as a national business registry (see column 7 in Table 1), thus enabling the calculation of weights to adequately reflect the population distribution.

Turning now to the non-census information, Table 1 shows that about one-fourth of the data input sources (i.e. 7 out of the 38) used in CompNet derives from samples (i.e. they do not include the entire population). For these specific data sources, we do not have adequate information on sampling schemes or possible stratification. However, the corresponding comparability issues seem to be rather contained. First, and as mentioned, apart from two countries (Poland and Hungary), all sample sources can be linked to a source covering the census information. Second, for four countries (Hungary, Italy, Poland

and Slovakia), there is a further indication of the appropriateness of the resulting database for making statistics since national statistical offices are involved in their compilation.

In addition, the analysis of the representativeness on the CompNet output database in Section 4 provides further confirmation that the overall representativeness is good. The assessment given there compares the share of firms and employees by sector and size class in the 6th vintage CompNet database with the shares derived from the officially published figures available from Eurostat, at the macro-sector and size-class level, as well as through a more granular test on the within-cell bias.

In principle, when appropriate weighting schemes are available, there is no threat to comparability when some countries have census and others sample data. The CompNet code uses a reweighting procedure that is based on the comparison to Eurostat, either in terms of the number of firms or employment. While this generic approach in principle improves the comparability across country, there are two caveats to be made. First, using the weights based on the number of firms assumes that the definition of the firm in the CompNet input sources is the same as that used by Eurostat. This issue is investigated in Subsection 3.1.4, and we will see that this is not the case for all countries. In addition, as census sources cover the entire population, no reweighting procedures are needed to produce statistics representative for the universe of firms. Therefore, future vintages could consider differentiating between census- and sample-based sources and apply weighting only to the latter. Please refer to Subsection 4.3.1 for more information on the reweighting procedure.

3.1.3 **Statistical treatment of the data**

This section documents the possible application of correction methods by data providers (or the institutes responsible for the source) to the original input datasets, in particular the treatment of missing values and the treatment of aberrant observations.

Treatment of missing values

According to standard statistical practice, missing values and non-response are replaced by estimates, based on, for example, known firm characteristics or historical information, a process referred to as “imputation”. Alternatively, missing values are sometimes kept zero. Such practices vary across sources and across countries, and may therefore distort comparability.

Imputations are done for the purpose of deriving aggregate statistics, but can have an impact on firm-level analysis. White, Reiter and Petrin (forthcoming), for example, show that common practices for the imputation of missing values lower considerably the variance of the variables in question. In the specific case of CompNet, this would have an impact on the robustness of the competitiveness analysis, which makes use of the variance of productivity and misallocation indicators. For analytical purposes, it is

therefore desirable to be able to revert to the original — not imputed — variables series, or when the original values are not available, that those imputed observations are flagged, to have the possibility to exclude them from specific calculations.

Table 4: Imputations of missing values and flagged values

Country	Source	Imputations	Flags
BE			
CZ	P501	No	
	TRADE	No	
	RES	No	
DE	AFiD	No	
DK	Acc. Stat.	Yes	Yes
	Gen. Ent.	No	
ES	CBA	No	
	CBB	No	
FI	SBS	Yes	No
	ITS	No	
FR	BRN	No	
	RSI	No	
HR	FINA	No	
HU	NAV	No	
	VR	No	
	Trade	No	
IT	ASIA	No	
	BIL	No	
	SCI	No	
	COE	No	
LT	F01	Yes	No
	BR	Yes	No
	CU	Yes	No
NL	SFO	Yes	No
	ABR	No	
PL	F01	No	
	F02	No	
PT	CBSD	No	

Table Continued

	IES	No	
RO	Bal. Sheet info	No	
	Trade	No	
SE	SBS	No	
	ITG	No	
	BR	No	
SL	AJPES	No	
SK	Reports	No	
	Register	No	
	Customs	No	

Source: Survey carried out among data providers
Notes: No information available for Czech Republic, Italy, Poland.

Table 4 shows that variance in the treatment of imputation methods is not a big issue in the CompNet dataset. Only the data from Finland, Denmark and the Netherlands (for these countries only one source is altered) and Lithuania (all three sources) are subject to imputation. In the Danish data, these imputations can be flagged.

Recommendations:

While for many countries imputation of missing values is not an issue, for future vintages, it is worthwhile to investigate the possibility to take into account the flagged observations in Danish data when running the analysis. Furthermore, it is necessary to investigate whether such observations should be treated differently in the common code, or whether they should perhaps be excluded beforehand. In the cases of Lithuania and Finland, it is not possible to flag the imputations; here, it is recommended to investigate the share of observations that has been imputed, which would allow to quantify the impact of their inclusion.

Treatment of aberrant observations

For statistical purposes, data may be cleaned of any aberrant observations (i.e. outliers). As in the case of missing values, practices for doing so vary. Observations are either dropped, or replaced by more realistic values, which can be estimated in different ways. Again, analogous to the case of imputation, it is desirable to either revert to the original values or identify the observations that were replaced to be able to exclude them from the analysis.

Table 5: Other correction routines applied to raw variables and whether they are flagged

Country	Source	Corrections	flags
BE			
CZ	P501	No	
	TRADE	No	
	RES	No	
DE	AFiD	No	
DK	Acc. Stat.	No	
	Gen. Ent.	No	
ES	CBA	Yes*	No
	CBB	Yes	No
FI	SBS	Yes	No
	ITS	No	
FR	BRN	No	
	RSI	No	
HU	NAV	Yes*	No
	VR	No	
	Trade	No	
HR	FINA	No	
IT			
LT	F01	Yes	No
	BR	Yes	No
	CU	Yes	No
NL	SFO	Yes	No
	ABR	No	
PL			
PT	CBSD	Yes	No
	IES	Yes	No
RO	Bal. Sheet info	No	
	Trade	No	
SE	SBS	No	
	ITG	No	
	BR	No	
SK	Reports	No	
	Register	No	
	customs	No	
SL	AJPES	No	

Notes: No information available for PL, IT.
Spain: Comprehensive set of outlier and consistency routines are applied to assess quality and consistency.
Hungary: Misreporting, currency checks are applied, unique stories are corrected by hand.
Source: Survey carried out among data providers.

As Table 5 shows, in the sources of five countries (Hungary, Lithuania, Netherlands, Portugal and Spain) data correction routines were applied (for Finland and Hungary, this is the case for one source only). None of the corresponding sources flag the values that have been altered.

This issue has an obvious negative impact on comparability, but we do not have information to quantify the magnitude of this effect. Against this background, we make four recommendations.

Recommendations:

- First, since the CompNet code employs its own outlier detection methods,[10] data providers are encouraged to start from unfiltered data whenever possible.
- When countries are unable to start from the unfiltered data, the CompNet code can be adjusted to differentiate the outlier correction and treat unfiltered data differently from data that has already been cleaned. This potentially means that countries' samples face different outlier treatment procedures. However, we argue that this will be an improvement over the situation where samples are cleaned by both the data provider and the common code.
- The current outlier routine in CompNet is symmetrical between the left and the right tail of the distribution. Given the fact that the distribution is asymmetrical, one could think of taking this into account in designing the outlier routine to prevent loss of useful information, particularly in the largest cells that are of interest.
- For analysis, the researcher is advised that in the current vintage patterns will potentially look smoother and less volatile for variables from sources in which outliers were corrected, as compared to the ones coming from sources in which they were not.

3.1.4 Documenting the unit of observation used

The words "firm", "business", "enterprise" and "company" are used interchangeably in everyday language. But, these words have very different meanings in actual statistical usage. In particular, Eurostat defines an enterprise as *"an organisational unit producing goods or services which has a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situated at more than one location. An enterprise may consist out of one or more legal units."* Business Registers typically consist of: *enterprises*, carrying out economic activities contributing to the gross domestic product (GDP); their *local units*; the *legal units* of which those enterprises consist; and *enterprise groups* (association of enterprises bound together by legal and/or financial links). In this report, the Eurostat definition of the enterprise level is the point of reference. Note,

[10] Please refer to the Annex for a detailed description of these outlier cleaning procedures in the CompNet code.

however, that in some countries different concepts of the firm are used simultaneously. In France, for instance, the NSI indicated that the legal unit and enterprise level coexist across different data sources and provide, depending on the definitions used, a different picture of the economy (Béguin and Hecquet, 2015). Another example is the Czech Republic, where the relevant unit of observation in the Czech Statistical Office is the *enterprise* but the business register instead tracks the *legal unit*. This, however, does not pose an internal comparability issue since the two definitions are assumed and actually do coincide.[11]

While the reporting unit is mostly consistent across the sources used within each country, not all countries use the enterprise (as defined above) as the reporting unit. Table 6 shows that in 6 out of 18 countries, the reporting unit refers to higher levels such as the enterprise level (Croatia, Finland, Hungary, Italy, Lithuania, the Netherlands, Romania, Slovakia and Sweden). For 12 data providers (Czech Republic, Denmark, Spain, France, Poland and Portugal), the reporting unit is the *legal unit*.

Table 6: Reporting the unit of observation across countries

Country	Reporting unit	Consolidation
BE		
DE	Legal unit	
CZ	Legal unit	Unconsolidated
DK	Legal unit	
ES	Legal unit	
FI	Enterprise	Unconsolidated
FR	Legal unit*	
HR	Enterprise	Unconsolidated
HU	Legal unit*	Unconsolidated
IT	Legal unit	
LT	Enterprise	Unconsolidated
NL	Enterprise group	Consolidated
PL	Legal unit	
PT	Legal unit	

[11] The Czech Republic Structural Business Statistics Methodology document, see http://ec.europa.eu/eurostat/ramon/nat_methods/SBS/SBS_Meth_CZ.pdf (accessed on 11 July 2018).

Table continued

RO	Enterprise	Unconsolidated
SE	Enterprise	Consolidated
SI	Legal unit	Unconsolidated
SK	Legal unit	Consolidated

Notes: France: despite coexisting levels of aggregation, information for CompNet stems from data sources which use the legal unit.
Hungary: information is derived from fiscal sources; fiscal reporting unit is the legal unit.

The information on the extent to which this definitional question affects comparability is scattered. According to Eurostat, *"only a very small share of enterprises consists of more than one legal unit",*[12] suggesting that the impact of this issue should be limited. Statistics Sweden and the Central Banks of Finland, the Czech Republic, Croatia and Romania confirmed this claim. In the case of Sweden, the NSI reported that about 40 enterprises consist of more than 1 legal unit. Nevertheless, selected CompNet data providers (the Netherlands and France) report that using either definition can lead to rather different results. In the Dutch data, about one-third of the enterprises consist of more than one legal entity. For France, existing research suggests that legal units with more than 5000 employees accounted for 13% of the total workforce in 2011, while on the other hand, enterprises with more than 5000 employees had 24% of the workforce (Béguin and Hecquet, 2015).

When it comes to identifying a way to solve the inherent risk posed here for comparability, one has to take into account that the unit of observation of the national data sources is often not a result from a choice but instead is determined by the way a country's administrative, judicial and tax systems are organised. For instance, in numerous countries (Hungary, Slovakia, Slovenia), the legal unit is derived because firms are grouped according to their fiscal reporting unit. The choice for the unit of observation is thus a result of the process of data collection and can in most cases not be adjusted. That being said, we understood from data providers that this issue is being researched. Statistics Sweden is investigating the largest business groups to define enterprises and their underlying units more precisely. Statistics Italy indicated that they are working on the creation of a Business Register that has the enterprise group as unit of observation rather than their current register with the legal unit.

A related issue is that of consolidation, which refers to the cancelling out of intra-firm flows, for example in a consolidated income statement. In general, enterprises that consist of more than one legal unit are

[12] See http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise (accessed on 21 March 2018).

typically large, and therefore these cases account for a relatively large share of economic activity. When intra-firm flows of goods and services are not cancelled out, this has an upward effect on production as compared to that apparent from consolidated enterprise data. Moreover, one may have to be careful to compare averages based on unconsolidated enterprise data with those based on legal units. For example, if a country has unconsolidated enterprise data, where turnover is reported as aggregated across legal units, the *average* turnover across *enterprises* is higher than would have been the case when the data had listed all *legal units* (that is, the same aggregate turnover would have been denominated by a higher number of reporting units).

As displayed in Table 6, data providers report a mixture of consolidated and unconsolidated data. Therefore, the issue of different reporting units represents a potentially relevant problem for cross-country comparability. However, we are not in a position to fully evaluate its impact, apart from the partial insights above. For instance, it is not clear *a priori* how and if correlations and cross-relations are affected.

