CompNet The Competitiveness Research Network

User Guide for the 7th Vintage of the CompNet Dataset

This Version: 13.07.2020

This User Guide has been updated by Marco Christophori and includes contributions by Matthias Mertens, Peter Haug, Tommaso Bighelli, Jonathan Deist, Johannes Amlung, Roman Blyzniuk, Marcel Lang (all Halle Institute for Economic Research) Tibor Lalinský, Sergio Inferrera and Ottavia Papagalli (all European Central Bank).

Table of Content

List of Tables

List of Figures

How to Become a User

To receive access, it is necessary to fill an online request form in the data section on the CompNet-homepage[1]. The Executive Committee of CompNet will review the request and inform you about their decision. The processing time can be reduced if you provide sound information about yourself and your research project. Additional information (such as a CV) can be very helpful. The applicant will normally be informed about the decision within two weeks. Please note the terms and conditions and other important regulations regarding the usage of the data, which are described in detail on the application page.

In case of acceptance, you will receive an email with the necessary credentials to log into the system and full access to the 4th (up to 2012), 5th (up to 2013), 6th (up to 2016) and 7th (up to 2017) data collection rounds for a period of six months. The renewal is possible at the end of this period. The user will be contacted two weeks before the termination regarding a potential renewal. Questions related to technically accessing the data can be directed to *fdz@iwh-halle.de*.

[1] https://www.comp-net.org/data/

1. Information Included in the User Guide

This user guide provides users of the 7th Vintage of the CompNet dataset with all necessary information to have an easy start with the dataset. The user guide represents the go-to guide for all dataset related questions.

Chapter 2 gives a detailed overview of the dataset and provides information on how to find the information of interest. It includes information on the available countries, time span, the target population and the naming convention of the data files and variables. Chapter 3 provides an overview of the caveats and possible limitation of this version of the CompNet dataset. Chapter 4 illustrates the differences between the 6th and the 7th Vintage[2]. The above chapters are augmented by an extensive appendix which provides detailed information, lists and tables on:

- List of variables
- Derivation of (semi-) parametric indicators[3]
- Country specific information
- Covered macro-sectors and two-digit NACE Rev. 2 sectors
- CompNet Data Collection
  - Harmonization of input data/data preparation
  - Confidentiality
  - Outlier routine
  - Validation
- Data provider data sources
- Instructions: How to become a user

The user guide is complemented by the "2020 CompNet Firm Productivity Report". This report focuses on the novelties of the 7th Vintage of the CompNet dataset and highlights them via series of example analyses.

When using the data, please cite as:

CompNet 2020, User Guide for the 7th Vintage CompNet Dataset

[2] Available in the complete version of this User Guide

[3] Indicators and variables are in many occasions used interchangeably, especially with respect to naming convention we do not distinguish between variables and indicators. However, in some specific cases indicators refer to more complex variables following certain assumptions or requiring more demanding calculations.

2. The 7th Vintage CompNet Dataset

This chapter introduces the reader to the technical information necessary to use the dataset. This does include knowledge about how the dataset is provided, about the applied naming conventions and information about the content of the different types of subsets included.

2.1 Sample, Time Range and Levels of Aggregation

The 7th Vintage of CompNet dataset is an unbalanced panel dataset which aims to cover 19 European countries.. The dataset includes a rich set of indicators from 6 different categories: productivity, finance, labour, competition, trade and others.

These variables are available for two samples: the "all" sample and the "20e" sample. The all sample includes all firms in the target population while the 20e sample includes only firms with 20 or more employees. The main reason for having two samples is that in some countries, firms are legally obliged to report their balance sheet data only when certain size thresholds are met.[4] As a result, besides the version of dataset covering all firms, CompNet decided to construct the separate version covering only firms with 20 and more employees. The 20e version of the dataset represents more homogeneous sample available for larger number of countries. The time spans covered by both versions are unbalanced, but most countries cover the period between 1999 and 2017.

Table 1 shows the samples and time spans available in the final version of the 7th Vintage of the CompNet dataset.

Table 1: Countries, Samples and Time Span

Country	Full Sample	20e	Time Span
Belgium	**X**	**X**	2003-2017
Croatia	**X**	**X**	2002-2017
Czech Republic	**X**	**X**	2005-2017
Denmark	**X**	**X**	2000-2016
Finland	**X**	**X**	1999-2017
France	**X**	**X**	2004-2016
Germany	**X** [a)]	**X**	2001-2017 [b)]

[4] These thresholds may vary across countries. For example, in Poland, only firms with more than 10 employees and in Slovakia firms with 20 employees report detailed accountings.

Hungary	**X**		2003-2017
Italy	**X**	**X**	2006-2016
Lithuania		**X**	2000-2017
Netherlands	**X**	**X**	2007-2017
Poland		**X**	2005-2017
Portugal	**X**	**X**	2004-2017
Romania	**X**	**X**	2005-2016
Slovakia		**X**	2000-2017
Slovenia	**X**	**X**	2002-2017
Spain	**X**	**X**	2008-2017
Sweden	**X**	**X**	2003-2016
Switzerland	**X**	**X**	2009-2017

Notes: a) For the German manufacturing business only firms >= 20 employees included; b) 2001-2002 include only Macro sector 1, 2017 only service sectors data. Macro sectors 1, 4,6,7,8 and 9 available for the period 2003-2017 with the exception of Subsector 75 available only from 2009. Macro sector 2 only partially available: sector 41 (2003-2007), sector 42 missing, sector 43 (2003-2008). Macro sectors 3 and 5 are missing.

Target Population:

The CompNet dataset covers non-financial corporations with at least 1 employee[5]. This definition is consistent with category S.11 in the European System of Accounts. It consists of institutional units which are independent legal entities and market producers, and whose principal activity is the production of goods and non-financial services (excluding sole proprietors). The non-financial corporation sector also includes non-financial quasi-corporations. The detailed information on sectors covered by the CompNet dataset is provided in Section 5.4.6 in the appendix.

[5] Please note that CompNet *input* includes all industries of the non-financial corporations (NFC) sector. The CompNet output however slightly deviates from this classification, excluding NACE rev.2 chapters B, D, E and division 95. Please refer to Table 19 in the appendix for a detailed overview of the NFC sectors included in CompNet.

Levels of Aggregation (or Dimensions):

Indicators available in the CompNet dataset are aggregated to different levels, e.g. according to different sector definitions or firm sizes. The available levels of aggregation are Country, Macro-Sector[6], Macro-Sector-Size-Class, Sector[7] and NUTS 2[8].

A detailed definition of the macro-sector and sector level aggregation can be found in section 5.4.6 in the appendix. The size-class definitions follow the Eurostat classification system:

Table 2 Size-Class Definitions

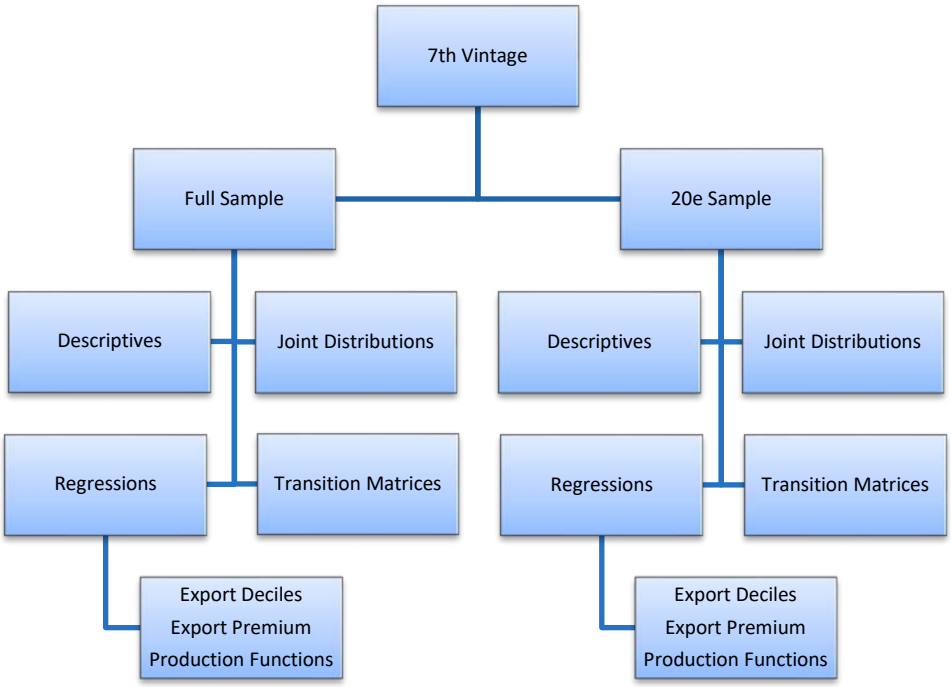
Size Class 1	Size Class 2	Size Class 3	Size Class 4	Size Class 5
1-9 empl.	10-19 empl.	20-49 empl.	50-250 empl.	>250 empl.

2.2 Structure of the 7th Vintage of the CompNet Dataset

The CompNet dataset consists of a relatively large number of data files saved in thematic folders. Each folder contains different datasets or regression output tables. Figure 1 shows the folder structure of the dataset.

[6] Corresponding to NACE Rev. 2 sections
[7] Corresponding to 2-digit NACE Rev. 2 sectors
[8] Corresponding to basic regions for the application of regional policies based on the Nomenclature of Territorial Units for Statistics (NUTS).

Figure 1: Structure of the Dataset



All files comprising the dataset are either Stata (.dta) or Excel files (.xls).[9] All files forming the dataset have unique names. The logic of the file naming convention, as well as the peculiarities of the subsets are described below.

The Naming Convention of the Data Files:

The data files are named according to the following basic structure:

Content_dimension_sample.dta

The three different labels represent different information regarding the data:

1. ***Content***: states the type of dataset and, if applicable, the main variable of interest
2. ***Dimension***: specifies the level of aggregation of the dataset
3. ***Sample***: indicates the weighted or unweighted sample on which the dataset is built:
    a. "all" includes all firms with at least 1 employee
    b. "20e" includes all firms with at least 20 employees

[9]The .dta files are compatible with Stata 13 or higher version and .xls files are compatible with MS Excel 2010 and higher. However, many statistical software packages are capable of importing and converting Stata files, e.g. in R by the R-package readstata13 or via RStudio.

Two Examples:

The dataset that contains general unconditional descriptive statistics for the weighted sample including all firms, at country level, can be found under the name:

unconditional_country_weighted_all_firms.dta

Accordingly, the joint distribution of the capital investment ratio (FR14[10]) at the macro-sector for the weighted 20e sample will be in:

jd_FR14_invest_k_country_weighted_20e_firms.dta

2.2.1 Descriptives

The descriptive section includes three different types of datasets, unconditional distributions (Unconditionals), productivity decompositions and misallocation measures (Decompositions), and productivity dispersions (Dispersions).

Unconditionals

The unconditional subsets contain the unconditional distributions of all productivity, finance, labour, trade, competition and other indicators available in CompNet's dataset and can be recognized by the prefix "*unconditional_dimension_weighted_sample"*. These subsets are available for the country, macro-sector, macro-sector-size-class, sector and nuts2 dimension both for the all and the 20e sample. For the first time, the 7th Vintage of the CompNet dataset also includes unweighted versions of two unconditional datasets. These datasets can be identified as follows: "*unconditional_dimension_**unweighted**_sample"*

Decompositions

The decompositions are datasets which contain the estimates of different allocative and dynamic efficiency measures. These can be recognized by the prefix "foster*_decomp" or "op_decomp"*. These stand for decompositions in line with Foster et al. (2006) or with Olley & Pakes (1996). Both decompositions are available for four different levels of aggregation: country, macro-sector, sector and nuts2. Among the indicators in this dataset, the user can find the decomposition of sector productivity into several components including the unweighted mean and covariance term between the variable of interest and the applied

[10] The identification code is explained in Section 2.3.

weight of economic activity (i.e. size). Further information on the computation of these indicators can be found in Section 5.3 of the appendix and the original articles.

Dispersions

The dispersion subsets are indicated through the prefix "*dispersion_"* and are estimated at the macro-sector and the sector level for both samples. They contain the within-sector dispersion of 46 productivity and competition variables, including labour productivity, capital productivity or De Loecker and Warzynski (2012) markups. The dispersion is calculated in the spirit of Kehrig (2011) to make dispersion within industries comparable among each other and over time. Differently from Kehrig, who is using the mean of the variance, CompNet uses the median of the standard deviation. This creates an estimate that is more robust to outliers. More information can be found in Section 5.3 of the appendix.

2.2.2 Joint Distributions

Joint distributions refer to conditional distributions, i.e. distributions of a large number of variables given a specific condition[11]. The condition can be either discrete or continuous. The dataset includes discrete values (0 and 1) of dummy variables or deciles of continuous variables. The list of available conditions varies across the CompNet dataset dimensions[12]. However, the list of variables summarised within the joint distributions is identical across the conditions and dimensions.

All conditional distributions of (summarized) variables available within a certain condition are stored in a unique data file. The name of the file follows the structure "jd_*varname_dimension*_weighted", where *varname* stands for the (condition) variable and *dimension* stands for the aggregation level. Similar to the descriptives, the 7th vintage also include unweighted joint distributions. Their names follow the structure "*jd_varname_dimension_**unweighted***".

An example of a discrete (condition) variable could be a dummy taking the value one for distressed firms and zero for non-distressed firms. The data file would then include all distributions of the (summarized) variables within a given dimension conditional upon the firm

[11] That is also the reason why the independent distributions of included variables in the descriptives are called *unconditionals.*

[12] See Section 5.1 in the appendix for an overview

being distressed, or not. If the split is done on the basis of a continuous variable, for example, labour productivity, the dataset would include the distribution of the indicators given the deciles of firms' productivity distribution within a given dimension.

The 7th Vintage of the CompNet dataset includes 253 joint distributions. Following the file name construction mentioned above, the available conditions can be read from the content part of the file names. The datasets of joint distributions are available at the country, macro-sector, macro-sector-size-class, sector or nuts2 dimensions for both, the all and the 20e sample. The full list of joint distributions available can be found in Section 5.1 in the appendix.

The following example is meant to illustrate the concept of a continuous joint distribution.

Example: *jd_LV21_l_country_weighted_20e.dta*

The dataset provides all indicators conditional on the continuous variable *l* (number of employees of the firms), which is indicated in the content part of the data file "*LV21_l*". Furthermore, this dataset uses the weighted 20e sample, meaning that all variables are based on the complete target population of firms with more than 20 employees.

Figure 2: Labour Productivity Distribution by Size Centiles in Poland

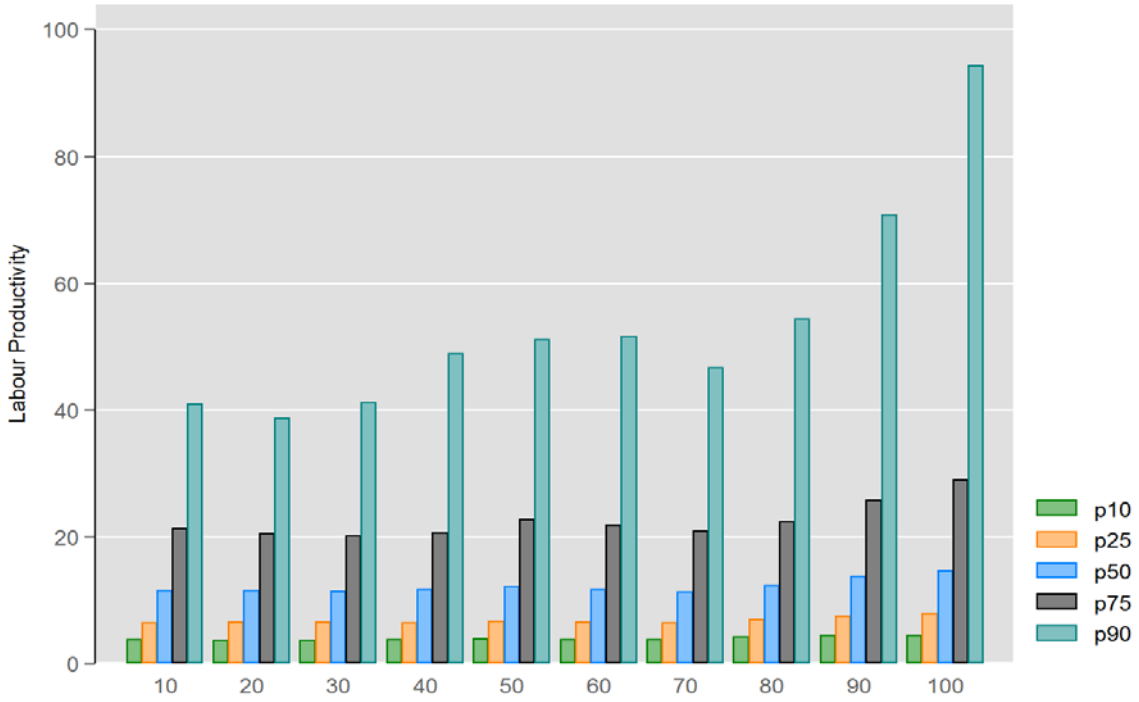


Figure 2 shows an example of a (continuous) joint distribution. In particular, it shows the labour productivity distribution of firms in different deciles of the firm size distribution in Poland in 2015. It becomes clear that there is a discontinuity in the productivity of firms by size. Up to the p70 centile of the size distribution, the productivity of firms does not change significantly with size. However, productivity increases continuously with size when considering firms in the upper 30 per cent of the size distribution.

2.2.3 Transition Matrices

Transition matrices aim to show the share and characteristics of firms with different growth performance over a three-year window. These datasets are indicated with the prefix "*transition_matrix*" and the currently available dimensions are country and sector.

To fulfil this aim, we restrict the sample to firms surviving over the three-year time period in question and define five size quintiles with this population. Firms are then categorized into transition categories according to their movement within these quintiles: E.g. a growing firm might be in size quintile 1 at the beginning of the three-year time period and be in quintile 3 at the end. We than gather summary statistics for firms in each category at the start and the end of the period.

The data file follows a similar structure as the joint distributions given that the condition used to split the sample is the indicator describing the transition of the firm along the size distribution. The condition is named "*TRmat_l_country"* in the transition matrix datasets. For example, for all firms that moved from the second labour quantile to the fifth the variable take the value "*from_q2_to_5*". Thereby researchers have the opportunity to investigate the share and characteristics of firms transitioning in different quintiles in terms of a number of employees.

The available transitions are: *from_q._to_., from_q._to_1, from_q._to_2, from_q._to_3, from_q._to_4, from_q._to_5, from_1._to_., from_q1_to_1, from_q1_to_2, …, from_q5_to_5.*

2.2.4 Regressions

In addition to the above-introduced datasets, each vintage of the CompNet dataset provides the user with a series of firm-level regression outputs, available as Excel files. For each country, the regression outputs cover results of production function estimations and for countries with trade data also regression results for export probability or export premium are available. Each of them is discussed in the following subsections.

Production Functions:

Production functions are estimated by pooling all firms operating in a given macro-sector or sector level, and by assuming either a Cobb-Douglas or translog production function. The output measure of the firm used in the regression is either real value-added or real revenue. At the same time several different combinations of input variables were considered. See Section 5.3.1 in the appendix for more details on the estimation procedure. Apart from the regression output files in Excel format, the coefficients and standard errors of the different specifications are also provided collectively in two dta files for each country.

Export-Probability:

The first set of export-related regression output files give the user the results of probit estimations of the probability to export on the productivity and firm size. There is a separate regression for every explanatory variable and country. They correspond to deciles of the different productivity and size variables. All regressions control for year dummies. Formally, the coefficients of the following model are estimated by maximum likelihood:

$$P(Dummy\_exp = 1|\mathbf{ct\_x}, \mathbf{szclass}, \mathbf{year}) = \Phi(\mathbf{ct\_x}\boldsymbol{\alpha} + \mathbf{szclass}\boldsymbol{\beta} + \mathbf{year}\boldsymbol{\gamma})$$

with **ct_x**[13] as the productivity centile dummies and **szclass** as the macro-sector-size-class dummies and $\Phi()$ as the standard normal cumulative distribution function. The observations are weighted by their inverse sampling probability defined as the theoretical number of firms within a certain macro-sector size class divided by the actual number of sample firms in that macro-sector size class. An overview of the productivity and size variables can be found in Section 5.1 in the appendix.

Export Premium:

The second bundle of export-related regression outputs deals with the question whether exporting firms are more productive than firms which are not exporting. Each file contains the result of three regressions with the following specifications of the dependent productivity variable: productivity (all firms), productivity change since the last period, and productivity (only non-exporters). The independent variables include the trade dummy variables sector, size class and crisis dummy as well as interaction terms. Just like in the probit estimation, the observations are weighted by their inverse sampling probability defined as the theoretical number of firms within a certain macro-sector-size-class divided by the actual number of sample firms in that size class. The regression model is then a simple pooled OLS-regression:

$$\mathbf{prod} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{SZ}\boldsymbol{\beta} \pm \mathbf{SEC}\boldsymbol{\gamma} + \mathbf{CR}\boldsymbol{\delta} + \mathbf{IT}\boldsymbol{\varepsilon} + \mathbf{e}$$

with **prod** as the vector of the chosen productivity measure (in logs), **X** as the matrix of various export dummies, **SZ** containing the size class dummies, the two-digit sector dummies **SEC, CR** for the crisis dummies and **IT** containing the interaction terms of **CR** and the export dummy,

[13] For technical reasons the reported productivity centile dummy names inside the excel files are still in line with the naming scheme of the 6th Vintage of the CompNet dataset. The excel file name does indicate the correct name.

and finally the error term **e**. The number of independent variables used depends on the model specification.

Table 3 shows the dependent and independent variables:

Table 3: Export Premium (Weighted OLS-Regressions)

Dependent Variables		Independent Variables
Identification Code	Variable Name	Variable Name
PV1	lnkprod_va	D_exp_3y
PV2	lnlprod_rev	D_exp_new
PV3	lnlprod_va	D_stop_exp
PV4	lnsr	D_exp_switch
PV5	lnsr_cs	D_exp
PE20	lntfp_rcd_in_ols_M	Sector
PE21	lntfp_rcd_in_ols_S	Size class
PE22	lntfp_rcd_ols_M	D_financial_crisis
PE23	lntfp_rcd_ols_S	
PE24	lntfp_rcd_wd_M	
PE25	lntfp_rcd_wd_S	
PE26	lntfp_rtl_ols_M	
PE27	lntfp_rtl_ols_S	
PE28	lntfp_rtl_vi_ols_M	
PE29	lntfp_rtl_vi_ols_S	
PE30	lntfp_rtl_vi_wd_M	
PE31	lntfp_rtl_vi_wd_S	
PE32	lntfp_rtl_wd_M	
PE33	lntfp_rtl_wd_S	
PE34	lntfp_vcd_ols_M	
PE35	lntfp_vcd_ols_S	
PE36	lntfp_vcd_wd_M	
PE37	lntfp_vcd_wd_S	

2.3 Naming Convention of Variables

The naming convention for variables in the 7th Vintage of the CompNet dataset has been completely restructured. All variables (or indicators) delivered in the CompNet dataset follow now a clearly defined naming convention. Each part of a variable name does contain coded information which is consistent across all variables. In addition, a code system was introduced

to clearly identify variables with a combination of 4 characters. This "identification code" can always be found at the beginning of each variable. These changes have been implemented to make the naming more consistent and to make the best use of the set limit of 32 characters for variable names in Stata. The data user will also find a short Stata code in the data section on the CompNet website that can be run to remove the CompNet identification code. This might be useful for individual analysis on a standard set of indicators.

7th Vintage Variable Naming Scheme:

Each name of every variables included in the 7th Vintage of the CompNet dataset follows the new naming scheme:

IdentificationCode_IndividualName_[Weightedby]_Suffix

The (4-digit) **identification code** is built with three elements: the thematic category, the variable type and numerical code:

CategoryTypeNumericalCode

Each of the elements can be directly identified and understood: the first letter represents the category; the second letter indicates the variable type and the last 2-digit combination represents the numbering. An overview of the potential content is provided in Table 4. The category corresponds to the already introduced categories defined as the topics covered by the CompNet dataset. The CompNet dataset includes a range of different types of variables. In this vintage, it is now possible to directly understand from the identification code whether a variable is e.g. an estimate or a ratio. The numerical code system is used as a sequential number system and applied within each category and variable group. For example, in the category "competition" the dataset includes 3 "ratio" type variables with the code CR00, CR01 and CR02. In the same category, the dataset also includes 10 "value" type variables represented by the identification codes PV00 up to PV09. In some cases, the dataset includes more than 99 variables for a given category and variable group. In such cases, the following sequence is applied: *{a0,a1,…,a9,b0,b1,…,b9,c0,c1…}*. Table 4 summaries the available details for the identification codes.