Recommendations:

- Since the unit of observation will not be changed in the short term, we recommend to identify for which countries this difference between units of observations is relevant to be able to assess the magnitude of possible incomparability. We concluded that for France, the Netherlands and Italy,[13] the difference might be considerable and thus cross-country comparability could be affected. For the Czech Republic, Sweden, Finland, Croatia and Romania, the difference between definitions of units of observation is negligible.

3.2 Linking and comparability of input sources over time

This section discusses the linking of different databases and associated potential issues of comparability. In particular, we will focus on combining information from separate data sources at the firm level, as well as on longitudinal linking, i.e. following firms over time. In addition, a short section discusses threats to comparability over time due to changes in the data collection.

[13] In Italy, the difference arises due to the existence of multiple VAT codes (legal units) for firms that are active in foreign markets.

3.2.1 Linking different data sources within countries

Nearly all countries participating in CompNet combine the input data from different sources, resulting in the database covering various areas, ranging from production to corporate finance and trade. Limitations on the ability of linking these firm-level data could potentially affect the quality and comparability of the database. However, all data providers report to be able to link the various sources in their country together based on a common firm identifier, which therefore does not cause any concern. Data providers did report varying ways of tracking firms across databases, such as legal identification numbers in case of Spain or tax codes as is the case with Hungary, but this should have no effect on comparability as long as identifiers are unique and linking is possible.

3.2.2 Longitudinal linkages

For the analysis of growth and business dynamics, it is necessary to be able to follow firms over time. While all countries report that there is a constant and unique firm identifier in the data that allows linkage across data sources (see Subsection 3.2.1), and also over time, some countries have pointed out that M&A, and conversely the splitting up, or other reorganisations of business units, also in the statistical process, are not explicitly taken into account. As for the majority of firms, this does not play a role and has on average only a moderate impact on the calculation of firm-level growth figures. This issue does, however, affect the analysis of business dynamics, for instance the growth and survival analysis of small firms, keeping in mind the particular growth patterns of these firms documented by e.g. Geurts and van Biesenbroeck (2014).

It is fair to say, however, that not many empirical studies actually take this into account, especially in a cross-country setting. However, most NSIs nowadays complement their business registers with business demography information, which is also reported to Eurostat. The OECD DynEmp and Multiprod exercises rely on such “event information” to distinguish actual firm births and deaths from artificial entry and exit in countries where this is possible. It has to be taken into account that this solution is only available in the case of well-equipped data providers with access to such event information. Since some data providers will not have the information needed for this approach, it risks introducing a solution for only a part of the participating countries.

Recommendations:

- It is recommended to take stock of whether data providers can distinguish actual firm births and deaths from artificial entry and exit, and if so to adjust the business dynamics analysis to take this information into account. From our survey, we conclude that at least the data providers from Hungary, the Netherlands, Slovakia, Italy, Sweden and Spain have this event information

available. To the contrary, the data providers from Finland, Croatia, Germany and the Czech Republic do not have this event information available.

3.2.3 Other known quality issues and breaks in input data

Acknowledging that the issues identified in the metadata surveys are necessarily non-exhaustive, it included open questions allowing data providers to report on any breaks due to changes in the data collection, deviations from the broad project definitions of the input variables and other known quality issues in the input datasets. Regarding breaks, some countries indeed report periods that are not fully comparable. By way of reference, responses to these answers have been consolidated into country notes that are included in the Annex, and they can be consulted by the database users.

Recommendations:

- A harmonised way of data providers updating CompNet on these breaks could prove to be a good exercise.

3.3 Assessment of comparability of input variable definitions

After looking at the characteristics of the input sources, this section turns to the definition of the variables available in each of the sources. Harmonising as much as possible the definition of the input variables across countries is an obvious starting point for comparability. However, not each data provider may have available information according to an "ideal" definition, if it exists, for example due to the fact that administrative and statistical information is collected for various purposes (not usually for economic analysis), with the exact definition depending on the goal of the measurement, the type of source and practical considerations regarding the availability of information to respondents. Administrative and tax systems also differ country by country, and sometimes the exact definition of a variable could therefore be slightly different under alternative regimes.

This section addresses the following aspects:

1. Currency units
2. Transformation to NACE Rev.2
3. Measurement of employment
4. Calculation of value added
5. Inclusion of taxes and subsidies on production variables
6. Timing of the variables
7. Other variable-specific issues

3.3.1 **Currency units**

An obvious requirement for comparability is that variables are denominated in the same unit of measurement. At the most basic level, monetary variables should be expressed in the same currency, and in the same magnitudes (thousands, millions et cetera). The CompNet data collection requires monetary variables to be denominated in thousands of Euros. All countries fulfil this requirement. For example, the data for the Czech Republic and Poland was converted from the national currencies to Euro by the data providers using the annual average exchange rate of the domestic currency versus the Euro, and similarly for pre-2015 data for Lithuania, and for Slovakia from before 2009. More information on the conversion of the 10 countries that have or have had different currencies can be found in the country-specific notes in Annex 8.5.

3.3.2 **Measurement of employment**

Employment can be reported in full-time equivalents (fte) or headcount. Moreover, some sources may include employees only, while other sources count all persons employed (including proprietors and unpaid working family members as well). Both potential differences affect the measurement of "per worker" variables such as labour productivity, and in addition the size-class classification, especially of smaller firms.

Table 7. Definitions of employment by country in input data

Country	Headcount	fte
Belgium		
Croatia		x
Czech Republic	x	
Denmark		x
Germany		x
Spain	x	
Finland		x
France	x	
Hungary	x	
Italy		x
Lithuania	x	
Netherlands		x
Poland	x	
Portugal	x	
Romania	x	
Slovakia		x

Table 8. Definitions of employment by country in input data

Country	Persons employed	Employees
Belgium		
Croatia		x
Czech Republic		x
Denmark		
Germany		x
Spain		x
Finland		x
France		x
Hungary		x
Italy	x	
Lithuania		x
Netherlands	x	
Poland	x	
Portugal		x
Romania		x
Slovakia		x

Table continued

Slovenia	x	Slovenia	x
Sweden	x	Sweden	x

Notes: Information drawn from survey.
Within countries, no heterogeneous definitions.

Notes: Information drawn from survey.
Within countries, no heterogeneous definitions.

The survey results reported in Table 7 suggest that there is no common definition used by the majority of countries, and that, in fact, there exists some heterogeneity among countries. As can be seen form Table 7, 8 countries indicated using the headcounts and 9 countries indicated using the fte definition. Among the latter group, Croatia, Italy, Sweden and Slovakia are found to also have information available on the headcount. For, Finland and the Netherlands this information is available only for a sub-sample of the population. Taking this into consideration, the common denominator is the headcount. Table 8 shows the countries that count all persons employed (including proprietors and unpaid working family members) and those countries that count the employees only. Only Italy, the Netherlands and Poland, count persons employed. Of those countries, only Italy has information available on the employees only as well. Overall, the employees-only definition is the most widely available.

While we are unable to quantify the impact on the results of using different definitions of employment, the direction of the effects is clear: the number of employees is always smaller than the persons employed and the number of fte is always smaller than the headcount. For example, assuming all else as equal, the labour productivity measure in Poland and Lithuania (based on headcount of persons employed) is expected to be lower compared to Sweden (based on fte and employees). Moreover, large firms are less affected as a marginal unit of labour has a smaller relative impact on total employment for these firms, and therefore the assignment to size classes is less affected, and the same holds for employment as the denominator in “per worker” variables. Finally, groups of countries with similar definitions can of course always be compared.

These caveats about the employment variable mainly concern the comparison of indicators in levels. It is likely that the comparison of growth rates as well as correlation analyses is less affected.

Recommendations:

- Taking into account the additional information the countries possess, the most commonly available measure across countries is the one based on the headcount and the employees-only definitions. Data providers are encouraged to, when the information allows them, switch to these specifications of the employment variable to reduce overall variance.

- For the purpose of productivity estimation ideally labour input is measured in terms of full-time equivalent and the persons employed definitions. Taking this into account it is recommended to investigate which countries have sources available containing that information.
- To overcome differences between fte and headcount, as well as between persons employed and employees, their ratios could be determined from (possibly industry-level) aggregate data, which could be used in the code as a scaling factor to make the employment figures more harmonised. Such a routine could be tested in countries where more than one measure of employment is available.
- In order to quantify the effect on the results, the CompNet code could be run for various definitions of employment in countries where this is available.

3.3.3 Calculation of value added

In national data sources, value added is usually a derived variable, meaning it is constructed from other variables collected in the data sources. Eurostat has a detailed description on which terms to include in this calculation.[14] This calculation is listed in column 1 of Table 9 and contains the following terms: turnover, capitalised production, other operating income, increases and decreases of stocks, purchases of goods and services, taxes linked to turnover and taxes linked to production. This subsection compares the value-added definitions used across the participating countries, as well the Eurostat definition.

Table 9: Value added, Eurostat calculation and CompNet calculations

Eurostat calculation value added		BE	SE	DK	CZ	DE	RO*	NL	HR	HU	IT	SK	FI	LT	FR	PT	ES	PL	SI
Turnover	+		Y		Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Capitalised production	+		Y		N		N	N	N	Y	Y	N	Y	Y	Y	Y	Y	N	N
Other operating income	+		Y		N		N**	N	N	Y	Y	N	Y	Y	N	Y	Y	N	N
Increases of stocks	+		Y		N		N	N	N	Y	Y	N	Y	Y	Y	Y	Y	N	N
Decreases of stocks	–		Y		N		N	N	N	Y	Y	N	Y	Y	N	Y	Y	N	N
Purchases of goods and services	–		Y		Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

[14] Please refer to Eur-Lex website for the Commission Regulation underlying the value added calculation: http://eur-lex.europa.eu/eli/reg/2009/250/oj (accessed 23 March 2018).

Table continued

Other taxes on products which are linked to turnover but not deductible	–		N		N		N	N	N	N	Y	N	Y	Y	N	N	N	N	N
Duties and taxes linked to production	–		Y		N		N	N	N	N	Y	N	Y	Y	N	Y	Y	N	N

Notes: *Romania, deviating from Eurostat, includes trade discounts into the value-added calculation.
** Romania, other operating income excluded except for subsidies related to turnover.
No information available for Belgium and Germany.

Table 9 shows that, as might be expected, most countries would be able to derive value added from turnover minus the purchases of goods and services. However, the inclusion of the various other sub-items varies across countries. The value-added definition matches the SBS standard in 4 countries (Hungary, Italy, Finland and Lithuania). The extent to which these deviations lead to differences in value-added levels depends on the share of a particular term in total value added. Compared to turnover and intermediate inputs, these sub-items will generally be small. The variance observed across the different value-added definitions is therefore likely to affect the comparability of levels of value added only mildly. In addition, growth rates are less affected.

Recommendations:

- Whenever possible use multiple output variables by way of robustness check
- In CompNet common code, next to the value-added constructed in the data sources, derive value added as turnover minus purchases of goods and services, which all countries have available
- Quantify the share of each of the value-added sub-items, for countries where these data are available.
- In follow-up metadata collections, obtain information about the components of turnover, in particular which of the elements of value added may be contained in turnover and are not available individually.

3.3.4 Definition and valuation of production variables

Another aspect is specific to the variables related to production (value added, turnover, intermediate inputs and labour cost). These variables may or may not include taxes and subsidies. For output, these taxes and subsidies could be either on products or production (i.e. valuation against market or basic prices, or at factor cost). Labour cost could include wages and salaries only, or the total of labour cost,

including social security contributions by the employer and wage tax. In addition, turnover and intermediate inputs may or may not include goods for resale. Variation in the inclusion of these different components could potentially affect comparability, although the extent to which depends on their relative magnitude, and growth rates and correlations are not likely to be affected.

Valuation of output and intermediate inputs

Table 10 shows which type of valuation is used in the various countries for turnover, value added and intermediate inputs. The following scheme is used to classify the responses to the metadata survey question on this issue:

Factor cost

Plus taxes on production, less subsidies on production

= Basic prices

Plus (non-VAT) taxes on products, less subsidies on products

= Market prices

Please note that since taxes/subsidies on production do not play a role in intermediate inputs, valuation at factor cost is not relevant for these variables.

In practice, Table 10 shows that, as reported by the data providers, sources provide several hybrid forms, where for example taxes on product/production are contained in the valuation, but subsidies have not been subtracted or vice versa. The valuation of turnover and value added is mostly consistent within countries, however, except in Lithuania and Denmark. More than half of the countries for which information is available report the use of turnover and value added against factor cost, though in three of these cases (Netherlands, Portugal, and Spain) the valuation does include taxes on products. Poland reports the use of market prices, although product taxes/subsidies are not taken into account.

Also for intermediate inputs, we observe great variety in the valuation reported. Four countries use market prices; 3 countries use basic prices, while the rest report the use some hybrid form, where either only taxes or only subsidies are included.

The degree to which these differences in valuation impact comparability depends largely on the magnitude of the pertinent taxes and subsidies, which may be product- and industry-specific (e.g. excise on alcohol and tobacco). When taxes and subsidies can be assumed to be more or less constant over time, however, growth rates should be, and correlations are likely to be unaffected, especially if the econometric analyses control for time- and sector-specific effects.

The question whether subsidies/taxes can be excluded/included to arrive at a “greatest common denominator” definition for each variable was answered negatively by data providers from Italy, Slovakia

and Finland. Given the fact that the valuation method is anchored in the data source structure, data providers do not have flexibility to adjust the valuation.

Recommendations:

- The information provided in the table is preliminary due to existing uncertainties regarding the relation of the question posed in the table and the accounting standards of, for instance, the IFRS. CompNet could draw upon existing accounting standards and define the valuation and definition of their variables based on these harmonised accounting standards.
- Quantification of the impact that different valuation methods have on values of turnover, value added and intermediate inputs, possibly through case studies for countries where different valuations are available.

Table 10: Valuation of turnover, intermediate inputs and value added

	Turnover					Value Added					Intermediate inputs		
	Non-VAT taxes on products added to basic price of product (e.g. excise)	Subsidies on products subtracted from basic price of product	Subsidies on production subtracted (factor cost valuation)	Taxes on production added (factor cost valuation)	Implied valuation	Non-VAT taxes on products added to basic price of product (e.g. excise)	Subsidies on products subtracted from basic price of product	Subsidies on production subtracted (factor cost valuation)	Taxes on production added (factor cost valuation)	Implied valuation	Non-VAT taxes on products added to basic price of product (e.g. excise)	Subsidies on products subtracted from basic price of product	Implied valuation
Belgium													
Croatia													
Czech Republic	No	No	No	No	Factor cost	No	No	No	No	Factor cost	No	No	Basic prices
Denmark	No	No	No	No	Factor cost	?	?	?	?		Yes	Yes	Market prices
Germany	Yes	Yes	Yes	Yes	Market prices	Yes	Yes	Yes	Yes	Market prices	Yes	Yes	Market prices
Finland	No	No	No	No	Factor cost	No	No	No	No	Factor cost	No	No	Basic prices

Table continued

France	No	No	No	No	Factor cost	No	No	No	No	Factor cost	No	No	Basic prices
Hungary	No	No	No	No	Factor costs	No	No	No	No	Factor costs	No	No	Basic prices
Italy	No	No	No	No	Factor cost	No	No	No	No	Factor cost*	Yes	Yes	Market prices
Lithuania	No	Yes	Yes	Yes	Basic prices corrected for product subsidies	No	No	No	No	Factor cost	No	No	Basic prices
Netherlands	Yes	No	No	No	Factor cost but corrected for product taxes	Yes	No	No	No	Factor cost but corrected for product taxes	Yes	Yes	Market prices
Poland	Yes	Yes	No	No	Market prices but not corrected for taxes/subsidies on production	Yes	Yes	No	No	Market prices but not corrected for taxes/subsidies on production	Yes	Yes	Market prices
Portugal	Yes	No	No	No	Factor cost but corrected for product taxes	Yes	No	?	?	?	Yes	No	Basic prices including product taxes
Romania													
Slovakia	No	Yes	Yes	No	Market prices corrected for subsidies	No	Yes	Yes	No	Market prices corrected for subsidies	No	Yes	Basic prices corrected for subsidies
Spain	Yes	No	No	No	Factor cost but corrected for product taxes	Yes	No	?	?	Factor cost but corrected for product taxes	Yes	No	Basic prices including product taxes

Table continued

Slovenia	No	No	No	No	Factor cost	No	No	No	No	Factor cost	Yes	No	Basic prices including product taxes
Sweden	No	Yes	No	No	Factor cost but corrected for product subsidies	No	Yes	No	No	Factor cost but corrected for product taxes	No	Yes	Basic prices corrected for subsidies

Notes: *In the Italian data, value added includes other revenues (that include subsidies on production) and other operating expenses are subtracted (taxes on products not deductible, duties and taxes linked to production).
No information available for Croatia, Belgium, Romania.

Labour taxes/subsidies

Turning to taxes and subsidies included in the labour cost variable, Table 11 shows that in none of the countries are subsidies included, except in Slovakia.

By contrast, data providers reported that the labour cost variable includes taxes in 7 countries (Croatia, Czech Republic, Italy, the Netherlands, Poland, Slovakia and Sweden).

Labour tax can be a relatively large fraction of gross wages in Europe. Therefore, when comparing levels of labour cost across countries or indicators based on this variable, it is wise to bear in mind these differences. However, again, when constant over time, growth rates and correlation analyses are likely to be affected only mildly, if at all.

Recommendations:

- If a single definition is not feasible, an assessment could be performed on the magnitude of subsidies and taxes in the total value of labour costs across countries. This information can be used to add a generic correction to the labour cost variable in the relevant countries.
- It is recommended to investigate the valuation of labour cost in light of accounting standards. In terms of accounting standards, considerable harmonisation efforts have been undertaken, and using an accounting definition will possibly lead to a common denominator.
- For future metadata exercises, it is recommended to explicitly survey the inclusion of contributions to social security, separately from retained wage taxes. In this respect, it is also worthwhile to investigate the different tax and social security system across the European countries

Table 11. Labour subsidies and taxes included

	Subsidies subtracted	Taxes added
Belgium*		
Croatia	No**	Yes
Czech Republic	No	Yes
Denmark	No	No
Finland	No	No
France	No	No
Germany	No	No
Hungary	No	No
Italy	No	Yes
Lithuania	No	No
Netherlands	No	Yes
Poland	No	Yes
Portugal*		
Romania	No	No
Slovakia	Yes	Yes
Slovenia	No	No
Spain*		
Sweden	No	Yes

Source: information from survey.

Notes: there is no within-country variation.

* For PT, BE and ES, no information available.

**In principle, subsidies are not included in Croatia, although full information on this is not available.

Goods for resale

Table 12 reports on whether turnover and intermediate inputs include goods for resale. All countries except France report including these goods. When using these variables for the calculation of value added, the treatment of goods for resale should be consistent, which is the case for all countries. (That is, value added is net of goods for resale when these goods are either both included or excluded from turnover and intermediate inputs, but not when included only in turnover or intermediate inputs.) From this we can conclude that the treatment of goods for resale does not pose any problems for comparability, only with respect to turnover and intermediate inputs in France, but not for the calculation of value added.

Recommendation:

- Investigate whether goods for resale can be included in turnover and intermediate inputs in France.

Table 12: Turnover and intermediate inputs, goods for resale included.

Country	Turnover	Intermediate inputs
Belgium		
Croatia		
Czech Republic	Yes	Yes
Denmark	Yes	*
Finland	Yes	Yes
France	No	No
Germany	Yes	Yes
Hungary	Yes	Yes
Italy	Yes	Yes
Lithuania	Yes	Yes
Netherlands	Yes	Yes
Poland	Yes	*
Portugal	Yes	Yes
Romania	Yes	Yes
Slovakia	Yes	Yes
Slovenia	Yes	Yes
Spain	Yes	Yes
Sweden	Yes	Yes

Notes: No information available for HR and BE and partial information available for DK and PL.

3.3.5 **Timing of variables**

Sources can differ in the way they deal with timing of variables. Some variables may refer to the status at the end-of-year (e.g. balance sheet data); other variables may be recorded as an average or total over the year. Differences across countries in the timing of variables impacts comparability. For example, average employment is lower than total employment over a year, with an obvious effect on

indicators drawing upon the number of workers.[15] In addition, there is the issue of seasonality. Consider an industry with strong seasonal differences in labour inputs. If a firm in such an industry was measured by average employment over the year, it would report a higher number of employees than when the number of employees is reported outside the busy season, usually the end of the accounting period.

Table 13, which is listed in the Annex, shows the reference period for all the raw variables for the countries that responded to the surveys. The answer categories are divided into three groups: a point-in-time (containing end of the year status and other dates), annual totals and annual averages, with any deviations listed in the table notes. For a few key variables such as labour costs, number of employees, value added and raw materials, the answers vary quite considerably. For the number of employees for instance, the answers in the survey range from annual average (Finland, Czech Republic, Slovakia, Romania) and annual totals (Hungary) to point-in-time (Lithuania, Netherlands, Portugal, Spain), which usually refers to 31 December. Overall, in most cases, variables refer to a point in time, specifically the end of the accounting period, with several countries leaving room for firms to have deviating book years, although these are likely to be exceptions.

Recommendations:

- More research can be done to investigate the impact of these differences on comparability. A first step could be to explore whether within countries, information can be gathered on variables using more than one definition, and to compare the levels and the impact on indicators using different timing.
- Another possibility to quantify the impact is to group countries that use the same definition, and subsequently assess differences in levels across these groups, and check whether certain correlations hold equally in each of the groups, or that certain results tend to be weaker or stronger for a particular definition of the variables involved. We will revisit this point in Section 5.3.

3.3.6 Other variable-specific issues

Accounting for the possibility that our survey questions regarding input variables were not exhaustive, we included questions addressing possible deviations from the CompNet definition. These variable-specific deviations are listed in each country's specific notes in the Annex. This information is also included in the metadata mapping introduced in Subsection 2.5.

[15] Such an effect on the ratio of two variables might be mitigated if the numerator variable has the same timing as the denominator.

4. CompNet 6th vintage: the output side

After considering the sources and variable definitions, this section considers the "output side", i.e. the actual CompNet cross-country dataset of micro-founded variables and indicators available to researchers, which will be compared with population information as published by Eurostat.[16] The following aspects are documented: (1) coverage versus published figures of Eurostat; (2) representativeness (before and after reweighting) versus the published figures of Eurostat; (3) comparison of CompNet indicators with other published indicators from various sources. While these aspects are indicative of the overall quality of the database, differences across countries touch upon comparability as well.

For the output side, by *coverage***,** we mean the share (or sub-population) that is covered with respect to the corresponding population figure (or sub-population) according to Eurostat, specifically in terms of number of firms, total employment or output. By *representativeness*, we mean the extent to which the various segments of the population (by firm size, industry and so on) are proportionally reflected.

Three preliminary caveats should be kept in mind, which are as follows:

1. First, the comparison should be restricted to the target population of the source. For example, when the financial sector is not included in the industries covered in the source, it should be excluded from the Eurostat figures as well.

2. Second, the *definitions* used in the source data may differ from the statistical concepts used in the Eurostat data:

- **Number of firms**: Eurostat publishes figures on the number of *enterprises*. As discussed in Subsection 3.1.4 and shown in Table 6, some of the CompNet data providers use sources with the *legal unit* as the basis for data collection.
- **Employment**: Eurostat publishes figures related to *persons employed*, i.e. "the total number of persons who work in the observation unit (inclusive of working proprietors, partners working regularly in the unit and unpaid family workers), as well as persons who work outside the unit who belong to it and are paid by it (e.g. sales representatives, delivery personnel, repair and maintenance teams)". In particular, this differs slightly from the number of employees, and one should be cautious about the definition of lower size classes, where counting working proprietors and so on directly impacts the firm's classification. The definition Eurostat uses is the headcount, i.e. "the total number of persons". Subsection 3.3.2 showed that some countries use fte to define

[16] For Europe, the official statistics on the number of enterprises, employment, by industry and size classes, are collected from the National Statistical Institutes (NSIs) and published on an annual basis by Eurostat.

their employment variable. This introduces a small variance with the Eurostat figures. Finally, Eurostat figures also refer to annual averages. In Subsection 3.3.2, it was shown that there is considerable variance in the way employment is reported in the input sources used. This not only affects the comparison of employment figures, but also of statistics by size class.

3. The CompNet data are subject to the outlier cleaning procedure. This routine reduces the number of observations that end up in the ultimate database, relative to the input sources.[17]

4.1 Firm and employee coverage compared to Eurostat

Table 14 shows the percentage of firms covered by CompNet, relative to the Eurostat Structural Business Statistics. On average, CompNet covers about 58% of the corresponding population of firms although with large cross-country variation, ranging from 11% of firms in Italy to about 90% in Slovakia and the countries providing only information on firms with more than 20 employees (for which we compare to the relevant size classes). Coverage in terms of firms, however, may not represent a problem *per se*, provided that the samples are representative.