Table 4 CompNet Identification Code

Categories:	Variable Types:	Numerical Code System (read vertically)					
C – Competitiveness	D – Dummies	00	06	a0	a6	b2	b8
F – Financial	E – Estimates[14]	01	07	a1	a7	b3	b9
L – Labour	G – Growth Rates	02	08	a2	a8	b4	c0
P – Productivity	R – Ratios	03	09	a3	a9	b5	c1
T – Trade	V – Values[15]	04	10	a4	b0	b6	c2
O – Other		05	…	a5	b1	b7	…

Coming back to the variable naming scheme**,** **IndividualName** stands for the abbreviation combination of the actual name of the individual variable. For example, the abbreviation combination mu_l_rcd_wd_S stands for the mark-up variable (mu) calculated with labour as an intermediate input (l) used in a revenue-based Cobb-Douglas production function (rcd) estimated with an instrumental variable approach as in Wooldridge (2009) (wd) for the sector level (S). An overview including the identification code, the individual variable name and a definition of all variables can be found in Section 5.2 in the appendix.

Weightedby indicates the weighting method used in the derivation of the variables for the decompositions provided in this vintage[16]. It always starts with a capital W and ends with the abbreviation of the weighting method, e.g. Wrrv which stands for "weighted by real revenue". If no weighting was applied to a specific variable, this step will be skipped and the next component of the variable name, i.e. suffix will be reported directly. Table 5 reports the different weighting methods used in the 7th Vintage of the CompNet dataset for the decomposition variables[17] including the applied abbreviation.

[14]Estimates are defined as any variable which is based on a production function
[15]Defined as a number that represents an amount
[16] For a detailed explanation review Section 5.3.2 in the appendix
[17] This should not be confused with the general weighting procedure that is applied to all indicators. See 5.4.3 in the appendix.

Table 5 Weighting Methods Used for the Decompositions

Abbreviation	Meaning
rrv	Real revenue
ntc	Nominal total costs
nlc	Nominal labour cost
rv	Nominal revenue
l	Labour
rva	Real value-added
nm	Nominal intermediate inputs
nvi	Nominal variable inputs
nen	Nominal energy inputs
nva	Nominal value-added

The dataset delivers a rich set of information for every variable included in the dataset. This includes the mean, percentiles or the number of firms with available information to compute the relevant indicator. The suffix denotes which information of the variable the user is currently looking at. The set of information delivered by the dataset and described by the suffixes is shown in Table 6.

Table 6: Suffixes

Suffix	Meaning
p1, p5, p10, p25, p50, p75, p90, p95, p99	Percentiles of the considered variable
mn	Mean of the considered variable
sd	Standard deviation of the considered variable
skew	The skewness of the considered variables
kurt	Kurtosis of the considered Variable
N	Number of observations
sw	Summed weights of the considered variable
umn	The unweighted mean of the considered variable
usw	Unweighted summed weights of the considered var.
wmn	Weighted mean of the considered variable
cov	The covariant term of the considered variable
wth	Indicates the Foster within component
btw	Indicates the Foster between component
agg	Indicates the aggregate foster term
disp	Kehring dispersion
tot	Population total or absolute value for whole population of a variable.

A complete overview of all CompNet dataset variables including their identification code, individual variable name and suffixes can be found in Section 5.2 in the appendix.

3. Important Notes on Using the Dataset

This chapter highlights a few important features of the 7th Vintage of the CompNet dataset and provides recommendations on how to deal with them while using the data. It is highly recommended to carefully review this section before starting an analysis with the 7th Vintage. Topics range from technical intricacies, to how to correctly interpret and combine provided information. The purpose is to help the user avoid "technical" mistakes in using the dataset and enable her to conduct sound work with the data.

It is important to stress that the 7th Vintage of the CompNet dataset addressed a multitude of caveats existing in the 6th Vintage of the CompNet dataset to improve the accuracy and comparability of many variables. An overview of these improvements is included in Section 4. More information about the comparability of the 6th Vintage of the CompNet dataset can be found in the CompNet cross-country comparability report (2018).

3.1 Comparison with National Accounts

CompNet indicators are aggregated from firm-level sources where the information is based on national taxation legislation, European legislation and accounting principles (e.g. GAAP). These different sources are consolidated into the national accounts according to the current national accounts standards of the European System of Accounts (ESA). The national accounts aggregated data differ significantly from the CompNet variables, first and foremost because the data stems from a wide variety of sources which also cover firms outside of the CompNet dataset's target population[18]. Hence, the two datasets might show similar patterns, but are vastly different because they measure different slices of economic activity.

3.2 General Notes

3.2.1 Sample and Population Figures

It is important that the data user is aware that the applied weighting procedure gears the descriptive statistics of the CompNet indicators towards describing the total population, and

[18] For more details see Section 2.1

not the underlying sample [19]. Consequently, one frequent question is how to compute aggregate statistics with the variables provided in the weighted datasets in 7th Vintage of the CompNet dataset.

To compute aggregate statistics for the underlying population, one needs to make use of variables ending on "sw", standing for "summed_weights". To calculate the total employment for the *population* of a given cell, it is sufficient to multiply the average employment "l_mn" by the variable "l_sw" (i.e. the implied number of firms in the *population*). This line of reasoning applies to all other variables. Variables with the suffix "N" show the relative number of firms in the cell with available information to construct the given variable. Consequently, the "tot" variable helps the data user to find out more about the sample.

If the researcher wants to collapse the dataset to a higher level of aggregation (for example, from the sector to the macro-sector level), one needs to use the "sw" variable to have population-representative weights. This is an important difference with respect to CompNet vintages **preceding** the 6th Vintage which had implemented the reweighting procedure only for the 20e sample, but not for the full sample.

3.2.2 Dummy Variables

The CompNet dataset contains many dummy variables included in all categories. The CompNet code is built to store in each output file an array of the percentiles and moments of the distribution of each variable. However, this operation is not applied for dummies due to the binary distribution. For these variables, the only descriptive statistic in the output files that provides useful information is the mean, i.e. the percentage of observations (firms) for which the variable is equal to 1. For example, the mean of "TD14_exp" equals the share of exporters in the given cell and will, therefore, take values between 0 and 1.

3.3 Comparability

The data collection process discussed in Section 5.4 in the appendix has three main advantages:

[19] This only applies to the weighted datasets. The unweighted datasets deliver statistics on the sample. For more information on the weighting procedure see Section 5.4.3.

(i) The dataset uses existing national datasets, with no need to undertake new and costly data collection efforts

(ii) Confidentiality of the micro information is fully protected by using the micro-aggregation technique

(iii) Member institutions participate actively in improving and using the dataset.

By using existing national data sources, the data collection process is less cost-intensive but has a considerable downside: there is limited ability to affect source characteristics such as sampling techniques, variable definitions, industry coverage and others. These characteristics may sometimes vary considerably across countries due to differences in economic structure and legal systems, i.e. tax codes and administrative procedures, or due to the discretion of the statistical office. These cross-country differences might limit cross-country comparability.

To redeem this disadvantage, CompNet provides documentation of source data differences to help data users deciding on their relative importance, as well as suggesting strategies to mitigate the potential biases of own estimations based on CompNet data. For that reason, CompNet has invested to produce detailed meta-data and to analyse the strengths and weaknesses of the data in terms of cross-country comparability. This documenting effort sets the CompNet dataset apart from other sources of granular data. To find out more about comparability in the context of the CompNet dataset, a careful review of the CompNet cross-country comparability report (2018) is highly recommended.

The causes of comparability limitations are divided into the country- and source-specific comparability issues as well as variable and indicator specific incomparability. The following sections discuss these causes and provide some apparent examples.

3.3.1 Countries and Source Data

The country and source-specific causes of incomparability refer to the fundamentals of the different data sampling methodology in each country. Some exemplary questions here are at what level of aggregation the information is captured, what industries are covered, whether firms are representative of the population in terms of macro-sector and size classes and whether there are significant breaks or changes affecting the quality of the underlying source. Note, however, that data providers update the whole time series every time they run the code (therefore, not only one extra year is added) in order to minimize breaks in the dataset

resulting from the addition of new indicators, change in sector classification or improvement of the underlying methodology.

Units of Observations

In a dataset containing micro information, firm-data can be gathered at different levels of aggregation, the so-called units of observation. Eurostat uses the enterprise level of observation[20]. The enterprise-level is used by a selected number of data providers of CompNet as well, but the majority use the legal unit, which is a lower level of aggregation[21]. The usage of different levels of aggregation matters because different data sources across countries will target a different 'slice' of the economy. Consolidation of the balance sheets also plays a role here; unconsolidated information at the enterprise level could inflate economic activity relative to consolidated enterprise information.

Representativeness

On a more fundamental note, it is important to have representative data for all different countries. Enough firms should be covered by the domestic data sources and more importantly, these firms should be representatively distributed across different size classes and macro-sectors. Although the coverage rates differ between countries, the overall assessment of the sample representativeness is very positive[22]. For details, see Section 5.4.7.

3.3.2 Variables and Indicators

The variable- and indicator specific sources of incomparability refer to possible differences between raw variable definitions. The common code sent out to data providers calculates the output indicators from the underlying raw variables. Hence, differences between the definitions of the input may cause large differences in the output of the code. For this vintage,

[20] "An organizational unit producing goods or services which have a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and can do so at more than one location. An enterprise may consist of one or more legal units."
[21] See appendix Section 5.4.4
[22] For an assessment see the CompNet cross-country comparability report (2018)

a new harmonized set of definitions has been used by all data providers and included 1st, 2nd and 3rd best variable definitions. Section 5.4.5 in the appendix contains detailed overviews:

- Table 17 includes information on all raw variables and their possible definitions
- Table 18 highlights the used definitions for each country included in the dataset.

4. Differences 6th Vintage vs 7th Vintage of the CompNet Dataset

The 7th Vintage of the CompNet Dataset includes some innovations respect to the previous editions to improve the availability and the quality of the data and their user-friendliness.

The list of the most important innovations is as follows:

- New Source of Raw Data: in this vintage we use different data sources for Italy and Portugal (see Table 16);
- A new naming scheme (see Section 2.2): a help file is available to facilitate the conversion from the old to the new naming scheme;
- New variables:
    - Production function estimation with Intangible fixed assets as input;
    - Trade variables available also with the threshold adjust version;
    - Trade variables differentiate from extra-EU and intra-EU trade;
    - Sample and population version of Herfindahl-Hirschman Index (HHI);
    - Job destruction and job birth rates;
    - A large set of trade dummies
    - Growth rates coded as decimals (1% increase equals 0.01)

5. Appendix

5.1 List of Data Folders

The following overview presents the available data folders and number of files in the 7th Vintage of the CompNet Dataset.

Table 7 Data Files Overview

Sample	Folder	(subfolder)	Number of files
20e_firms	**Descriptives**		17
	JointDistributions		259
	Regressions	Export_deciles	285
		Export_premium	286
		Production_Functions	288
	Transmatrices		2
all_firms	**Descriptives**		17
	JointDistributions		259
	Regressions	Export_deciles	256
		Export_premium	257
		Production_Functions	270
	Transmatrices		2

5.2 Detailed Variable Overview

The definition of all output variables can be found in the following tables. Each table includes the variables for one category.

5.2.1 Competition Variables

Table 8 Competition Variables

Competition				
ID	Variable category	Numerical code	Variable Name	Definition
Estimates				
C	E	00	dm_rcd_ols_M	Indicator of firms' labour market power, based on the OLS estimation of revenue-based Cobb-Douglas production function at the mac-sector level
C	E	01	dm_rcd_ols_S	Indicator of firms' labour market power, based on the OLS estimation of revenue-based Cobb-Douglas production function at the sector level
C	E	02	dm_rcd_wd_M	Indicator of firm's labour market power, based on the Wooldridge estimation of revenue-based Cobb-Douglas production function at the mac-sector level
C	E	03	dm_rcd_wd_S	Indicator of firms' labour market power, based on the Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
C	E	04	dm_rtl_ols_M	Indicator of firms' labour market power, based on the OLS estimation of revenue-based translog production function at the mac-sector level
C	E	05	dm_rtl_ols_S	Indicator of firms' labour market power, based on the OLS estimation of revenue-based translog production function at the sector level
C	E	06	dm_rtl_wd_M	Indicator of firms' labour market power, based on the Wooldridge estimation of revenue-based translog production function at the macro-sector level
C	E	07	dm_rtl_wd_S	Indicator of firms' labour market power, based on the Wooldridge estimation of revenue-based translog production function at the sector level

Competition				
ID	Variable category	Numerical code	Variable Name	Definition
C	E	08	mu_l_rcd_ols_M	D.W.[23] markup-term given the firm labour input decision, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
C	E	09	mu_l_rcd_ols_S	D.W. markup-term given the firm labour input decision, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
C	E	10	mu_l_rcd_wd_M	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
C	E	11	mu_l_rcd_wd_S	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
C	E	12	mu_l_rtl_ols_M	D.W. markup-term given the firm labour input decision, derived from OLS estimation of revenue-based translog production function at the macro-sector level
C	E	13	mu_l_rtl_ols_S	D.W. markup-term given the firm labour input decision, derived from OLS estimation of revenue-based translog production function at the sector level
C	E	14	mu_l_rtl_wd_M	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
C	E	15	mu_l_rtl_wd_S	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of revenue-based translog production function at the sector level

[23] Following De Loecker and Warzynsky (2012), see Section 5.3.5 for details.

Competition				
ID	Variable category	Numerical code	Variable Name	Definition
C	E	16	mu_l_vcd_ols_M	D.W. markup-term given the firm labour input decision, derived from OLS estimation of value added-based Coob-Douglas production function at the macro-sector level
C	E	17	mu_l_vcd_ols_S	D.W. markup-term given the firm labour input decision, derived from OLS estimation of value added-based Coob-Douglas production function at the sector level
C	E	18	mu_l_vcd_wd_M	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of value added-based Coob-Douglas production function at the macro-sector level
C	E	19	mu_l_vcd_wd_S	D.W. markup-term given the firm labour input decision, derived from Wooldridge estimation of value added-based Coob-Douglas production function at the sector level
C	E	20	mu_m_rcd_ols_M	D.W. markup-term given the firm intermediate input decision, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
C	E	21	mu_m_rcd_ols_S	D.W. markup-term given the firm intermediate input decision, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level revenue-based Cobb-Douglas
C	E	22	mu_m_rcd_wd_M	D.W. markup-term given the firm intermediate input decision, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
C	E	23	mu_m_rcd_wd_S	D.W. markup-term given the firm intermediate input decision, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level l

Competition				
ID	Variable category	Numerical code	Variable Name	Definition
C	E	24	mu_m_rtl_ols_M	D.W. markup-term given the firm intermediate input decision, derived from OLS estimation of revenue-based translog production function at the macro-sector level
C	E	25	mu_m_rtl_ols_S	D.W. markup-term given the firm intermediate input decision, derived from OLS estimation of revenue-based translog production function at the sector level
C	E	26	mu_m_rtl_wd_M	D.W. markup-term given the firm intermediate input decision, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
C	E	27	mu_m_rtl_wd_S	D.W. markup-term given the firm intermediate input decision, derived from Wooldridge estimation of revenue-based translog production function at the sector level
C	E	28	mu_vi_rtl_vi_ols_M	D.W. markup-term given the firm intermediate and labour input decision, derived from OLS estimation of revenue-based translog production function at the macro-sector level
C	E	29	mu_vi_rtl_vi_ols_S	D.W. markup-term given the firm intermediate and labour input decision, derived from OLS estimation of revenue-based translog production function at the sector level
C	E	30	mu_vi_rtl_vi_wd_M	D.W. markup-term given the firm intermediate and labour input decision, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
C	E	31	mu_vi_rtl_vi_wd_S	D.W. markup-term given the firm intermediate and labour input decision, derived from Wooldridge estimation of revenue-based translog production function at the sector level
Ratios				

Competition				
ID	Variable category	Numerical code	Variable Name	Definition
C	R	00	top_rev_sam_C	Ratio of Top 10 firms' revenue to total revenue at the country level
C	R	01	top_rev_sam_M	Ratio of Top 10 firms' revenue to total revenue at the macro-sector level
C	R	02	top_rev_sam_S	Ratio of Top 10 firms' revenue to total revenue at the sector level
			Values	
C	V	00	hhi_rev_pop_C	Hirschman-Herfindahl index of market concentration at the country level based on the firm population
C	V	01	hhi_rev_pop_M	Hirschman-Herfindahl index of market concentration at the macro-sector level based on the firm population
C	V	02	hhi_rev_pop_N	Hirschman-Herfindahl index of market concentration on nuts2 level based on the firm population
C	V	03	hhi_rev_pop_S	Hirschman-Herfindahl index of market concentration at the sector level based on the firm population
C	V	04	hhi_rev_sam_C	Hirschman-Herfindahl index of market concentration at the country level based on the firm sample
C	V	05	hhi_rev_sam_M	Hirschman-Herfindahl index of market concentration at the macro-sector level based on the firm sample
C	V	06	hhi_rev_sam_N	Hirschman-Herfindahl index of market concentration at the nuts2 level based on the firm sample
C	V	07	hhi_rev_sam_S	Hirschman-Herfindahl index of market concentration at the sector level based on the firm sample

5.2.2 Productivity Variables

Table 9 Productivity Variables

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
Estimates				
P	E	00	elk_rtl_ols_M	Output elasticity of capital, derived from the OLS estimation of revenue-based translog production function at the macro-sector level
P	E	01	elk_rtl_ols_S	Output elasticity of capital, derived from the OLS estimation of revenue-based translog production function at the sector level
P	E	02	elk_rtl_vi_ols_M	Output elasticity of capital, derived from OLS estimation of revenue-based translog production function with variable inputs at the macro-sector level
P	E	03	elk_rtl_vi_ols_S	Output elasticity of capital, derived from OLS estimation of revenue-based translog production function with variable inputs at the sector level
P	E	04	elk_rtl_vi_wd_M	Output elasticity of capital derived from Wooldridge estimation of revenue-based translog production function with variable inputs at the macro-sector level
P	E	05	elk_rtl_vi_wd_S	Output elasticity of capital, derived from Wooldridge estimation of revenue-based translog production function with variable inputs level at the sector level
P	E	06	elk_rtl_wd_M	Output elasticity of capital, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	07	elk_rtl_wd_S	Output elasticity of capital, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	08	ell_rtl_ols_M	Output elasticity of labour, derived from OLS estimation the revenue-based translog production function at the macro-sector level
P	E	09	ell_rtl_ols_S	Output elasticity of labour, derived from OLS estimation of revenue-based translog production function at the sector level.

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	10	ell_rtl_wd_M	Output elasticity of labour, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	11	ell_rtl_wd_S	Output elasticity of labour, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	12	elm_rtl_ols_M	Output elasticity of intermediates, derived from OLS estimation of revenue-based translog production function at the macro-sector level
P	E	13	elm_rtl_ols_S	Output elasticity of intermediates, derived from OLS estimation of revenue-based translog production function at the sector level
P	E	14	elm_rtl_wd_M	Output elasticity of intermediates, derived from Wooldridge of revenue-based translog production function at the macro-sector level
P	E	15	elm_rtl_wd_S	Output elasticity of intermediates, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	16	elvi_rtl_vi_ols_M	Output elasticity of variable inputs[24], derived from OLS estimation of revenue-based translog production function at the macro-sector level
P	E	17	elvi_rtl_vi_ols_S	Output elasticity of variable inputs, derived from OLS estimation of revenue-based translog production function at the sector level
P	E	18	elvi_rtl_vi_wd_M	Output elasticity of variable inputs, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	19	elvi_rtl_vi_wd_S	Output elasticity of variable inputs, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	20	lntfp_rcd_in_ols_M	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas

[24] For the definition see Section 5.3.1

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
				production function with intangibles at the macro-sector level
P	E	21	lntfp_rcd_in_ols_S	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas production function with intangibles at the sector level
P	E	22	lntfp_rcd_ols_M	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	23	lntfp_rcd_ols_S	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	24	lntfp_rcd_wd_M	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	25	lntfp_rcd_wd_S	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	26	lntfp_rtl_ols_M	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based translog production function at the macro-sector level
P	E	27	lntfp_rtl_ols_S	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based translog production function at the sector level
P	E	28	lntfp_rtl_vi_ols_M	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based translog production function with variable inputs at the macro-sector level
P	E	29	lntfp_rtl_vi_ols_S	Logarithm of the total factor productivity, derived from OLS estimation of revenue-based translog production function with variable inputs at the sector level
P	E	30	lntfp_rtl_vi_wd_M	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based translog production function with variable inputs at the macro-sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	31	lntfp_rtl_vi_wd_S	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based translog production function with variable inputs at the sector level
P	E	32	lntfp_rtl_wd_M	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	33	lntfp_rtl_wd_S	Logarithm of the total factor productivity, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	34	lntfp_vcd_ols_M	Logarithm of the total factor productivity, derived from OLS estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	35	lntfp_vcd_ols_S	Logarithm of the total factor productivity, derived from OLS estimation of value-added based Cobb-Douglas production function at the sector level
P	E	36	lntfp_vcd_wd_M	Logarithm of the total factor productivity, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	37	lntfp_vcd_wd_S	Logarithm of the total factor productivity, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the sector level
P	E	38	mpk_rcd_ols_M	Marginal product of capital, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	39	mpk_rcd_ols_S	Marginal product of capital, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	40	mpk_rcd_wd_M	Marginal product of capital, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	41	mpk_rcd_wd_S	Marginal product of capital, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	42	mpk_rtl_ols_M	Marginal product of capital, derived from OLS estimation of revenue-based based translog production function at the macro-sector level
P	E	43	mpk_rtl_ols_S	Marginal product of capital, derived from OLS estimation of revenue-based translog production function at the sector level
P	E	44	mpk_rtl_vi_ols_M	Marginal product of capital, derived from OLS estimation of revenue-based based translog production function with variable inputs at the macro-sector level
P	E	45	mpk_rtl_vi_ols_S	Marginal product of capital, derived from the OLS estimation of revenue-based based translog production function with variable inputs at the sector level
P	E	46	mpk_rtl_vi_wd_M	Marginal product of capital, derived from the Wooldridge estimation of revenue-based based translog production function with variable inputs at the macro-sector level
P	E	47	mpk_rtl_vi_wd_S	Marginal product of capital, derived from Wooldridge estimation of revenue-based based translog production function with variable inputs at the sector level
P	E	48	mpk_rtl_wd_M	Marginal product of capital derived from Wooldridge estimation of revenue-based based translog production function at the macro-sector level
P	E	49	mpk_rtl_wd_S	Marginal product of capital, derived from Wooldridge estimation of revenue-based based translog production function at the sector level
P	E	50	mpk_vcd_ols_M	Marginal product of capital, derived from the OLS estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	51	mpk_vcd_ols_S	Marginal product of capital, derived from OLS estimation of value-added based Cobb-Douglas production function at the sector level
P	E	52	mpk_vcd_wd_M	Marginal product of capital, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the macro-sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	53	mpk_vcd_wd_S	Marginal product of capital, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the sector level
P	E	54	mpl_rcd_ols_M	Marginal productivity of labour, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	55	mpl_rcd_ols_S	Marginal productivity of labour, derived from the OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	56	mpl_rcd_wd_M	Marginal product of labour, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	57	mpl_rcd_wd_S	Marginal product of labour, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	58	mpl_rtl_ols_M	Marginal product of labour, derived from OLS estimation of revenue-based translog production function at the macro-sector level
P	E	59	mpl_rtl_ols_S	Marginal product of labour, derived from OLS estimation of revenue-based translog production function at the sector level
P	E	60	mpl_rtl_wd_M	Marginal product of labour, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	61	mpl_rtl_wd_S	Marginal product of labour, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	62	mpl_vcd_ols_M	Marginal productivity of labour, derived from OLS estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	63	mpl_vcd_ols_S	Marginal productivity of labour, derived from OLS estimation of value-added based Cobb-Douglas production function at the sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	64	mpl_vcd_wd_M	Marginal product of labour, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	65	mpl_vcd_wd_S	Marginal product of labour, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the sector level
P	E	66	mpm_rcd_ols_M	Marginal product of intermediates, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	67	mpm_rcd_ols_S	Marginal product of intermediates, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	68	mpm_rcd_wd_M	Marginal product of intermediates, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	69	mpm_rcd_wd_S	Marginal product of intermediates, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	70	mpm_rtl_ols_M	Marginal product of intermediates, derived from OLS estimation of a revenue-based translog production function at the macro-sector level
P	E	71	mpm_rtl_ols_S	Marginal product of intermediates, OLS estimation of a revenue-based translog production function at the sector level
P	E	72	mpm_rtl_wd_M	Marginal product of intermediates, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	73	mpm_rtl_wd_S	Marginal product of intermediates, derived from Wooldridge estimation of revenue-based translog production function at the sector level
P	E	74	mpvi_rtl_vi_ols_M	Marginal product of variable inputs, derived from OLS estimation of a revenue-based translog production function at the macro-sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	75	mpvi_rtl_vi_ols_S	Marginal product of variable inputs, derived from OLS estimation of a revenue-based translog production function at the macro-sector level
P	E	76	mpvi_rtl_vi_wd_M	Marginal product of variable inputs, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	77	mpvi_rtl_vi_wd_S	Marginal product of variable inputs, derived from Wooldridge estimation of revenue-based translog production function at the macro-sector level
P	E	78	ps_rcd_ols_M	Petrin-Sivadasan Gap, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	79	ps_rcd_ols_S	Petrin-Sivadasan Gap, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	80	ps_rcd_wd_M	Petrin-Sivadasan Gap, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	81	ps_rcd_wd_S	Petrin-Sivadasan Gap, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	82	ps_rtl_ols_M	Petrin-Sivadasan Gap, derived from OLS estimation of revenue based translog production function at the macro-sector level
P	E	83	ps_rtl_ols_S	Petrin-Sivadasan Gap, derived from OLS estimation of revenue based translog production function at the sector level
P	E	84	ps_rtl_wd_M	Petrin-Sivadasan Gap, derived from Wooldridge estimation of revenue based translog production function at the macro-sector level
P	E	85	ps_rtl_wd_S	Petrin-Sivadasan Gap, derived from Wooldridge estimation of revenue based translog production function at the sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	86	ps_vcd_ols_M	Petrin-Sivadasan Gap, derived from OLS estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	87	ps_vcd_ols_S	Petrin-Sivadasan Gap, derived from OLS estimation of value-added based Cobb-Douglas production function at the sector level
P	E	88	ps_vcd_wd_M	Petrin-Sivadasan Gap, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	89	ps_vcd_wd_S	Petrin-Sivadasan Gap, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the sector level
P	E	90	rts_rtl_ols_M	Returns to scale, derived from OLS estimation of revenue based translog production function at the macro-sector level
P	E	91	rts_rtl_ols_S	Returns to scale, derived from OLS estimation of revenue based translog production function at the sector level
P	E	92	rts_rtl_vi_ols_M	Returns to scale, derived from OLS estimation of value-added based translog production function with variable inputs at the macro-sector level
P	E	93	rts_rtl_vi_ols_S	Returns to scale, OLS estimation of value-added based translog production function with variable inputs at the sector level
P	E	94	rts_rtl_vi_wd_M	Returns to scale, derived from Wooldridge estimation of a value-added based translog production function with variable inputs at the macro-sector level
P	E	95	rts_rtl_vi_wd_S	Returns to scale, derived from Wooldridge estimation of value-added based translog production function with variable inputs at the sector level
P	E	96	rts_rtl_wd_M	Returns to scale, derived from Wooldridge estimation of value-added based translog production function at the macro-sector level