In terms of employment coverage analogous to the firm coverage, we observe large country heterogeneity. This ranges from 25% in the Spain to 86% in Slovakia. On average, the CompNet 6th vintage dataset covers around 59% of the corresponding population of employees.

Table 14: Firms and employee coverage ratios relative to Eurostat

Country	Employment (%)	Number of firms (%)
Belgium	44	19
Croatia	52	38
Czech Republic	72	72
Denmark	53	87
Finland	50	45
France	57	41
Germany**	45	13
Hungary	57	44
Italy	39	11
Lithuania	69	37
Netherlands	35	18

[17] Please refer to the Annex for a detailed description of the outlier cleaning procedure as well as the confidentiality procedure present within the CompNet code.

Table continued

Poland	75	74
Portugal	56	31
Romania	68	76
Slovakia*	86	90
Slovenia	50	28
Spain	25	15
Sweden	40	32

Note: representativeness is measured in 2011, number in parenthesis refer to the figures in Eurostat Structural Business Statistics. Due to data availability, coverage for Germany and Spain is measured in 2012 and Slovakia 2014.
* based on the 20e sample.
** based only on manufacturing.

4.2 Employment coverage across industries versus Eurostat

Table 15[18] shows that CompNet covers, on average, a rather large share of total employment as measured by published figures of Eurostat, although there is some heterogeneity across countries, and the coverages ratios of manufacturing are higher compared to construction and services. The latter observation underlines that reweighting should be carried out using industry- and indicator-specific weights, as is currently the case in the CompNet code.

Table 15: Employee coverage of CompNet vs Eurostat across macro-sectors

Country	Manufacturing (%)	Construction (%)	Services (%)
Belgium	71.9	46.3	70.4
Croatia	85.5	52.7	-
Denmark	58.0	65.2	92.9
Finland	65.8	54.1	71.5
France	71.6	69.4	85.6
Hungary	72.7	58.7	84.1
Italy	60.7	35.3	53.6
Lithuania	98.2	76.9	-
Netherlands	61.2	42.2	54.0
Portugal	80.2	60.1	65.3

[18] We would like to point out to the reader that table 14 and table 15 are based upon different dimensions of the CompNet dataset. Where the coverage figures of table 14 are based on the "country level", the coverage figures of table 15 are based on the "macro-sector level". This means that the outlier procedures differ across these two dimensions and causes that a weighted average of the figures in table 15 will not necessarily coalesce with the figures in table 14.

Table continued

Romania	78.7	71.9	-
Slovenia	81.3	38.4	91.2
Spain	40.4	28.6	36.8
Sweden	44.1	47.3	69.0
Czech Republic*	72.9	25.0	68.0%
Germany**	45.0	-	-
Poland*	70.6	29.7	64.5
Slovakia*	79.8	21.8	87.3

Notes: coverage is measured in 2011, number in parenthesis refer to the figures in Eurostat Structural Business Statistics.
* Figures rely on the 20e sample.
**Information only available for the manufacturing sector.

4.3 Representativeness versus Eurostat

Next to coverage, a defining feature determining database quality is how these observations are spread across the economy — i.e. the representativeness of the data. In fact, the number of observations can be increased, but when newly added observations are concentrated in certain industries only, overall representativeness will deteriorate.

To this end, the CompNet 6th vintage uses a reweighting procedure, which will be discussed briefly in Subsection 4.3.1. The following sections discuss in turn the representativeness of the output database in three ways, i.e. across size classes (Subsection 4.3.2), across industries (Subsection 4.3.3), and within-cell (Subsection 4.3.4), which is a test for the characteristics of the sampled firms in a given sector size-class combination.

4.3.1 CompNet reweighting procedure

The CompNet dataset aims to enable researchers to look beyond simple aggregations in the firm population. Thus, all eventual indicators should be understood as attributes of the underlying firm population and not of a sample, unless explicitly stated otherwise. To achieve this, CompNet output is weighted with so-called inverse probability weights: using data available from Eurostat, the number of firms in a given size class and NACE Rev. 2 macro-sector is gathered.[19] From the shares of each cell, the sampling probabilities of firms in all sector size-class combinations are calculated. Comparing these sampling probabilities to the actual shares in the CompNet data, weights have been calculated that are

[19] In some cases, Eurostat does not provide the firm population data needed for this exercise. In that case, the data provider is asked to provide the population information. This was, for example, the case in Croatia.

applied to adjust the CompNet figures to match the population shares. This reweighing procedure hinges on the assumption that there is no selection into reporting *within* these bins. Important here is that since some indicators require multiple variables and/or lagged values, these weights are constructed separately for every indicator. The weights may therefore differ considerably across indicators.

The rationale to implement this centralised weighting method using Eurostat is to apply a homogeneous weighting procedure to all data sources as well as to save additional work from the data providers. However, following the conclusions of the Section 3, applying this method introduces two potential distortions. First, a potential problem of this approach is that the aggregation level (macro-sector size class in the case of CompNet) might not be the level at which inclusion probabilities of the survey vary. In other words, if the sample is drawn to be representative of the population, but is done so on a different dimension than the macro-sector size class, the current weighting method will not be able to correct adequately for this.

Second, the weighting is done based on information from the structural business statistics published by Eurostat. As indicated in the introduction to Section 4, statistical concepts used by the national data sources may differ from the ones used in Eurostat, which introduces a bias in the weighting procedure.

In the case of Croatia, Eurostat does not provide enough information to derive the weights. Instead, they supply weights drawn from the FINA, which in essence functions as a business registry. This decentralised alternative to the weighting procedure ensures that weights are derived along the same dimensions and using the same statistical definitions as the source that provides the data. It can therefore be a worthwhile venture for CompNet to consider this decentralised weighting procedure.

One requirement for this is that data providers have the national business registries available to derive the weights from. Apart from Croatia, the Czech Republic, Slovakia, Italy, the Netherlands, Germany and Slovenia registers are available to use for this procedure.

Recommendations:

- Investigate whether for the remaining countries business registers are available.
- Consider the change from centralised to decentralised weighting procedure using the national business registries.

4.3.2 Representativeness across size classes after reweighting

Table 16 shows the share of employment across size classes between CompNet and the population (again, sourced from Eurostat). On average, the size-class shares in CompNet dataset are close to those in Eurostat, with some exceptions (Sweden, Czech Republic and Portugal). Table 17 shows the share of enterprises across size classes between CompNet and the population derived from Eurostat.

The observed differences we saw in Table 16 are less observed in Table 17. Risks of incomparability under this dimension appear to be limited, although for the countries mentioned, the results are probably biased towards specific size classes. Comparability across size classes is likely to be further improved when the employment variables on which the size-class classification is based are more aligned across countries, and compared to the Eurostat definition.

Table 16: Share of employment across size classes

Country / size classes	1	2	3	4	5
Belgium	21,5% (26,3%)	12,8% (7,8%)	20,3% (12,4%)	24,4% (16,8%)	20,8% (36,5%)
Croatia	27,2% (9,3%)	13,9% (13,0%)	17,4% (15,1%)	26,4% (27,8%)	14,9% (34,6%)
Denmark	41,1% (23,0%)	14,0% (9,6%)	17,8% (12,6%)	19,3% (21,6%)	7,57% (33,1%)
Finland	28,7% (28,2%)	14,0% (8,73%)	18,5% (11,2%)	24,9% (18,4%)	13,7% (33,3%)
France	30,4% (25,7%)	14,5% (8,1%)	19,2% (11,3%)	24,9% (15,9%)	10,8% (38,9%)
Hungary	37,2% (37,1%)	15,1% (8,6%)	15,4% (9,3%)	20,5% (16,7%)	11,6% (28,1%)
Italy	23,0% (41,0%)	18,3% (11,8%)	21,2% (10,8%)	25,7% (14,2%)	11,6% (21,8%)
Lithuania	23,4% (28,8%)	13,9% (11,1%)	20,2% (15,7%)	29,4% (23,0%)	12,8% (21,1%)
Netherlands	16,9% (26,2%)	13,4% (8,5%)	20,1% (11,5%)	30,1% (20,9%)	19,2% (32,8%)
Portugal	36,5% (32,1%)	16,4% (11,8%)	19,4% (13,7%)	19,8% (18,4%)	7,66% (23,8%)
Romania	29,3% (21,7%)	13,3% (8,2%)	18,4% (12,4%)	28,2% (23,3%)	10,5% (34,2%)
Slovenia	24,3% (37,0%)	11,7% (10,0%)	16,3% (8,5%)	28,6% (22,6%)	18,8% (21,8%)
Spain	33,3% (37,7%)	17,2% (9,54%)	20,2% (11,4%)	17,2% (14,6%)	11,8% (26,6%)
Sweden	39,8% (21,9%)	17,7% (9,7%)	22,3% (13,4%)	17,9% (20,0%)	2,01% (34,8%)
Czech Republic*	-	-	16,2% (16,5%)	38,3% (32,8%)	45,3% (50,5%)
Germany*	-	-	5,06% (7,3%)	27,5% (24,7%)	67,2% (53,4%)

Table continued

Poland*	-	-	13,4% (13,6%)	40,2% (34,4%)	46,2% (51,9%)
Slovakia*	-	-	13,9% (14,6%)	34,3% (32,9%)	51,7% (52,4%)

Note: representativeness is measured in 2011, number in parenthesis refer to the figures in Eurostat Structural Business Statistics
* Figures rely on the 20e sample

Table 17: Representativeness in terms of number of firms

Country / size classes	1	2	3	4	5
Belgium	79,0% (88,4%)	10,4% (5,6%)	7,3% (4,0%)	2,8% (1,6%)	0,3% (0,4%)
Croatia	83,2% (69,1%)	9,1% (16,3%)	5,0% (8,7%)	2,3% (4,7%)	0,3% (1,1%)
Denmark	91,5% (84,4%)	4,7% (7,8%)	2,7% (4,8%)	1,0% (2,4%)	0,1% (0,5%)
Finland	86,4% (84,5%)	7,1% (8,0%)	4,3% (4,7%)	1,9% (2,2%)	0,2% (0,5%)
France	83,6% (87,9%)	8,9% (6,2%)	5,2% (3,9%)	2,1% (1,6%)	0,2% (0,4%)
Hungary	89,5% (93,9%)	6,3% (3,4%)	2,9% (1,6%)	1,2% (1,0%)	0,1% (0,2%)
Italy	68,7% (90,5%)	17,8% (5,9%)	9,3% (2,4%)	3,7% (1,0%)	0,3% (0,2%)
Lithuania	76,4% (87,0%)	11,8% (6,2%)	7,7% (4,1%)	3,6% (2,3%)	0,3% (0,3%)
Netherlands	74,5% (82,4%)	12,5% (8,6%)	8,4% (5,4%)	4,0% (3,0%)	0,4% (0,5%)
Portugal	85,3% (88,9%)	8,5% (6,3%)	4,5% (3,2%)	1,6% (1,4%)	0,1% (0,2%)
Romania	84,5% (79,3%)	8,0% (9,7%)	4,9% (6,6%)	2,3% (3,6%)	0,2% (0,7%)
Slovenia	85,1% (91,3%)	7,37% (4,5%)	4,7% (2,5%)	2,4% (1,5%)	0,3% (0,3%)
Spain	83,1% (89,6%)	10,0% (5,8%)	5,3% (3,1%)	1,5% (1,2%)	0,1% (0,2%)
Sweden	86,9% (89,1%)	7,50% (5,6%)	4,25% (3,4%)	1,2% (1,5%)	0,02% (0,3%)
Czech Republic*	-	-	54,6% (58,5%)	37,4% (34,0%)	7,9% (7,4%)
Germany*	-	-	29,9% (12,9%)	49,0% (14,0%)	20,0% (3,5%)
Poland*	-	-	48,3% (53,5%)	43,1% (38,2%)	8,5% (8,3%)

Table continued

Slovakia*	-	-	51,5% (55,3%)	39,6% (36,6%)	8,8% (8,0%)

Note: Representativeness is measured in 2011, number in parenthesis refers to the figures in Eurostat Structural Business Statistics,

* figures rely on the 20e Sample.