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	E	97	rts_rtl_wd_S	Returns to scale, derived from Wooldridge estimation of value-added based translog production function at the macro-sector level
P	E	98	tfp_rcd_ols_M	Total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	99	tfp_rcd_ols_S	Total factor productivity, derived from OLS estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	a0	tfp_rcd_wd_M	Total factor productivity, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the macro-sector level
P	E	a1	tfp_rcd_wd_S	Total factor productivity, derived from Wooldridge estimation of revenue-based Cobb-Douglas production function at the sector level
P	E	a2	tfp_rtl_ols_M	Total factor productivity, derived from OLS estimation of revenue based translog production function at the macro-sector level
P	E	a3	tfp_rtl_ols_S	Total factor productivity, derived from OLS estimation of revenue based translog production function at the sector level
P	E	a4	tfp_rtl_wd_M	Total factor productivity, derived from OLS estimation of revenue based translog production function at the macro-sector level
P	E	a5	tfp_rtl_wd_S	Total factor productivity, derived from OLS estimation of revenue based translog production function at the sector level
P	E	a6	tfp_vcd_ols_M	Total factor productivity, derived from OLS estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	a7	tfp_vcd_ols_S	Total factor productivity, derived from OLS estimation of value-added based Cobb-Douglas production function at the sector level

ID	Variable category	Numerical code	Variable Name	Definition
Productivity				
P	E	a8	tfp_vcd_wd_M	Total factor productivity, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the macro-sector level
P	E	a9	tfp_vcd_wd_S	Total factor productivity, derived from Wooldridge estimation of value-added based Cobb-Douglas production function at the sector level
Growth Rates				
P	G	00	lnlprod_rev_1y	One-year growth rate of the logarithm of real revenues-based labour productivity, in decimals
P	G	01	lnlprod_va_1y	One-year growth rate of the logarithm of real value added-based labour productivity, in decimals
P	G	02	lntfp_rcd_ols_M_1y	One-year growth rate of the logarithm of total factor productivity derived from OLS estimation of revenue-based Cobb-Douglas production function at the macro-sector level, in decimals
Values				
P	V	00	kprod_va	Capital productivity defined as real value-added divided by nominal capital
P	V	01	lnkprod_va	Logarithm of capital productivity defined as real value-added divided by nominal capital
P	V	02	lnlprod_rev	Logarithm of labour productivity defined as real revenue divided by number of employees.
P	V	03	lnlprod_va	Logarithm of labour productivity defined as defined as real value-added divided by nominal capital.
P	V	04	lnsr	Logarithm of Solow residual derived from a Cobb-Douglas production function using 2/3 labour and 1/3 real capital weights
P	V	05	lnsr_cs	Logarithm of Solow residual derived from a Cobb-Douglas production function using cost shares weights
P	V	06	lprod_rev	Labour productivity defined as real revenue divided by number of employees
P	V	07	lprod_va	Labour productivity defined as real value-added divided by number of employees

Productivity				
ID	Variable category	Numerical code	Variable Name	Definition
P	V	08	solowres	Solow residual derived from a Cobb-Douglas production function using 2/3 labour and 1/3 real capital weights
p	V	09	solowres_cs	Solow residual derived from a Cobb-Douglas production function using cost shares weights

5.2.3 Labour Variables

Table 10 Labour Variables

Labour				
ID	Variable category	Numerical code	Variable Name	Definition
Dummies				
L	D	00	high_growth	Dummy equal 1 if firm labour growth is at least 20% in the last three years, and 0 otherwise
L	D	01	t10_l_C	Dummy equal 1 if firm is among the top 10 employers at the country level, and 0 otherwise
L	D	02	t10_l_M	Dummy equal 1 if firm is among the top 10 employers at the macro-sector level, and 0 otherwise
Ratios				
L	R	00	lc_rev	Wage share defined as nominal labour costs divided by nominal revenue
L	R	01	lc_va	Wage share defined as nominal labour costs divided by nominal value-added
L	R	02	tertshare	Share of employees with tertiary education
L	R	03	ulc	Unit labour costs defined as nominal labour costs divided by real value-added
Growth Rates				
L	G	00	avg_wage_1y	One-year growth rate of average labour cost per employee

Labour				
ID	Variable category	Numerical code	Variable Name	Definition
L	G	01	firm_1y	One-year midpoint growth rate of labour
L	G	02	firm_neg_1y	One-year midpoint growth rate of labour, labour being equal to 0 or negative
L	G	03	firm_pos_1y	One-year midpoint growth rate of labour, labour being positive
L	G	04	l_1y	One-year growth rate of number of employees
L	G	05	l_3y	Three-year growth rate of number of employees
Values				
L	V	00	avg_wage	Wage defined as nominal labour cost divided by number of employees
L	V	06	jcr_sam_C	Job construction rate defined as the size-weighted average of positive firm growth rate in the given sample at the country level[25]
L	V	07	jcr_sam_M	Job construction rate defined as the size-weighted average of positive firm growth rate in the given sample at the macro-sector level
L	V	08	jcr_sam_MS	Job construction rate defined as the size-weighted average of positive firm growth rate in the given sample at the macro-sector size-class level
L	V	09	jcr_sam_N	Job construction rate defined as the size-weighted average of positive firm growth rate in the given sample at the nuts2 level
L	V	10	jcr_sam_S	Job construction rate defined as the size-weighted average of positive firm growth rate in the given sample at the sector level
L	V	16	jdr_sam_C	Job destruction rate defined as the size-weighted average of negative firm growth rate in the given sample at the country level

[25] Following Davis et al. (1996), see Section 5.3.6 for details.

Labour				
ID	Variable category	Numerical code	Variable Name	Definition
L	V	17	jdr_sam_M	Job destruction rate defined as the size-weighted average of negative firm growth rate in the given sample at the macro-sector level
L	V	18	jdr_sam_MS	Job destruction rate defined as the size-weighted average of negative firm growth rate in the given sample at the macro-sector size-class level
L	V	19	jdr_sam_N	Job destruction rate defined as the size-weighted average of negative firm growth rate in the given sample at the nuts2 level
L	V	20	jdr_sam_S	Job destruction rate defined as the size-weighted average of negative firm growth rate in the given sample at the sector level
L	V	21	l	Labour defined as number of employees
L	V	22	wage_premium_pop_S	Wage premium defined as a percentage deviation from sector median wage given the firm population
L	V	23	wage_premium_sam_S	Wage premium defined as a percentage deviation from sector median wage given the firm sample

5.2.4 Trade Variables

Table 11 Trade Variables

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
Dummies				
T	D	00	2w_exterior	Dummy equal 1 if extra-EU trade (i.e. both exports and imports) is higher than intra-EU trade, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	01	2w_exterior_adj	Dummy equal 1 if extra-EU trade (threshold adjusted version) is higher than intra-EU trade, and 0 otherwise
T	D	02	2w_extersale	Dummy equal 1 if extra-EU exports are higher than intra-EU exports, and 0 otherwise
T	D	03	2w_extersale_adj	Dummy equal 1 if extra-EU exports are higher than intra-EU exports (threshold adjusted version), and 0 otherwise
T	D	04	2w_extra	Dummy equal 1 if extra-EU trade is positive, and 0 otherwise
T	D	05	2w_extra_adj	Dummy equal 1 if extra-EU trade (threshold, adjusted version) is positive, and 0 otherwise
T	D	06	2w_interior	Dummy equal 1 if intra-EU trade is higher than extra-EU trade, and 0 otherwise
T	D	07	2w_interior_adj	Dummy equal 1 if intra-EU trade (threshold adjusted version) is higher than extra-EU trade, and 0 otherwise
T	D	08	2w_intersale	Dummy equal 1 if intra-EU exports are higher than extra-EU exports, and 0 otherwise
T	D	09	2w_intersale_adj	Dummy equal 1 if intra-EU exports are higher than extra-EU exports (threshold adjusted version), and 0 otherwise
T	D	10	2w_intra	Dummy equal 1 if intra-EU trade is positive, and 0 otherwise
T	D	11	2w_intra_adj	Dummy equal 1 if intra-EU trade (threshold adjusted version) is positive, and 0 otherwise
T	D	12	2w_total	Dummy equal 1 if intra-EU and extra-EU trade are positive, and 0 otherwise
T	D	13	2w_total_adj	Dummy equal 1 if intra-EU and extra-EU trade (threshold adjusted version) are positive, and 0 otherwise
T	D	14	exp	Dummy equal 1 if exports are positive, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	15	exp_adj	Dummy equal 1 if exports (threshold adjusted version) are positive, and 0 otherwise
T	D	16	exp_adj_con2	Dummy equal 1 if exports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	17	exp_adj_con3	Dummy equal 1 if exports (threshold adjusted version) are positive for three consecutive years, and 0 otherwise
T	D	18	exp_adj_net	Dummy equal 1 if exports are higher than imports (threshold adjusted versions), and 0 otherwise
T	D	19	exp_adj_new1	Dummy equal 1 if exports (threshold adjusted version) are positive in the current year and equal 0 in the previous year, and 0 otherwise
T	D	20	exp_adj_new2	Dummy equal 1 if exports (threshold adjusted version) are positive in the current and the next year whilst equal 0 in the previous, and 0 otherwise
T	D	21	exp_adj_non2	Dummy equal 1 if exports (threshold adjusted version) equal 0 in two consecutive years, and 0 otherwise
T	D	22	exp_adj_non3	Dummy equal 1 if exports (threshold adjusted version) equal 0 in three consecutive years, and 0 otherwise
T	D	23	exp_adj_stp1	Dummy equal 1 if exports (threshold adjusted version) are positive in the current year whilst equal 0 next year, and 0 otherwise
T	D	24	exp_adj_stp2	Dummy equal 1 if exports (threshold adjusted version) are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	25	exp_adj_swi1	Dummy equal 1 if exports (threshold adjusted version) are positive in the current year whilst