4.3.3 Representativeness across industries after reweighting

Using the weights in terms of number of firms by indicator, industry and size class would, in principle, fully align the industry shares in the CompNet database with those in the Eurostat data. However, because of outlier correction and confidentially routines, some discrepancies may arise. Table 17 shows that the construction sector is slightly overrepresented, in contrast to the services sector, which is slightly under sampled. Table 18 shows that the same holds true for the representativeness in terms of employment shares across industries. Therefore, in terms of industry coverage, using the reweighted figures does not raise any concerns about comparability.

Table 18: Representativeness in terms of number of firms across industries

Country	Manufacturing	Construction	Services
Belgium	11,5% (7,28%)	18,0% (51,8%)	70,4% (40,9%)
Croatia	15,2% (11,4%)	13,0% (75,4%)	71,7% (13,0%)
Denmark	7,98% (7,55%)	16,3% (31,9%)	75,6% (60,4%)
Finland	12,7% (7,17%)	19,5% (47,5%)	67,6% (45,2%)
France	11,3% (7,24%)	20,2% (51,0%)	68,3% (41,7%)
Hungary	11,9% (8,56%)	11,8% (19,5%)	76,1% (71,8%)
Italy	23,6% (13,2%)	15,8% (36,1%)	60,5% (50,6%)
Lithuania	11,3% (9,37%)	10,7% (25,5%)	77,8% (65,1%)
Netherlands	9,85% (4,92%)	10,5% (61,8%)	79,5% (33,1%)
Portugal	13,4% (10,1%)	14,0% (26,6%)	72,4% (63,2%)
Romania	12,3% (13,7%)	11,6% (25,0%)	76,0% (61,2%)
Slovenia	15,9% (12,5%)	12,9% (30,3%)	71,0% (57,1%)
Spain	14,5% (9,42%)	16,8% (34,3%)	68,5% (56,2%)
Sweden	10,0% (6,83%)	16,1% (37,9%)	73,7% (55,2%)

Table continued

Czech Republic*	43,0% (47,6%)	10,3% (15,7%)	46,5% (36,6%)
Germany**	-	-	-
Poland*	37,2% (50,9%)	12,1% (18,1%)	50,5% (30,9%)
Slovakia*	40,0% (53,4%)	11,3% (17,2%)	48,6% (29,2%)

Note: representativeness is measured in 2011, number in parenthesis refer to the figures in Eurostat Structural Business Statistics. *Figures rely on the 20e Sample. ** Information only available for the manufacturing sector

Table 19: Representativeness in terms of employment across industries

Country	Manufacturing	Construction	Services
Belgium	11,5% (7,28%)	18,0% (51,8%)	70,4% (40,9%)
Croatia	15,2% (11,4%)	13,0% (75,4%)	71,7% (13,0%)
Czech Republic	7,98% (7,55%)	16,3% (31,9%)	75,6% (60,4%)
Finland	12,7% (7,17%)	19,5% (47,5%)	67,6% (45,2%)
France	11,3% (7,24%)	20,2% (51,0%)	68,3% (41,7%)
Hungary	11,9% (8,56%)	11,8% (19,5%)	76,1% (71,8%)
Italy	23,6% (13,2%)	15,8% (36,1%)	60,5% (50,6%)
Lithuania	11,3% (9,37%)	10,7% (25,5%)	77,8% (65,1%)
Netherlands	9,85% (4,92%)	10,5% (61,8%)	79,5% (33,1%)
Portugal	13,4% (10,1%)	14,0% (26,6%)	72,4% (63,2%)
Romania	12,3% (13,7%)	11,6% (25,0%)	76,0% (61,2%)
Slovenia	15,9% (12,5%)	12,9% (30,3%)	71,0% (57,1%)
Spain	14,5% (9,42%)	16,8% (34,3%)	68,5% (56,2%)
Sweden	10,0% (6,83%)	16,1% (37,9%)	73,7% (55,2%)
Czech Republic*	43,0% (47,6%)	10,3% (15,7%)	46,5% (36,6%)

Table continued

Germany**	-	-	-
Poland*	37,2% (50,9%)	12,1% (18,1%)	50,5% (30,9%)
Slovakia*	40,0% (53,4%)	11,3% (17,2%)	48,6% (29,2%)

Note: representativeness is measured in 2011, number in parenthesis refer to the figures in Eurostat Structural Business Statistics.
*Figures rely on the 20e Sample.
** Information only available for the manufacturing sector

4.3.4 Within-cell firm size bias

Within-cell firm size bias could arise when the size of firms included in a given industry size-class combination (i.e. the "cell") differs on average from the firms in the population for that cell. For example, firms in the sample may, on average, be larger compared to the relevant population, i.e. larger firms are overrepresented. The CompNet code reweights the indicators to replicate the proportions of the cells in the population in terms of number of firms, but when within-cell bias is present this could still lead to a distorted view. Moreover, when this bias is asymmetrical across countries this causes sampling induced variance in indicators.

Table 20 compares the average number of employees of firms in a given macro-sector size-class combination in CompNet relative to the population provided by Eurostat. Overall, the comparison made in the table shows that in the CompNet dataset, there is limited within-cell bias with respect to firm size. In particular, for larger firms, the CompNet data shows lower employment on average compared to the population. A possible explanation for this could be the fact that the CompNet codes include outlier procedures that exclude firms with relatively high levels of employment (Annex 8.2).

Table 20: Representativeness in sector*size classes by country

Belgium					Croatia				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.64 (13.24)	31.23 (30.90)	104.9 (103.8)	598.4 (736.3)	**Manufacturing**	13.44 (13.22)	30.22 (29.88)	106.1 (102.2)	541.6 (548.5)
Construction	13.42 (13.23)	30.31 (29.94)	82.15 (97.46)	312 (514)	**Construction**	13.56 (13.35)	30.21 (29.71)	90.44 (99.08)	361.5 (587)
Services	13.48 (13.27)	30.37 (30.06)	90.19 (98.68)	718.2 (1301.9)	**Services**	13.25 (12.59)	29.84 (26.34)	91.45 (84.95)	387.3 (463.0)

Table continued

Denmark				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	14.09 (13.29)	31.34 (30.87)	98.45 (97.56)	585.6 (779.0)
Construction	13.80 (13.42)	29.83 (29.42)	80.15 (89.66)	n.a.
Services	13.83 (13.44)	29.90 (26.58)	87.52 (97.87)	410.0 (720.3)

Finland				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	14.08 (13.28)	31.20 (30.29)	99.61 (102.1)	501.1 (790.4)
Construction	13.78 (13.77)	29.15 (29.20)	78.66 (90.14)	n.a.
Services	13.84 (14.98)	30.16 (31.54)	90.61 (116.1)	391.4 (886.1)

Hungary				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.56 (13.50)	30.44 (30.43)	90.04 (95.89)	433.1 (500.2)
Construction	13.05 (13.01)	29.52 (29.46)	79.06 (91.88)	376.7 (703.5)
Services	13.11 (13.08)	29.61 (29.68)	82.72 (93.70)	685.3 (997.6)

France				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.51 (13.98)	31.16 (34.57)	103.2 (111.3)	511.8 (846.6)
Construction	13.35 (14.80)	29.94 (32.36)	80.32 (100.9)	509.0 (845.7)
Services	13.30 (16.31)	30.60 (35.93)	95.09 (122.9)	410.7 (1763.)

Italy				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	14.07 (13.33)	31.13 (30.08)	95.51 (96.78)	435.0 (722.3)
Construction	13.66 (12.90)	29.73 (28.81)	75.62 (85.76)	n.a.
Services	13.68 (12.94)	30.41 (29.69)	92.41 (97.66)	525.3 (1167.)

Lithuania				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.65 (13.41)	30.98 (29.91)	98.80 (80.12)	436.9 (441.2)
Construction	13.52 (13.50)	29.97 (30.02)	89.33 (96.26)	374.2 (409.4)
Services	13.23 (13.18)	29.65 (32.31)	86.93 (78.39)	522.9 (784)

Table continued

Netherlands					Portugal				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.70 (14.92)	30.17 (34.59)	95.09 (108.4)	566.3 (606.6)	**Manufacturing**	13.56 (13.50)	30.44 (30.43)	90.04 (95.89)	433.1 (500.2)
Construction	13.54 (14.67)	29.82 (32.19)	89.94 (96.75)	408.3 (770.4)	**Construction**	13.05 (13.01)	29.52 (29.46)	79.06 (91.88)	376.7 (703.5)
Services	13.51 (18.20)	30.03 (39.19)	93.92 (130.9)	560.9 (1278.)	**Services**	13.11 (13.08)	29.61 (29.68)	82.72 (93.70)	685.3 (997.6)
Romania					**Slovenia**				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.73 (13.73)	31.10 (31.08)	106.0 (106.1)	537.2 (698.6)	**Manufacturing**	14.12 (13.38)	31.69 (27.19)	109.4 (104.3)	658.5 (569.8)
Construction	13.44 (13.44)	30.27 (30.16)	89.91 (98.29)	n.a.	**Construction**	13.89 (13.12)	30.32 (29.29)	83.58 (93.38)	n.a.
Services	13.20 (13.32)	29.65 (29.86)	93.98 (101.1)	n.a.	**Services**	13.69 (13.20)	30.11 (16.35)	92.07 (85.63)	409.3 (706.0)
Spain					**Sweden**				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	13.53 (13.38)	29.81 (29.89)	94.26 (99.93)	641.7 (685.0)	**Manufacturing**	13.49 (15.21)	30.65 (33.77)	92.07 (112.1)	326.5 (811.0)
Construction	13.19 (13.71)	29.06 (29.82)	76.53 (94.31)	538.6 (856.5)	**Construction**	13.25 (15.09)	29.04 (32.86)	66.48 (95.86)	n.a.
Services	13.33 (13.19)	29.33 (29.43)	91.86 (100.0)	735.3 (1153.)	**Services**	13.31 (14.49)	29.41 (34.27)	72.69 (111.3)	495.6 (997.6)
Poland*					**Slovakia***				
Sector		**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	n.a.	32.96 (30.34)	111.4 (108.4)	536.3 (652.5)	**Manufacturing**	n.a.	32.34 (30.04)	108.5 (108.4)	678.6 (742.9)
Construction	n.a.	31.45 (29.08)	99.30 (97.49)	415.8 (644.0)	**Construction**	n.a.	31.21 (30.86)	92.77 (93.00)	512.8 (597.2)

Table continued

Services	n.a.	30.93 (29.19)	101.6 (102.4)	732.0 (909.5)	**Services**	n.a.	31.41 (29.75)	97.76 (96.06)	724.9 (785.3)
Germany*					**Czech Republic***				
Sector	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**	**Sector**	**10–19 employees**	**20–49 employees**	**50–249 employees**	**>250 employees**
Manufacturing	n.a.	114.1 (106.4)	685.2 (909.1)	34.35 (34.14)	**Manufacturing**	n.a.	31.42 (30.62)	109.5 (106.1)	564.8 (663.6)
Construction	n.a.	n.a.	n.a.	n.a.	**Construction**	n.a.	29.61 (29.23)	95.36 (92.53)	478.3 (681.6)
Services	n.a.	n.a.	n.a.	n.a.	**Services**	n.a.	29.79 (29.25)	100.7 (98.12)	634.7 (1053.)

This is supported by the observation that the within-cell bias is mostly present for the larger size class macro-sector combinations and roughly has the same magnitude in all countries.

In conclusion, within-cell firm, size bias with respect to firm size may be an issue for higher size classes. However, as all countries seem to be subject to this bias, it may be of less relevance as far as comparability is concerned. It is recommended to further investigate the relation to the treatment of outliers in the next vintage.

4.4 Validation of trends versus other datasets

As for the validation of the output indicators, a separate CompNet team has compared the characteristics, i.e. time evolutions and values of the computed indicators in CompNet against other established datasets. This exercise is described extensively in the new CompNet Cross-Country report[20] where all major CompNet indicators are validated. Table 21 shows which CompNet indicators are validated and to which datasets they are compared to. As a test for the comparability of output indicators to other sources, please find the extensive validation in the cross-country report.

[20] The CompNet report focussing on the new data vintage titled *The new cross-country database of CompNet: novelties and main indicators*, has been published simultaneously with this report. See Lopez-Garcia (2018).

Table 21: Data validation against other datasets

CompNet indicator topic	Validation source
Productivity indicators	EUKLEMS
Unit labour costs	Eurostat
Indicators capturing distressed firms	ORBIS, SAFE survey
Mark-up indicators	Literature
Import and Export information	CEPII BACI
Indicators covering financials topics (e.g. Investment)	Eurostat
Indicators related to wages	Eurostat

4.5 Other topics concerning output side

This section discusses remaining issues that affect cross-country comparability on the output side of the dataset.