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
				equal 0 in the previous and the next year, and 0 otherwise
T	D	26	exp_adj_t10_C	Dummy equal 1 if firm's export value belongs to the top 10 values within a given country, and 0 otherwise
T	D	27	exp_adj_t10_S	Dummy equal 1 if firm's export value (threshold adjusted version) belongs to top 10 values within a given sector, and 0 otherwise
T	D	28	exp_con2	Dummy equal 1 if exports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	29	exp_con3	Dummy equal 1 if exports are positive for three consecutive years, and 0 otherwise
T	D	30	exp_ex	Dummy equal 1 if extra-EU exports are positive, and 0 otherwise
T	D	31	exp_ex_adj	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive, and 0 otherwise
T	D	32	exp_ex_adj_con2	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	33	exp_ex_adj_con3	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive for three consecutive years, and 0 otherwise
T	D	34	exp_ex_adj_new1	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive in the current year whilst equal 0 in the previous year, and 0 otherwise.
T	D	35	exp_ex_adj_new2	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive in the current and the next year whilst equal 0 in the previous year, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	36	exp_ex_adj_non2	Dummy equal 1 if extra-EU exports (threshold adjusted version) equal 0 in two consecutive years, and 0 otherwise
T	D	37	exp_ex_adj_non3	Dummy equal 1 if extra-EU exports (threshold adjusted version) equal 0 in three consecutive years, and 0 otherwise
T	D	38	exp_ex_adj_stp1	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive in the current year and equal 0 next year, and 0 otherwise
T	D	39	exp_ex_adj_stp2	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	40	exp_ex_adj_swi1	Dummy equal 1 if extra-EU exports (threshold adjusted version) are positive in the current year whilst equal 0 in the previous and the next year, and 0 otherwise
T	D	41	exp_ex_adj_t10_C	Dummy equal 1 if extra-EU firm's export value (threshold adjusted version) belongs to the top 10 values in a given country, and 0 otherwise
T	D	42	exp_ex_adj_t10_S	Dummy equal 1 if extra-EU firm's export value (threshold adjusted version) belongs to the top 10 values in a given sector, and 0 otherwise
T	D	43	exp_ex_con2	Dummy equal 1 if extra-EU exports are positive for two consecutive years, and 0 otherwise
T	D	44	exp_ex_con3	Dummy equal 1 if extra-EU exports are positive for three consecutive years, and 0 otherwise
T	D	45	exp_ex_new1	Dummy equal 1 if extra-EU exports are positive in the current year and equal 0 in the previous year, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	46	exp_ex_new2	Dummy equal 1 if extra-EU exports are positive in the current and the next year whilst equal 0 in the previous year, and 0 otherwise
T	D	47	exp_ex_non2	Dummy equal 1 if extra-EU exports equal 0 for two consecutive years, and 0 otherwise
T	D	48	exp_ex_non3	Dummy equal 1 if extra-EU exports equal 0 for three consecutive years, and 0 otherwise
T	D	49	exp_ex_stp1	Dummy equal 1 if extra-EU exports are positive in the current year whilst equal 0 next year, and 0 otherwise
T	D	50	exp_ex_stp2	Dummy equal 1 if extra-EU exports are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	51	exp_ex_swi1	Dummy equal 1 if extra-EU exports are positive in the current year whilst 0 in the previous and the next year, and 0 otherwise
T	D	52	exp_ex_t10_C	Dummy equal 1 if firm's extra-EU export value belongs to the top 10 within a given country, and 0 otherwise
T	D	53	exp_ex_t10_S	Dummy equal 1 if firm's extra-EU export value belongs to the top 10 values within a given sector, and 0 otherwise
T	D	54	exp_in	Dummy equal 1 if intra -EU exports are positive, and 0 otherwise
T	D	55	exp_in_adj	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive, and 0 otherwise
T	D	56	exp_in_adj_con2	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	57	exp_in_adj_con3	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive for three consecutive year, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	58	exp_in_adj_new1	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive in the current year whilst equal 0 in the previous year, and 0 otherwise
T	D	59	exp_in_adj_new2	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive in the current and the next year whilst equal 0 in the previous year, and 0 otherwise
T	D	60	exp_in_adj_non2	Dummy equal 1 if intra-EU exports (threshold adjusted version) equal 0 for two consecutive years, and 0 otherwise
T	D	61	exp_in_adj_non3	Dummy equal 1 if intra-EU exports (threshold adjusted version) equal 0 for three consecutive years, and 0 otherwise
T	D	62	exp_in_adj_stp1	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive in the current year and equal 0 next year, and 0 otherwise
T	D	63	exp_in_adj_stp2	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	64	exp_in_adj_swi1	Dummy equal 1 if intra-EU exports (threshold adjusted version) are positive in the current year whilst equal 0 in the previous and the next year, and 0 otherwise
T	D	65	exp_in_adj_t10_C	Dummy equal 1 if firm's intra-EU export value (threshold adjusted version) belongs to the top 10 values within a given country, and 0 otherwise
T	D	66	exp_in_adj_t10_S	Dummy equal 1 if firm's intra-EU export value (threshold adjusted version) belongs to the top 10 values within a given sector, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	67	exp_in_con2	Dummy equal 1 if intra-EU exports are positive for two consecutive years, and 0 otherwise
T	D	68	exp_in_con3	Dummy equal 1 if intra-EU exports are positive for three consecutive years, and 0 otherwise
T	D	69	exp_in_new1	Dummy equal 1 if intra-EU exports are positive in the current year and equal 0 in the previous year, and 0 otherwise
T	D	70	exp_in_new2	Dummy equal 1 if intra-EU exports are positive in the current and the next year whilst equal 0 in the previous year, and 0 otherwise
T	D	71	exp_in_non2	Dummy equal 1 if intra-EU exports equal 0 for two consecutive years, and 0 otherwise
T	D	72	exp_in_non3	Dummy equal 1 if intra-EU exports equal 0 for three consecutive years, and 0 otherwise
T	D	73	exp_in_stp1	Dummy equal 1 if intra-EU exports are positive in the current year whilst equal 0 next year, and 0 otherwise
T	D	74	exp_in_stp2	Dummy equal 1 if intra-EU exports are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	75	exp_in_swi1	Dummy equal 1 if intra-EU exports are positive in the current year whilst equal 0 in the previous and the next year, and 0 otherwise
T	D	76	exp_in_t10_C	Dummy equal 1 if firm's intra-EU export value belongs to the top 10 values within a given country, and 0 otherwise
T	D	77	exp_in_t10_S	Dummy equal 1 if firm's intra-EU export value belongs to the top 10 values within a given sector, and 0 otherwise
T	D	78	exp_net	Dummy equal 1 if exports are higher than imports, and 0 otherwise
T	D	79	exp_new1	Dummy equal 1 if exports are positive in the current year and equal 0 in the previous year, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	80	exp_new2	Dummy equal 1 if exports are positive in the current and the next year whilst equal 0 in the previous, and 0 otherwise
T	D	81	exp_non2	Dummy equal 1 if exports equal 0 for two consecutive years, and 0 otherwise
T	D	82	exp_non3	Dummy equal 1 if exports equal 0 for three consecutive years, and 0 otherwise
T	D	83	exp_stp1	Dummy equal 1 if exports are positive in the current year whilst equal 0 next year, and 0 otherwise
T	D	84	exp_stp2	Dummy equal 1 if exports are positive in the current and the previous year whilst equal 0 in the next year, and 0 otherwise
T	D	85	exp_swi1	Dummy equal 1 if exports are positive in the current year whilst equal 0 in the previous and the next year, and 0 otherwise
T	D	86	exp_t10_C	Dummy equal 1 if firm's export value belongs to the top 10 values within a given country, and 0 otherwise
T	D	87	exp_t10_S	Dummy equal 1 if firm's export value belongs to the top 10 values within a given sector, and 0 otherwise
T	D	88	imp	Dummy equal 1 if imports are positive, 0 otherwise
T	D	89	imp_adj	Dummy equal 1 if imports (threshold adjusted version) are positive, and 0 otherwise
T	D	90	imp_adj_con2	Dummy equal 1 if imports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	91	imp_adj_con3	Dummy equal 1 if imports (threshold adjusted version) are positive for three consecutive years, and 0 otherwise
T	D	92	imp_adj_new2	Dummy equal 1 if imports (threshold adjusted version) are positive in the current and the

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
				next year whilst 0 in the previous year, and 0 otherwise
T	D	93	imp_adj_t10_C	Dummy equal 1 if firm's import value belongs to the top 10 values within a given country, and 0 otherwise
T	D	94	imp_adj_t10_S	Dummy equal 1 if firm's import value (threshold adjusted version) belong to the top 10 values within a given sector, and 0 otherwise
T	D	95	imp_con2	Dummy equal 1 if imports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	96	imp_con3	Dummy equal 1 if imports are positive for three consecutive years, and 0 otherwise
T	D	97	imp_ex	Dummy equal 1 if extra-EU imports are positive, 0 otherwise
T	D	98	imp_ex_adj	Dummy equal 1 if extra-EU imports (threshold adjusted version) are positive, and 0 otherwise
T	D	99	imp_ex_adj_con2	Dummy equal 1 if extra-EU imports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	a0	imp_ex_adj_con3	Dummy equal 1 if extra-EU imports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	a1	imp_ex_adj_t10_C	Dummy equal 1 if firm's extra-EU import value (threshold adjusted version) belongs to the top 10 values within a given country, and 0 otherwise
T	D	a2	imp_ex_adj_t10_S	Dummy equal 1 if firm's extra-EU import value (threshold adjusted version) belongs to the top 10 values within a given sector, and 0 otherwise
T	D	a3	imp_ex_con2	Dummy equal 1 if extra-EU imports

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
				are positive for two consecutive years, and 0 otherwise
T	D	a4	imp_ex_con3	Dummy equal 1 if extra-EU imports are positive for three consecutive years, and 0 otherwise
T	D	a5	imp_ex_t10_C	Dummy equal 1 if firm's extra-EU import value belongs to the top 10 values within a given country, and 0 otherwise
T	D	a6	imp_ex_t10_S	Dummy equal 1 if firm's extra-EU import value belongs to the top 10 values within a given sector, and 0 otherwise
T	D	a7	imp_in	Dummy equal 1 if intra -EU imports are positive, 0 otherwise
T	D	a8	imp_in_adj	Dummy equal 1 if intra-EU imports (threshold adjusted version) are positive, 0 otherwise
T	D	a9	imp_in_adj_con2	Dummy equal 1 if intra-EU imports (threshold adjusted version) are positive for two consecutive years, and 0 otherwise
T	D	b0	imp_in_adj_con3	Dummy equal 1 if intra-EU imports (threshold adjusted version) are positive for three consecutive years, and 0 otherwise
T	D	b1	imp_in_adj_t10_C	Dummy equal 1 if firm's intra-EU import value (threshold adjusted version) belongs to the top 10 values within a given country, and 0 otherwise
T	D	b2	imp_in_adj_t10_S	Dummy equal 1 if firm's intra-EU import value (threshold adjusted version) belongs to the top 10 values within a given sector, and 0 otherwise
T	D	b3	imp_in_con2	Dummy equal 1 if intra-EU imports are positive for two consecutive years, and 0 otherwise