4.5.1 Productivity estimation

The productivity indicators incorporated in CompNet can be divided into two groups: non-parametric and parametric. To the former belong the labour productivity and the Solow residual, which are in essence, respectively, a profitability measure and an indicator of production that is not explained by inputs. Given the limited complexity of these two indicators, they are to a large extent comparable across countries and industries.

This is different for the parametric productivity indicators, which are more problematic to compare across countries and industries. For instance, the total factor productivity indicators (TFP) are estimated using production function and industry-specific output elasticities. In general, there is an ongoing debate on how to estimate TFP for industries other than manufacturing, because the functional form of the production function (the mixture of inputs, the flexibility of inputs and the kind of inputs, for example) to estimate TFP may be vastly different for industries other than manufacturing, making the validity of these estimates questionable. Following this, when we want to make cross-industry or cross-country comparisons of TFP estimates, it is important to keep in mind that these figures are computed based on estimated industry-specific output elasticities. The CompNet code estimates these outputs at various levels of aggregation such as the two-digit sector or the one-digit macro-sector.[1] Moving outside this industry or level of aggregation at which the estimation is done means shifting to another estimated

[1] Please refer to the CompNet User guide (2018) for a detailed explanation on how TFP and other productivity indicators are computed.

output elasticity that can account for level differences or jumps of the TFP measure. Strictly speaking, these level differences make it hard to compare TFP levels across industries, or, more generally, across different samples the estimation is done on.

Recommended solutions are to normalise or demean the data rather than to look at the levels. If you are truly interested in level differences of productivity, the more conservative approach would be to look at labour productivity and Solow residual rather than the parametric productivity indicators.

4.5.2 Country-specific factors determining the validity of an indicator

The CompNet indicators are defined and constructed to capture the intended economic phenomenon across all countries. However, considering that countries are inherently different, country-specific factors can be such that the intended indicator does not capture what it is intended to capture at construction. An example here is the indicators capturing distressed firms or "Zombie firms" in Italy. To adequately capture these distressed firms in Italy, it is important to include amortisation since the country-specific circumstances make this amortisation a defining feature of these firms. It is not possible to provide an exhaustive list of these types of examples, but researchers are encouraged to investigate the country-specific circumstances surrounding the indicator of interest.

5. Improvements and recommendations

This section provides recommendations for improvement in the light of future vintages of the CompNet database. For the most part, these recommendations follow from the comparability assessment above.

5.1 The data collection process

The basis for cross-country comparability is laid in the data collection process. CompNet provides an extensive user guide for the output side of the database and instructions for smooth running of the code. While much of the information for the input stage can be taken from this documentation, comprehensive instruction for assembling the input data is currently lacking. The working group recommends the following:

- Prepare and circulate to data providers an improved description of the input variables. When a single definition is not possible or desirable, a list of possibilities should be provided and, if possible, the prioritised definition. This will raise awareness among data providers about different possibilities and measurement issues, and help in making the most appropriate choices. The choice for the preferred definition could also be justified against the backdrop of different regimes such as the economic competitiveness literature or instead accounting standards such as the generally accepted accounting principles (GAAP). To assess in more

detail the impact of differing definitions, the data could be expanded to include more than one definition for selected key variables, for example headcount and fte for employment. Another improvement would be to list the input variables in a more logical order and to highlight relations between variables, in particular if a certain variable is the sum of other ones, for example the difference between short-term and long-term debts. For balance sheet and profit and loss information, a presentation in terms of a balance sheet or profit/loss statement could help in this respect. Overall, in the past, CompNet tried to define variables more broadly to ensure availability at the data providers. However, to improve the dataset further, the network would benefit from precisely defined input variables.

- Instruct country teams to:
    - Provide specific information on the inclusion of industries and size classes (e.g. exclusion of financial sector and exclusion of sole proprietors)
    - Give a preference to the use of raw input data or to explicitly flag when data have been altered, as opposed to using outlier corrections and imputations
    - Share best practices to other data providers that could improve cross-country comparability, i.e. the methodology used to transform sector classification systems to NACE Rev.2 and using the annual average exchange rate to transform monetary variables to Euros.
- Ask country teams to, whenever possible, distinguish actual entry and exit of firms from changes in the organisational structures or statistical events that deter following a business unit over time.
- Ask country teams to provide information on the population of firms. It should be decided in the future that weights are to be calculated from the population micro-data, rather than based on the Eurostat population totals.

5.2 Code alterations

Some recommendations following the assessment of cross-country comparability can be implemented through the CompNet software.

- Following the observation that the definition of a firm can differ across countries, and that not all sources use the Eurostat enterprise definition, using the Eurostat population numbers may not be the ideal way to calculate the sampling weights used for reweighting. Rather, to stay closer to the actual input data, data providers could include a source with the relevant population data that can be used to derive sampling weights that match the definition of the firm. The calculation of the weights would then have to be implemented in the code. Another recommendation would be to investigate employment-based weights as an alternative to weights based on firm counts.

- When countries are able to flag imputations or other adjustments to the raw data such as outlier corrections, the code could be altered to run with and without such observations, or differentiate the treatment of original and adjusted information in other ways.
- In addition, the outlier detection/correction procedure could be adjusted to differentiate between data sources that provide the raw input data versus those sources that have already been cleaned before feeding it into the CompNet code. Since data providers actually see the firm-level information, they are in a better position to accurately target outliers.
- After investigation of the possibility to correct specific firm-level variables using macro-data, associated correction modules could be implemented in the code. Similarly, the magnitude of taxes and subsidies could possibly be derived from aggregate data, providing a way to correct the valuation of firm-level variables such as turnover and value added.
- It is recommended to take stock of whether data providers can distinguish actual firm births and deaths from artificial entry and exit, and if so to adjust the business dynamics analysis to take this information into account.

5.3 Guidance to potential data users

Empirical researchers know that data will never be perfect. As long as measurement error is random, and uncorrelated with the true disturbance of the economic model, there is not much to worry about. Nevertheless, if measurement error varies systematically with specific units, or groups of units, then the analysis could become biased, and it is necessary to take these biases into account in the analysis or restrict the analysis to units that can be mutually compared.

General precautions

Section 3 has described in detail the metadata by country, source and variable. As described in Subsection 2.5, this information can be mapped to any specific analysis using the linkable metadata tables that will be provided with the CompNet database. Listing all the variables used in an analysis, one can link to the metadata tables and compare the metadata across countries and different subsamples. This allows researchers to distinguish sub-samples that are fully and less comparable, and everything in between. An overview of areas of caution is provided in this report. In this way, the analysis can be subject to different sensitivity and robustness checks. Moreover, it becomes clear where differences might be attributed to real economic or institutional differences, and where such statements are not opportune due to the nature of the data.

Controlling for time-invariant measurement error and time effects

Given the panel data structure of the CompNet database, the researcher has the ability to employ panel data techniques to mitigate the effect of measurement error through regression analyses. That is, the

part of the measurement error that is constant over time, and specific to either countries, industries or size classes, can be controlled for by modelling it through "fixed" (or random) effects. Considering that differences in the data characteristics across countries are probably often persistent over time, the assumption of time invariance does not seem implausible. In addition, the researcher can control for time effects through the inclusion of year dummies. Thus, a regression specification may look like

$$y_{isct} = \beta x_{isct} + \sum\nolimits_{isct} \theta_{isct} d_{isct} + \varepsilon_{isct}$$
$$= \beta x_{isct} + \sum\nolimits_{i} \theta_i d_i + \sum\nolimits_{s} \theta_s d_s + \sum\nolimits_{c} \theta_c d_c + \sum\nolimits_{t} \theta_t d_t + \varepsilon_{isct}$$

where *i* indexes industries, *s* size classes, *c* countries and *t* years; *y* is the dependent variable, *x* the regressors, ε the equation disturbance; d_x denotes the four sets of dummies corresponding to dimension $x = \{i,s,c,t\}$, and θ_x the coefficient dummies. In case of breaks in the data for specific countries, dummies for countries and specific time spans, appropriate dummies can be added.

Using metadata information in the empirical specification to enhance comparability

Modelling the fixed effects in panel data estimation is a broadly accepted way to take into account possible biases and measurement errors. However, given the metadata endeavour of the CompNet, there is an opportunity to do more. In fact, this allows the researcher to look a bit deeper into the issues that have plagued the international comparability of firm-level data research, such as those described in this report, and mitigate their effects.

From the metadata survey, we know specific characteristics about the data, by variable, source and country. Using this information, countries that are similar concerning a specific issue can be grouped. In many cases, this information can then also be translated into binary or categorical variables. Exploiting the cross-country variance in these variables, they can then be added to the empirical specification, to control for biases stemming from specific differences in the metadata characteristics. As an example, suppose that for the variables in an analysis, some countries have sample data and some have census data, and in addition some have used imputations and others do not. Then the specification given above could be expanded including binary variables indicating the type of source, as well as the use of imputation. In general, denoting the indicators based on the metadata variables as *m*, the above specification can be expanded to

$$y_{isct} = \beta x_{isct} + \sum\nolimits_{isct} \theta_{isct} d_{isct} + \sum\nolimits_{j} \rho_j m_j + \varepsilon_{isct}$$

In addition, we note that not only does this approach provide a way to control for any reduced comparability issues, it also provides a way to roughly quantify the average extent of the effect on comparability. That is, the significance and magnitude of the coefficients ρ_j on the additional control

variables can be interpreted as the effect of a specific issue on the comparability of the dependent variable across countries.

Finally, any known incomparabilities could also be used to the advantage of the researcher, that is, to test robustness. For example, if a certain correlation is consistent across countries with census or sample data, with outlier correction or not, or variations in the exact definitions of the variables used, then the correlation seems robust to these differences in the data. In situations of a break in the data collection or a shift in definitions within a data source, using the information from this report, the researcher might be able to abuse this known incomparability over time to investigate the sensitivity of the results to the regime change, by carrying out sub-period analyses. In these ways, and depending on the research question, variation in the nature of the data across countries and time might be exploited by researchers for robustness checks.

6. Conclusions

While aiming at full comparability is a process, the report finds that the CompNet dataset has strong fundamentals to rely on. First, the national data sources contain a significant share of sources based on definitions and statistical guidelines set by the EU, which require substantial harmonisation efforts. Second, most sources are fully or partly under the responsibility of national statistical institutes, which is a guarantee of top methodological standards. Third, when sources do not draw from census information (i.e. are based on surveys), for virtually all countries, they can be linked to the census population; this allows the possibility of *ex-post* reweighing to ensure that the database out of that specific source is adequately representative of the population.

These fundamentals result in a 6th CompNet vintage covering on average about 40% of the corresponding population of firms (drawn from Eurostat), which is high in comparison with other datasets. Admittedly, the firms' coverage of the CompNet dataset varies across countries, but this does not appear to represent a problem, since overall the country samples are representative. In particular, the CompNet dataset captures the various segments of the firm-size population well and in line with Eurostat, albeit with some exceptions. In addition, the cross-country report on 6th vintage CompNet data provides a comparison of CompNet output with other published data sources as a general validation of the indicators. Overall, risks of incomparability under the dimensions of coverage, representativeness and validation appear to be limited.

On a number of issues, work is ongoing. For instance, more investigation is needed on (i) the ability to use national business registries for decentralising the weighting procedure (now 7 countries indicated that business registers are available for this transition), (ii) the impact different units of observation (e.g.

enterprises versus legal entities) have, in among others, Germany, Poland and Spain and (iii) different procedures and solutions being used at the country level to follow business dynamics (e.g. entry, exit, mergers), on comparability and what the common denominator is here.

Going forward, the report sets up an ambitious manageable agenda to further improve the comparability of the dataset over time. Besides the above-mentioned issues requiring more analysis, some of these improvements can already be achieved — at least partly — along two dimensions:

(1) By simply revising the common code. In this context, plans are already in place to (i) implement correction routines for cross-country variation in variables such as value added, employment and labour taxation, as well as (ii) adjust the common code to better match the cleaning and weighting procedures at the network level, with the ones already adopted at the country level and (iii) start pilot runs of the common code to deepen the common denominator.

(2) At the network level by providing precise instructions and definitions to all current and future data providers to have clarity on the content of the variables and to share best practices among data providers better.

The priority should be now to implement the recommendations in this report for the next vintage, thus solidifying the accuracy of the CompNet micro-aggregated dataset.

7. References

Ark, B. and Jäger, K., 2017. *Recent Trends in Europe's Output and Productivity Growth Performance at the Sector Level, 2002-2015*. International Productivity Monitor, (33), pp.8-23.

Ali, A., Klasa, S., & Yeung, E. (2014). *Industry concentration and corporate disclosure policy*. Journal of Accounting and Economics, 58(2-3), 240-264.

Altomonte, C. and Aquilante, T., 2012. *The EU-EFIGE/Bruegel-unicredit dataset* (No. 2012/13). Bruegel working paper.

Altomonte, C., Barba Navaretti, G., Di Mauro, F. and Ottaviano, G., 2011. *Assessing competitiveness: how firm-level data can help* (No. 2011/16). Bruegel Policy Contribution.

Airaksinen, A., Berezowska, J., Djahangriri, N., Edelhofer, E., Redecker, M., Zupan, G. (2013) *Final Report of the ESSnet on Linking of Microdata to Analyse ICT Impact.*

Awano, G., Bartelsman, E., Hagsten, E., Kotnik, P., Polder, M. (2012). *ESSnet on Linking of Microdata on ICT Usage.* Eurostat Grant Agreement 50701.2010.001-2010.578.

Bartelsman, E. and M. Doms (2000), *Understanding productivity: Lessons from*

longitudinal microdata, Journal of Economic Literature (38): 569-594.

Bartelsman, E., Haltiwanger, J. and Scarpetta, S., 2004. *Microeconomic evidence of creative destruction in industrial and developing countries.*

Bartelsman, E. J., Hagsten, E., & Polder, M. (forthcoming). *Micro moments database for cross-country analysis of ICT, innovation, and economic outcomes.* Journal of Economics and Management Strategy.

Béguin, J., and Hecquet, V. *An economic definition of enterprises for a clearer vision of France's economic fabric*, published in an Insee collection "Enterprises in France, Edition 2015"

Berlingieri, G., Blanchenay, P., Calligaris, S. and Criscuolo, C., 2017. *The Multiprod project: A comprehensive overview.* OECD Science, Technology and Industry Working Papers, 2017(4), p.1.

Caves, R. (1998), *Industrial Organization and New Findings on the Turnover and*

Mobility of Firms, Journal of Economic Literature (36): 1947-1982.

Dai, R., (2012). *International accounting databases on wrds: Comparative analysis.*

Ferrando, A., and Ruggieri, A., *Financial Constraints and Productivity: Evidence from Euro Area Companies* (July 9, 2015). ECB Working Paper No. 1823. Available at SSRN: https://ssrn.com/abstract=2628770

Geurts, K. and Van Biesebroeck, J., 2016. *Firm creation and post-entry dynamics of de novo entrants*. International Journal of Industrial Organization, 49, pp.59-104.

Hagsten, E., Iancu, D., Kotnik, P. (2013) *Quality of Linked Firm-Level and Micro-Aggregated Datasets:*

The Example of the ESSLait Micro Moments Database

Kalemli-Ozcan, S., Sorensen, B., Villegas-Sanchez, C., Volosovych, V., & Yesiltas, S. (2015). *How to construct nationally representative firm level data from the ORBIS global database* (No. w21558). National Bureau of Economic Research.

Lopez-Garcia, P. (2016). *The use of firm-based information for an enhanced assessment of competitiveness: Pros and cons of AMADEUS and COMPNET.* Internal note, European Central Bank.

Lopez-Garcia, P. (2018) *The new cross-country database of CompNet: novelties and main stylised facts,* European Central Bank

Mayer, T. and Ottaviano, G.I., 2008. *The happy few: The internationalisation of european firms. Intereconomics*, 43(3), pp.135-148.

Ribeiro, S. P., Menghinello, S., & De Backer, K. (2010). *The OECD ORBIS database: Responding to the need for firm-level micro-data in the OECD*. OECD Statistics Working Papers, 2010(1), 1.

8. Annex

8.1 Commercially available databases

The ORBIS database (AMADEUS in European context), compiled by the Bureau van Dijk Electronic Publishing, is a commercially constructed database that contains administrative data on 130 million firms worldwide. ORBIS provides financial and balance sheet data stemming from business registers collected by local Chambers of Commerce. The coverage of ORBIS varies by country due to differences in filing requirements for the business registries of the local Chambers of Commerce and due to confidentiality-related issues.

CompuStat is another commercially available database. It is constructed by S&P Global Market Intelligence. This database focuses on firms listed in stock exchanges. By definition, this database therefore captures the right-sided tail of the size distribution, and refrains from covering the smaller firms. Its country coverage is highest for North America.

These databases, as well as others, try to capture the granularity needed for adequate research but leave room for improvements in the field of coverage across countries, across the size distribution and the included indicators.

8.2 Outlier procedures CompNet Code

The CompNet code applies two routines that affect the raw variables before being fed into the actual indicator computation. The first routine loops through the main raw variables eliminating impossible values. The second routine focusses on assessing implausible values along a few criteria, and deletes them if the criteria do not hold. We discuss each routine in more depth.

8.2.1 Impossible values

The ratio behind the impossible values routine is the focus on preserving as much useful information as possible. Hence, small violations in certain accounting identities are not judged and treated as being data inconsistencies. We could test whether the difference between turnover and intermediate inputs, which should be equal to value added, holds in our datasets. However, we observe small violations of this identity. This can be explained by the plurality of data providers and heterogeneous underlying data sources. Instead of applying invasive accounting routines, we rely on our outlier treatment to filter out mismeasured values.

Therefore, the first routine investigates the raw variables provided by the national counterparts on the basis of accounting identities. The content of the following variables are deleted if they show negative values: turnover, capital, labour, totals assets, cash holdings, long-term debt holdings, trade credit, trade debt, interest payments, other fixed assets, current assets, dividend payments and depreciations. The

single observation is therefore treated as missing by the code. In a similar fashion, the interest payments and debts are checked. If the debt value is smaller than interest payments, both observations are turned to missing values.

8.2.2 Outlier dropping

The second routine focusses on mismeasured values and identifies them as outliers. Previous vintages of data collection taught us a trade-off. The outlier procedure must not affect or distort aggregate results by limiting the number of observations used for the indicator calculation. But still, it must be strict enough to correctly filter out values that can be identified as outliers. A factor that further complicates the creation of this code is that the routine is written without fine-tuning it to an individual data source.

Before the routine starts, the data is split into bins according to the two-digit sector and year. Within these bins consequently, three checks are applied.

1. Is a value more than three standard deviations away from the median?
2. Is a value in the top or bottom 1 percentile?
3. Is the growth of a value with respect to the previous year in the top or bottom 1 percentile?

If all of these conditions are fulfilled, the value is set to missing. Literature labels this as a *lenient routine*. Given the quality of the data sources and the institutions behind them, this lenient routine can be justified. The outlier procedure is assessed after each round of data collection and will possibly be strengthened in future vintages.

8.3 Confidentiality procedure

Although the literature has long recognised that firm-level data delivers crucial information about a wide range of phenomena, economic research based on these data has been so far hampered by issues of confidentiality and comparability. As a result, the CompNet data collection and indicator construction process has been designed in such a way that both issues are resolved. We describe the CompNet confidentiality procedure in two parts, one part focussing on the raw firm-level data and another part covering the eventual output of the code, the output indicators. Both parts contribute to the fact that the user of the final data will not be able to uniquely identify individual firms based on the aggregated data.

8.3.1 Raw variables

The conditions of dealing with firm-level information and the obligations surrounding confidentiality differ across countries and across member institutions. Given the large heterogeneous amount of data providers in the CompNet project, the process of raw variable compiling is decentralised. The CompNet secretariat and the individual data providers work together intensively in compiling the dataset, but the code is run in a decentralised way in each of the respective institutions. This means that no individual

firm-level data is made available to the secretariat of CompNet. In this way, each member institution can satisfy their individual confidentiality constraints.

8.3.2 **Output Indicators**

The second aspect of the confidentiality procedure is ensuring that the eventual output indicators leave no room for identifying individual firms. Also in this regard, each member institution can have individually specified conditions to satisfy. In the CompNet code is included a specific routine, which is run in the final stage of the computation, that checks the eventual output cells. This routine includes thresholds for the minimum number of observations to guarantee that no individual firm can be identified and tests for statistical dominance. If a cell is based on a limited amount of underlying micro-observations, making the identification of individual firms possible, the cell is dropped. This information is not eliminated from the total distribution; it is only left out of the specific cell. The second test is the test for statistical dominance. It includes thresholds for the largest permissible size share a single observation takes on in a given cell.

These thresholds can be set *a priori* by the data providers to satisfy their country- or institution-specific conditions. The following parameters can be chosen:

1. The minimal number of observations for the 1% and 99% percentiles can be adjusted.
2. The minimal number of observations for the 5% and 95% percentiles can be adjusted.
3. The parameter for statistical dominance can be adjusted. This is the largest permissible share an observation takes on in a cell

8.4 Tables

Table 13: Reference period across raw variables and countries

	1. Capital (Tangible fixed assets)	2. Raw materials (intermediate inputs)	3. Labour cost	4. Value added	5. Number of employees	6. Turnover	7. Unadjusted export value	8. Threshold adjusted export value	9. Import value	10. Total assets	11. Cash and cash equivalents	12. Cash flow (from profit/loss statement)	13. Profit/loss	14. Interest paid (or financial charges)	15. Long-term debt	16. Short-term debt	17. Total inventories	18. Depreciation	19. Trade credit (accounts payable)	20. Trade debt (accounts receivable)	21. Current liabilities	22. Non-current liabilities	23. Shareholder funds (equity)	24. Profits and losses before taxes	25. Other current assets	26. Other non-current liabilities	27. Other fixed assets	28. Intangible fixed assets	29. Current assets	30. Other current liabilities	31. Total fixed assets	32. Dividends
Finland	p	p	p	p	a	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p		p		p		
Czech Republic	p	t	t	t	a	t	t	t	t	p	t	t	t	t	p	p	p	t	p	p			p	t			p	p			p	t
Hungary	p	t	t	t	t	t	t		t	p	p	t	t	t	p	p	p	t	p	p	p	p	p	p	p	p	p	p	p	p	p	p
Lithuania	p	t	t	t	p	t	t		t	p	p	t	t	t	p	p	p	t	p	p	p	p	p	t	p	p	p	p	p	p	p	t
Netherlands	p	t	t	t	a	t				p	p	p	p	p	p	p	p	p	p	p	p	p	p	t	p	p	p	p	p	p	p	t
Portugal	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p
Slovakia	p	t	t	t	a	t	t		t	p		t	t	t		p	p	p	p	p			p	t				p	p		p	t
Spain	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p
Romania	p	t	t	t	a	t	t		t	p	p	t	t	t	p	p	p	t	p	p		p	p	t	p	p	p	p	p		p	
Sweden	p	t	t	t	t	t		t	t	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p
Denmark	p	t	t	t	t	t	t		t	p	p	t	t	t	p	p	p	t	p	p	p	p	p	t	p	p	p	p	p		p	t
Slovenia	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p

Note: p= point-in-time; a = annual average; t = annual total

Table 24: Non-financial sectors covered in CompNet

NACE rev.2 section	Description	Description
C	Manufacturing	Manufacture of food products
		Manufacture of beverages
		Manufacture of tobacco products
		Manufacture of textiles
		Manufacture of wearing apparel
		Manufacture of leather and related products
		Manufacture of wood and products of wood and cork, except furniture
		Manufacture of paper and paper products
		Printing and reproduction of recorded media
		Manufacture of chemicals and chemical products
		Manufacture of basic pharmaceutical products and pharmaceutical preparations
		Manufacture of rubber and plastic products
		Manufacture of other non-metallic mineral products
		Manufacture of basic metals
		Manufacture of fabricated metal products, except machinery and equipment
		Manufacture of computer, electronic and optical products
		Manufacture of electrical equipment
		Manufacture of machinery and equipment
		Manufacture of motor vehicles, trailers and semitrailers
		Manufacture of other transport equipment
		Manufacture of furniture
		Other manufacturing
		Repair and installation of machinery and equipment
F	Construction	Construction of buildings
		Civil engineering
		Specialised construction activities
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
		Wholesale trade, except of motor vehicles and motorcycles
		Retail trade, except of motor vehicles and motorcycles
H	Transportation and storage	Land transport and transport via pipelines
		Water transport
		Air transport
		Warehousing and support activities for transportation

Table continued

		Postal and courier activities
I	Accommodation and food service activities	Accommodation
		Food and beverage service activities
J	Information and communication	Publishing activities
		Motion picture, video and television programme production, sound recording and music publishing
		Programming and broadcasting activities
		Telecommunications
		Computer programming, consultancy and related activities
		Information service activities
L	Real estate activities	Real estate activities
M	Professional scientific and technical activities	Legal and accounting activities
		Activities of head offices; management consultancy activities
		Architectural and engineering activities; technical testing and analysis
		Scientific research and development
		Advertising and market research
		Other professional, scientific and technical activities
		Veterinary activities
N	Administrative and support service activities	Rental and leasing activities
		Employment activities
		Travel agency, tour operator and other reservation service and related activities
		Security and investigation activities
		Services to buildings and landscape activities
		Office administrative, office support and other business support activities

8.5 Country-specific information

8.5.1 Denmark

Table 25: Country-specific information: Denmark

Source	Subject	Information
Acc. Stat.	Breaks	Sole proprietorships are excluded in the CompNet sample.