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	D	b4	imp_in_con3	Dummy equal 1 if intra-EU imports are positive for three consecutive years, and 0 otherwise
T	D	b5	imp_in_t10_C	Dummy equal 1 if firm's intra-EU import value belongs to the top 10 values within a given country, and 0 otherwise
T	D	b6	imp_in_t10_S	Dummy equal 1 if firm's intra-EU import value belongs to the top 10 values within a given sector, and 0 otherwise
T	D	b7	imp_new2	Dummy equal 1 if imports are positive in the current and the next year whilst equal 0 in the previous year, and 0 otherwise
T	D	b8	imp_t10_C	Dummy equal 1 if firm's import value belongs to the top 10 values within a given country, and 0 otherwise
T	D	b9	imp_t10_S	Dummy equal 1 if firm's import value belongs to the top 10 values within a given sector, and 0 otherwise
Growth Rates				
T	G	00	exp_1y	One-year growth rate of export value (threshold adjusted version)
T	G	01	exp_adj_1y	One-year growth rate of export value (threshold adjusted version)
Ratios				
T	R	00	exp_adj_pop_C	Ratio of export value (threshold adjusted version) to population total export value for the country
T	R	01	exp_adj_pop_S	Ratio of export value (threshold adjusted version) to population total export value for the sector
T	R	02	exp_adj_rev	Ratio of export value (threshold adjusted version) to revenue

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	R	03	exp_adj_sam_C	Ratio of export value (threshold adjusted version) to sample total export value for the country
T	R	04	exp_adj_sam_S	Ratio of export value (threshold adjusted version) to sample total export for the sector
T	R	05	exp_adj_va_rev	Estimate of value added in export (threshold adjusted version) based on share of value added in revenue
T	R	06	exp_ex_adj_pop_C	Ratio of extra-EU export value (threshold adjusted version) to population total extra-EU export value for the country
T	R	07	exp_ex_adj_pop_S	Ratio of extra-EU export value (threshold adjusted version) to population total extra-EU export value for the sector
T	R	08	exp_ex_adj_rev	Ratio of extra-EU export value (threshold adjusted version) to revenue
T	R	09	exp_ex_adj_sam_C	Ratio of extra-EU export value (threshold adjusted version) to sample total extra-EU export for the country
T	R	10	exp_ex_adj_sam_S	Ratio of extra-EU export value (threshold adjusted version) to sample total extra-EU export for the sector
T	R	11	exp_ex_adj_va_rev	Estimate of value added in extra-EU export (threshold adjusted version) based on share of value added in revenue
T	R	12	exp_ex_pop_C	Ratio of extra-EU export value to population total extra-EU export value for the country
T	R	13	exp_ex_pop_S	Ratio of extra-EU export value to population total extra-EU export value for the sector
T	R	14	exp_ex_rev	Ratio of extra-EU export to revenue
T	R	15	exp_ex_sam_C	Ratio of extra-EU export value to sample total extra-EU export value for the country
T	R	16	exp_ex_sam_S	Ratio of e extra-EU export value to sample total extra-EU export value for the sector

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	R	17	exp_ex_va_rev	Estimate of value added in extra-EU export based on share of value added in revenue
T	R	18	exp_in_adj_pop_C	Ratio of intra-EU export value (threshold adjusted version) to population total intra-EU export for the country
T	R	19	exp_in_adj_pop_S	Ratio of intra-EU export value (threshold, adjusted version) to population total intra-EU export for the sector
T	R	20	exp_in_adj_rev	Ratio of intra-EU export value (threshold adjusted version) to turnover (nominal)
T	R	21	exp_in_adj_sam_C	Ratio of intra-EU export value (threshold adjusted version) to sample total intra-EU export for the country
T	R	22	exp_in_adj_sam_S	Ratio of intra-EU export value (threshold adjusted version) to sample total intra-EU export for the sector
T	R	23	exp_in_adj_va_rev	Estimate of value added in intra-EU export (threshold adjusted version) based on share of value added in revenue
T	R	24	exp_in_pop_C	Ratio of intra-EU export value to population total intra-EU export for the country
T	R	25	exp_in_pop_S	Ratio of intra-EU export value to population total intra-EU export for the sector
T	R	26	exp_in_rev	Ratio of intra-EU export value to revenue
T	R	27	exp_in_sam_C	Ratio of intra-EU export value to sample total intra-EU export for the country
T	R	28	exp_in_sam_S	Ratio of intra-EU export value to sample total intra-EU export for the sector
T	R	29	exp_in_va_rev	Estimate of value added in intra-EU export based on share of value added in revenue
T	R	30	exp_pop_C	Ratio of export value to population total export for the country
T	R	31	exp_pop_S	Ratio of export value to population total export for the sector

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	R	32	exp_rev	Ratio of export value to revenue
T	R	33	exp_sam_C	Ratio of export value to sample total export for the country
T	R	34	exp_sam_S	Ratio of export value to sample total export for the sector
T	R	35	exp_va_rev	Estimate of value added in export based on share of value added in revenue
T	R	36	imp_adj_pop_C	Ratio of import value (threshold adjusted version) to population total import value for the country
T	R	37	imp_adj_pop_S	Ratio of import value (threshold adjusted version) to population total import value for the sector
T	R	38	imp_adj_rev	Ratio of import value (threshold adjusted version) to revenue
T	R	39	imp_adj_sam_C	Ratio of import value (threshold adjusted version) to sample total import value for the country
T	R	40	imp_adj_sam_S	Ratio of import value (threshold adjusted version) to sector total import value for the country
T	R	41	imp_ex_adj_pop_C	Ratio of extra-EU import value (threshold adjusted version) to population total extra-EU import value for the country
T	R	42	imp_ex_adj_pop_S	Ratio of extra-EU import value (threshold adjusted version) to population total extra-EU import value for the sector
T	R	43	imp_ex_adj_rev	Ratio of extra-EU import value (threshold adjusted version) to revenue
T	R	44	imp_ex_adj_sam_C	Ratio of extra-EU import value (threshold adjusted version) to sample total extra-EU import value for the country

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	R	45	imp_ex_adj_sam_S	Ratio of extra-EU import value (threshold adjusted version) to sample total extra-EU import value for the sector
T	R	46	imp_ex_pop_C	Ratio of extra-EU import value to population total extra-EU import value for the country
T	R	47	imp_ex_pop_S	Ratio of extra-EU import value to population total import value for the sector.
T	R	48	imp_ex_rev	Ratio of extra-EU import value to revenue
T	R	49	imp_ex_sam_C	Ratio of extra-EU import value to sample total extra-EU import value for the country
T	R	50	imp_ex_sam_S	Ratio of extra-EU import value to sample total extra-EU import value for the sector
T	R	51	imp_in_adj_pop_C	Ratio of intra-EU import value (threshold adjusted version) to population total intra-EU import value for the country
T	R	52	imp_in_adj_pop_S	Ratio of intra-EU import value (threshold adjusted version) to population total intra-EU import value for the sector
T	R	53	imp_in_adj_rev	Ratio of intra-EU import value (threshold adjusted version) to revenue
T	R	54	imp_in_adj_sam_C	Ratio of intra-EU import value (threshold adjusted version) to sample total intra-EU import value for the country
T	R	55	imp_in_adj_sam_S	Ratio of intra-EU import value (threshold adjusted version) to sample total intra-EU import value for the sector
T	R	56	imp_in_pop_C	Ratio of intra-EU import value to population total import value for the country.
T	R	57	imp_in_pop_S	Ratio of intra-EU import value to population total intra-EU import value for the sector
T	R	58	imp_in_rev	Ratio of intra-EU import value to revenue
T	R	59	imp_in_sam_C	Ratio of intra-EU import value to sample total intra-EU import value for the country

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	R	60	imp_in_sam_S	Ratio of intra-EU import value to sample total intra-EU import value for the sector
T	R	61	imp_pop_C	Ratio of import value to population total import value for the country
T	R	62	imp_pop_S	Ratio of import value to population total import value for the sector
T	R	63	imp_rev	Ratio of import value to revenue
T	R	64	imp_sam_C	Ratio of import value to sample total import value for the country
T	R	65	imp_sam_S	Ratio of import value to sample total import value for the sector
T	R	66	imp_exp	Import intensity defined as imports divided by exports, proxy for Global Value Chain participation
T	R	67	imp_exp_adj	Import intensity defined as imports divided by exports (threshold adjusted version), proxy for Global Value Chain participation
Values				
T	V	00	dom_sale	Domestic sales defined as revenue minus exports
T	V	01	dom_sale_adj	Domestic sales defined as revenue minus exports (threshold adjusted version)
T	V	02	exp	Export value (nominal exports)
T	V	03	exp_adj	Export value (threshold adjusted version)
T	V	04	exp_ex	Extra-EU export value
T	V	05	exp_ex_adj	Extra-EU export value (threshold adjusted version)
T	V	06	exp_in	Intra-EU export value
T	V	07	exp_in_adj	Intra-EU export value (threshold adjusted version)
T	V	08	imp	Import value
T	V	09	imp_adj	Import value (threshold adjusted version)
T	V	10	imp_ex	Extra-EU import value

Trade				
ID	Variable category	Numerical code	Variable Name	Definition
T	V	11	imp_ex_adj	Extra-EU import value (threshold adjusted version)
T	V	12	imp_in	Intra-EU import value
T	V	13	imp_in_adj	Intra-EU import value (threshold adjusted version)

5.2.5 Finance Variables

Table 12 Finance Variables

Finance				
ID	Variable category	Numerical code	Variable Name	Definition
Dummies				
F	D	00	absconstr	Dummy equal 1 if subject to absolute credit constraints, and 0 otherwise
F	D	01	safe	Dummy equal 1 if subject to credit constraints based on SAFE score, and 0 otherwise
F	D	02	t10_rev_C	Dummy equal 1 if firm is among Top10 revenue firms at the country level, and 0 otherwise
F	D	03	t10_rev_M	Dummy equal 1 if firm is among Top10 revenue firms at the mac-sector level, and 0 otherwise
F	D	04	t10_rev_S	Dummy equal 1 if firm is among Top10 revenue firms at the sector level, and 0 otherwise
F	D	05	zombie_intcov	Dummy equal 1 if interest payments exceed operational profit for three years, whilst profit

Finance				
ID	Variable category	Numerical code	Variable Name	Definition
				is positive and no high labour growth, and 0 otherwise
F	D	06	zombie_intcov_np	Dummy equal 1 if interest payments exceed operational profit for three years and no high labour growth, and 0 otherwise
F	D	07	zombie_negprof	Dummy equal 1 if negative profit for three years and no high labour growth, and 0 otherwise
Growth Rates				
F	G	00	rev_1y	One-year growth rate of nominal revenue
F	G	01	rk_1y	One-year growth rate of real capital
F	G	02	rk_3y	Three-year growth rate of real capital
Ratios				
F	R	00	capcost_m	Ratio of capital costs to intermediate input expenditures
F	R	01	cash_ta	Ratio of cash to total assets
F	R	02	cashflow_ta	Ratio of cash flow to total assets
F	R	03	collateral_ta	Ratio of capital to total assets
F	R	04	costcov	Ratio of revenue to labour costs and intermediate input expenditures
F	R	05	costcov_vi	Ratio of revenue to labour costs, intermediate input expenditures and capital costs
F	R	06	depr_ta	Ratio of depreciation to total assets
F	R	07	div_ta	Ratio of dividends to total assets
F	R	08	equity_debt	Ratio of equity to debt
F	R	09	equity_ta	Ratio of equity to total assets
F	R	10	fingap	Financial gap defined as the ratio of investment (change in nominal capital plus depreciation) net of cash flow to revenue
F	R	11	ifa_k	Ratio of intangible fixed assets to capital
F	R	12	inte_debt	Ratio of interest paid to average debt (based on current and previous year)