8.5.2 Finland

Table 26: Country-specific information: Finland

Source	Subject	Information
SBS	Breaks	2005–2006, change in taxation records data. This change had an effect on the variable content of the data. Only aggregate-level variables are available, no breakdowns anymore.

Table continued

SBS	Breaks	2012–2013, renewal of the business statistics causes a break in the data. Main changes relate to Standard Industrial Classification, harmonisation of turnover and personnel data and deduction rules.
SBS	Quality issue	Ability to follow business dynamics such as M&As and firm entry and exit is not captured in the data.
SBS	NACE transformation.	Classification keys and other derivation methods used to transform rev.2 pre-2009
SBS	Variable deviation	Capital contains tangible and intangible fixed assets
SBS	Variable deviation	Raw materials are constructed subtracting value added from turnover
SBS	Variable deviation	Long- and short-term debts are defined as total debt. No difference between them

8.5.3 Czech Republic

Table 27: Country-specific information: Czech Republic

Source	Subject	Information
P501	Breaks	Threshold between full coverage/survey was 20 employees before 2008
TRADE	Breaks	Reporting threshold for intra-EU transactions was 4 million CZK per year before 2008, 8 million CZK after
P501	NACE transformation	Data back to 2005 were already assigned NACE 2 category by CZSO based on correspondence tables. Data for 2003–2004 were matched separately at 2-digit level
P501	Usage	Used by Czech Statistical Office to construct National accounts and for Structural Business Statistics
TRADE	usage	Used by Czech Statistical Office to construct foreign trade statistics
-	Currency conversion	Conversion based on annual average CZK/EUR exchange rate
P501	Variable deviation	VA and intermediate inputs do not account for cost of services
P501	Variable deviation	Cash and cash equivalents only contains actual cash, no bank deposits

8.5.4 Hungary

Table 28: Country-specific information: Hungary

Source	Subject	Information
NAV	Breaks	Jump in number of small firms is due to a change in accounting system (many firms had to create double-entry bookkeeping for tax returns. Birth year information is derived externally from Business register.
NAV	Quality issue	Variables not needed on tax reports can be misreported, for instance number of employees, tangible assets. Reporting of employee numbers is significantly lower than mandatory information
NAV	NACE transformation	There are four classification variables in our dataset handling NACE classification from 1992 to 1997, 1997 to 2003, 2003 to 2008 and 2008 onward. Consistency in NACE classification made in a special do-file based on methodology of Hungarian NSI.
	Currency Conversion	Conversion based on annual average HUF/EUR exchange rate.

8.5.5 Italy

Table 29: Country-specific information: Italy

Source	Subject	Information
ASIA	Breaks	In 2008, the methodology to estimate unpaid workers changed.
SIA	NACE transformation	Correspondence table of enterprises with NACE in 2007 and 2008 at a micro level. For large companies, a routine based on main export product is applied.

8.5.6 **Lithuania**

Table 30: Country-specific information: Lithuania

Source	Subject	Information
CU	Breaks	Since 2004, Customs declarations (including import and export statistics) within EU are managed via EU's Intrastat system; there are reporting exemptions applicable to Lithuanian firms based on annual import and export values (different every year), while total values of imports and exports are not affected significantly (exemptions do not make more than 5% of country-level import/export values), foreign trade statistics for smaller Lithuanian firms might be underreported significantly.
F01	NACE Transformation	The assignment has been done by National Statistical Office of Lithuania.
-	Currency conversion	Pre-2015, conversion is based on annual average LITAS/EUR exchange rate.
F01	Variable deviation	Reporting thresholds are present since 2004; they are different every year; no adjustments are made to uncover the true export value by the firm.

8.5.7 **Netherlands**

Table 31: Country-specific information: Netherlands

Source	Subject	Information
SFO	Breaks	The definition of small firms changes over time: 2000 < 12,5 million balance sheet total; 2001–2010 < 23 million; 2011–2014 = 40 million. This affects the source of the firm: large firms are surveyed; small firms are collected from tax register.
SFO	Breaks	Trade debt and credit are not available before 2005.
ABR	Breaks	2006: redesign of Business Register (redefinition of firm units); 2013: introduction of substantial number of new firms (not actual births).
ABR	Quality issue	The ability to follow business dynamics and/or determine firm age is limited because of M&A and changes in business ids for statistical purposes.
ABR	Quality issue	Employment is rounded to integers in most years; up to 2005 the employment data are less reliable.
SFO	NACE transformation	Three-tier approach: 1. One-to-one correspondence NACE rev 2 to NACE rev 1; 2. use 2009 info in earlier years; 3. derivation algorithm of statistical department responsible for financial data.
BR	Variable deviation	Employment in BR concerns persons employed and refers to only those persons on the own payroll (no secondments or agency workers). It is in fte and includes owners, family members only if they are on the payroll. It is noted that the employment characteristic is a proxy used to derive the size-class classification, and not the actual fte (which is available only for units of economic activity and on a sample basis).
SFO	Variable deviation	Turnover at purchaser prices (includes non-VAT taxes/subsidies).
SFO	Variable deviation	Firm birth: not actual birth but registration dates, no correction for M&A or other reasons for "fake" birth and death.

8.5.8 **Poland**

Table 32: Country-specific information: Poland

Source	Subject	Information
F01	NACE transformation	The dataset includes NACE rev 2 since 2005. The 2005 to 2008 codes have been provided alongside NACE rev 1 codes.

8.5.9 **Portugal**

Table 33: Country-specific information: Portugal

Source	Subject	Information
CBSD	Breaks	From 2000 to 2005, data were collected through the Central Balance Sheet Database (CBSD) annual survey. CBSD annual survey covered about 17,500 corporations each year. The statistics compiled from this annual survey included about 15,000 non-financial corporations (NFC) and, for that period, there is a bias towards large companies. It is important to note that data from 2000 to 2005 presents an array of characteristics suggesting that an analysis of the results should be undertaken with caution. This is mostly because the indicators must not be interpreted as corresponding to the overall results for the Portuguese NFC. For instance, the overall results of the corporations reflect, in terms of activity, a better coverage of large corporations as well as of corporations in the “manufacturing,” “electricity, gas and water” and “transport and communication” sectors. On the other hand, there is less coverage in the “trade and repairs” sector. This conclusion is reached

Table continued

		by comparing the proportion of turnover in the corporations in the sample with that of the total population of NFC.
IES	Breaks	It comprises the information on the annual accounts of corporations (cover all NFC, as well as other institutional sectors). Regarding NFC, IES comprises all resident enterprises with a commercial, industrial or agricultural nature as principal activity as well as non-resident entities with a permanent establishment in Portugal.
CBSD	Checks	Internal quality controls are applied to individual data in order to assure the internal consistency of information by companies and the consistency with other internal sources.
IES	Checks	Internal quality controls are applied to individual data in order to assure the internal consistency of information by companies and the consistency with other internal sources.

8.5.10 **Slovakia**

Table 33: Country-specific information: Slovakia

Source	Subject	Information
Reports	Breaks	Euro adoption in Slovakia (original data for the period before 2009 in SKK-Slovak koruna)
Register	Breaks	Euro adoption in Slovakia (original data for the period before 2009 in SKK-Slovak koruna)
Customs	Breaks	Intrastat thresholds' increases in 2007 and 2009. Euro adoption in Slovakia (original data for the period before 2009 in SKK-Slovak koruna).
Reports	NACE transformation	Two-tier approach: 1. One-to-one correspondence NACE rev 2 and NACE rev 1; 2. use 2009 info in earlier years
-	Currency conversion	Pre-2009, conversion is based on annual average SKK/EUR exchange rate.
Reports	Variable deviation	Profit/loss: net income (profit/loss − income tax)
Reports	Variable deviation	Cash flow: net income + depreciation
Reports	Variable deviation	Short-term debt: bank loans
Reports	Variable deviation	Current assets: total assets − total fixed assets

8.5.11 **Spain**

Table 34: Country-specific information: Spain

Source	Subject	Information
CBA	Breaks	Sample includes firms with turnover but not employment. These firms were excluded by CBSO in the previous period and will be included in future deliveries
CBA	Breaks	With respect to NACE, there are no breaks in the series due to the NACE rev2: firms' NACE has been well connected in the historical series
CBA	Quality	Firms that do not satisfy comprehensive internal consistency, or consistency checks with external sources, are excluded
CBA	Quality	The strict application of the recommendation of the Task Force on Head Offices, Holding companies and SPEs had meant that certain holdings, without autonomy of decision, developing these activities (under 642 NACE), have been included in NFC sector.

8.5.12 **Romania**

Table 35: Country-specific information: Romania

Source	Subject	Information
Trade	Breaks	Implementation of Intrastat methodology since 2007, resulting in a lower number of exporters and importers, changing of Intrastat reporting thresholds for imports in 2013, to ron 500 000 (approx. €113 000), from ron 300 000 in 2012 (approx. €67 000)
Bal. sheet info	Quality	The ability to follow business dynamics and/or firm age is limited because no care has been taken to take into account M&A and changes in business ids for statistical purposes
Bal. sheet info	NACE transformation	Three-tier approach: 1. Use post-2008 info on previous years; 2. use one-to-one correspondence between NACE 1 and NACE 2; 3. in case of one-to-many correspondence between NACE 1 and NACE 2, choose NACE 2 sector with largest employment in 2009.
Bal. sheet	Variable deviation	Intermediate inputs: raw materials and consumables expenses + other material costs + utilities expenses + goods for resale expenses

Table continued

Bal. sheet	Variable deviation	Cash flow: gross profit (= total revenue − total expenses) + depreciation
Bal. sheet	Variable deviation	Profit/loss: operating profit = operating revenue − operating expenses
Bal. sheet	Variable deviation	Depreciation: Depreciation on tangible and intangible assets
Bal. sheet	Variable deviation	Trade debt: Trade debt (Accounts payable) = Debt related to purchased goods and services
Bal. Sheet	Variable deviation	Non-Current liabilities: Long-term debt + Other non-current liabilities = Long-term debt + Provisions
Bal. Sheet	Variable deviation	Profit losses before taxes: Gross profit = Total revenue − Total expenses
Bal. Sheet	Variable deviation	Other current assets: Current assets − Trade credit − Total inventories
Bal. Sheet	Variable deviation	Other fixed assets: Financial assets
Bal. Sheet	Variable deviation	Birth rates: Registration date; no dealing with M&As
	Currency conversion	Conversion based on annual average RON/EUR exchange rate.

8.5.13 **Croatia**

Table 36: Country-specific information: Croatia

Source	Subject	Information
FINA	Breaks	Not relevant, since the data provider mapped data series over structural breaks into a consistent time series. (The accounting law changed 4 times during the CompNet collection period, creating subsamples for years 2002–2007, 2008–2009, 2010–2015 and 2016+)
FINA	Quality	It is a standard and widely used firm-level source, e.g. Bureau van Dijk uses it for their database compilations, as well as domestically for various statistical reports
FINA	Quality	FINA compiles the raw firms' balance sheets and P&L on the “as-is” base, leading sometimes to poor data quality. Quality and consistency checks are done on the data.
FINA	NACE transformation	The reclassification done by the National Statistical Office of Croatia
	Currency conversion	Midpoint exchange rate at Dec. 31

8.5.14 **Sweden**

Table 37: Country-specific information: Sweden

Source	Subject	Information
ITG	Breaks	The threshold in the Intrastat survey has been changed a few times during the period. 2003: Arrivals and Dispatches: 1.5 million SEK, 2005: Arrivals 4.5 and dispatches 2.2 million SEK, 2009: Dispatches 4.5 million SEK, 2015: Arrivals 9.0 million SEK
	Currency Conversion	Conversion based on annual average SEK/EUR exchange rate.