Finance				
ID	Variable category	Numerical code	Variable Name	Definition
F	R	13	inv_rev	Ratio of inventories to revenue
F	R	14	invest_k	Ratio of investment (change in nominal capital plus depreciation) to nominal capital in the previous year
F	R	15	lc_capcost	Ratio of labour costs to capital costs
F	R	16	lc_l	Ratio of labour costs to labour
F	R	17	lc_m	Ratio of labour costs to intermediate input expenditures
F	R	18	leverage	Ratio of debt to total assets
F	R	19	op_inte	Ratio of operating profits to interest payments
F	R	20	pcm_kfix	Price cost margin excluding capital costs (assumed fixed)
F	R	21	pcm_kvar	Price cost margin including capital costs
F	R	22	profitmargin	Ratio of operating profit to revenue
F	R	23	rd_costs	Ratio of R&D expenditures to total costs
F	R	24	rd_m	Ratio of R&D expenditures to intermediate input expenditures
F	R	25	rev_capcost	Ratio of revenue to capital costs
F	R	26	rev_ener	Ratio of revenue to energy input expenditures
F	R	27	rev_lc	Ratio of revenue to labour costs
F	R	28	rev_lc_m	Ratio of revenue to sum of labour costs and intermediate input expenditures
F	R	29	rev_m	Ratio of revenue to intermediate input expenditures
F	R	30	rk_l	Ratio of real capital to labour
F	R	31	roa	Return on assets defined as the ratio of operating profit to average total assets (based on current and previous year)
F	R	32	trade_credit	Ratio of accounts payable to total assets
F	R	33	trade_debt	Ratio of accounts receivable to total assets
F	R	34	va_ener	Ratio of value added to energy costs
F	R	35	va_rev	Ratio of value added to revenue

Finance				
ID	Variable category	Numerical code	Variable Name	Definition
Values				
F	V	00	capcost	Capital costs defined as the sum of depreciation, interest paid and imputed interest on equity
F	V	01	debt	Long-term debt plus short-term debt
F	V	02	debt_fin	Financial debt defined as the sum of current and noncurrent liabilities excluding creditors (accounts payable)
F	V	03	n_ener	Nominal energy inputs
F	V	04	nk	Nominal capital stock
F	V	05	nlc	Nominal labour costs
F	V	06	nm	Nominal intermediate input expenditures
F	V	07	nrd	Nominal R&D expenditures
F	V	08	nrev	Nominal revenue
F	V	09	nrev_sq	Nominal revenue squared
F	V	10	nva	Nominal value added
F	V	11	nva_pos	Positive nominal value added
F	V	12	nvi	Sum of nominal intermediate input expenditures and nominal labour costs
F	V	13	rifa	Real intangible assets
F	V	14	rk	Real capital stock
F	V	15	rlc	Real labour costs
F	V	16	rm	Real intermediate input expenditures
F	V	17	rrev	Real revenue
F	V	18	rva	Real value added
F	V	19	rva_pos	Positive real value added
F	V	20	ta	Total assets
F	V	21	y_zombie_intcov	Number of consecutive years for being zombie based on intcov definition
F	V	22	y_zombie_intcov_np	Number of consecutive years for being zombie based on intcov_np definition

Finance				
ID	Variable category	Numerical code	Variable Name	Definition
F	V	23	y_zombie_negprof	Number of consecutive years for being zombie based on negprof definition

5.2.6 Other Variables

Table 13 Other Variables

Other				
ID	Variable category	Numerical code	Variable Name	Definition
Dummies				
O	D	00	exit	Dummy equal 1 if firm exits the market in the current or next year, and 0 otherwise
O	D	01	firm_age_medium	Dummy equal 1 if firm age is more than 5 and less than 25 years, and 0 otherwise
O	D	02	firm_age_new	Dummy equal 1 if firm age is less than 3 years, and 0 otherwise
O	D	03	firm_age_old	Dummy equal 1 if firm age is at least 25 years, and 0 otherwise
O	D	04	firm_age_young	Dummy equal 1 if firm age is at least 3 years and at most 5 years, and 0 otherwise
O	D	05	foreign_own	Dummy equal 1 if more than 50% of the shares are controlled by foreign owners, and 0 otherwise
O	D	06	legal_form_1	Dummy equal 1 if limited liability company or limited liability partnership, and 0 otherwise
O	D	07	publ_own	Dummy equal 1 if more than 50% of the shares are controlled by government (directly or indirectly), and 0 otherwise
Values				
O	V	00	firm_age	Age of firm in years
O	V	01	firm_age_atexit	Age of exiting firm in years
O	V	02	years_till_exit	Number of years before exiting the market

5.3 Derivation of Indicators (More Complex Variables)

This section discusses the calculation and theoretical background of a selected number of more complex variables. Specifically, productivity indicators, zombie indicators, indicators of financial constraints, the Dobbelaere-Mairesse indicator of labour market imperfections, the Petrin-Sivadsan gap indicator, markups and job creation and destruction rate indicators.

5.3.1 Production Function Estimation, TFP, and Marginal Products

Several indicators within the CompNet database rely on production function estimation techniques. Among others, these include measures of productivity, markups, and allocative efficiency measures. Given the importance of the production function estimation for the CompNet database, we will discuss the applied methodology briefly in this section before we describe the indicators derived from the recovered production function parameters.

We estimate several different types of production functions in the two-digit sector, the macro-sector, and the country level. In most cases, we rely on a Cobb-Douglas specification. For a few variables, however, we also estimate a translog type of production function as this allows for time-varying output elasticities. As input variables, capital, labour, intermediates & variable inputs[26] are used in different combinations. Besides that, we apply gross-output and value-added production models. For the sake of brevity, we cover only the value-added production function estimation in this section.

The value-added production function is given in logs by:

$$(1) \quad rva_{it} = \theta^k k_{it} + \theta^l l_{it} + a_{it} + \varepsilon_{it},$$

where rva_{it} is real value-added, k_{it} is the real book value of net capital, l_{it} is total employment, and a_{it} *denotes hicks-neutral productivity*. ε_{it} is an i.i.d. error term and θ^x denotes the output elasticity of input $x = \{k_{it}, l_{it}\}$ To control for productivity and potential endogeneity concerns, we apply a control function approach as in Olley and Pakes (1996) and Levinsohn and Petrin (2003). We base our control function on firms' intermediate input

[26] Defined as a single weighted combination of labour and intermediate inputs.

decision (m_{it}) Assuming that productivity evolves according to a Markov process and plugging in our control function for productivity, we can rewrite equation (3) as:

$$(2) \quad rva_{it} = \theta^k k_{it} + \theta^l l_{it} + g_{it-1}(k_{it-1}, m_{it-1}) + \xi_{it} + \varepsilon_{it},$$

where ξ_{it} denotes the innovation in productivity. We approximate $g_{it-1}(k_{it-1}, m_{it-1})$ with a third-order polynomial in all of its variables. Following Wooldridge (2009), we estimate the production function in one step using lagged values of employment as instruments for its contemporaneous values. When estimating the production function, we also control for a full set of time dummies.

In order to obtain consistent estimates with sufficient degrees of freedom, we require a minimum of 100 observations per sector and year.

Having estimated the output elasticities from the production function, we can calculate total factor productivity in the following way:

$$(3) \quad TFP_{it} = a_{it} = rva_{it} - (\theta^k k_{it} + \theta^l l_{it}).^{27}$$

Using the estimated coefficients of capital and labour, i.e. θ^k and θ^l, it is possible to estimate the marginal product of both inputs:

$$(4) \quad MRPK_{it} = \frac{\theta^k{}_1 rva_{it}}{k_{it}}$$

$$(5) \quad MRPL_{it} = \frac{\theta^l rva_{it}}{l_{it}}$$

Table 14 showcases the inputs and methods used for the production function estimation.

Table 14 Production Function (Weighted Two-Step Instrumental Variable Regression

Dependent Variables	Polynomials in lagged K, L M		Production function
FV17_rrev FV18_rva	*ln_K_l1* *ln_M_l1* *k2_l1* *m2_l1* *k3_l1*	*m3_l1* *km_l1* *k2m_l1* *km2_l1*	Translog Cobb-Douglas

[27] In addition, a markup adjusted TFP following Gal (2013) is available as well.

5.3.2 Allocative Efficiency: Static and Dynamic

Static Allocative Efficiency (Olley and Pakes, 1996)

Olley and Pakes introduced a very simple-to-compute indicator of allocative efficiency measured by the covariance between productivity and size, usually labelled as "OP gap".

Let y_{st} be productivity in industry *s* at time *t*, measured as a weighted average of firm-level productivity ω_{it}, with shares of industry size as weights.

The productivity of industry s can be decomposed as:

(1) $$y_{si} = \sum_{i \in S} \theta_{it}\omega_{it} = \bar{\omega}_{st} + \sum_{i \in S}(\theta_{it} - \bar{\theta})(\omega_{it} - \bar{\omega}_{st})$$

where *S* is the set of firms belonging to an aggregation level s, θ_{it} and ω_{it} represent size and productivity of firm *i* at time *t*, respectively, $\theta_{st} = \sum_{i \in S}(\theta_{it} - \bar{\theta})(\omega_{it} - \bar{\omega}_{st})$ and ω_{st} represent the unweighted mean size and productivity of aggregation level *s* at time *t*, respectively.

The decomposition splits the weighted average of firm productivity into two components: the unweighted mean and the covariance between productivity and size. The latter is often interpreted as a measure of allocative efficiency as it reflects the extent to which firms with higher than average productivity have a greater market share in terms of size.

In this vintage, the decompositions were not only calculated for productivity, but also for a range of other firm characteristics. This allows for a simple way to introduce weighted averages (i.e. aggregate equivalents) for a series of other variables. An overview of available decompositions is given in Table 15. We compute decompositions at the country, sector, nuts2 and macro-sector level.

Table 15 Overview of Decompositions

op_decomp_	country_	weighted_sample
	mac_sector_	weighted_sample
	nuts2_	weighted_sample
	sector_	weighted_sample
foster_decomp_	mac_sector_	weighted_sample
	nuts2_	weighted_sample

	sector_	weighted_sample

Dynamic allocative efficiency (Foster, Haltiwanger, and Krizan, 2006)

The covariance between size and productivity provides a snap-shot of market allocative efficiency, that is, of how resources are allocated at a certain moment in time.

A complementary way of exploring the question is looking at how resources move between two points in time across firms, hoping that they will be released from low productive/exiting units and reallocated to more productive/entering firms.

Let, as before, y_{st} be aggregation level *s* productivity at time *t*, measured as a weighted average of firm-level productivity ω_{it}, with shares of industry size as weights. Following Foster et al. (2006), the change in productivity of industry s from time t-k to time t can be decomposed as:

$$(2) \qquad \Delta y_{st} = \sum_{i\in C} \theta_{i,t-k}\Delta\omega_{it} + \sum_{i\in C} \omega_{i,t-k} - \widehat{\omega}_{s,t-k})\Delta\theta_{it} + \sum_{i\in C} \Delta\theta_{it}\ \Delta\omega\_it + \sum_{i\in N} \theta_{it}\left(\omega_{it} - \widehat{\omega}_{s,t-k}\right) - \sum_{i\in X} \theta_{i,t-k}\left(\omega_{i,t-k} - \widehat{\omega}_{s,t-k}\right)$$

Where Δ is the differential operator between *t-k* and *t*; *C* denotes continuing firms, *N* denotes entering firms, and *X* denotes exiting firms; ϑ_{it} and $\omega_{i,t}$ represent size and productivity of firm *i* at time *t*, respectively, ϑ_{st} and ω_{st} represent the weighted mean size and productivity of aggregation level *s* at time *t*, respectively. The first three terms capture the contribution of within-firm dynamics, between-firm dynamics and a covariance-term between $\omega_{i,t}$ and the size of firms to the change aggregate in productivity, y_{st}, respectively. The last two terms capture the contribution of entering and exiting firms. In our database, we only compute the first three terms as we do not have reliably information for entry and exit across a large set of countries. The sum of the latter two terms can, however, be recovered by subtracting the first three terms from the aggregate value. We advise, however, to carefully interpret this residuum as entry and exit might also refer to sample entry and exit instead of true entry and exit. A large value in the residuum term may thus reflect a large rotation in the firm sample. We compute this decomposition at the country, sector, nuts2 and macro-sector level.

Petrin-Sivadasan Gap (Petrin and Sivadasan 2013)

Petrin and Sivadasan (2013) proposed a measure of labour's allocative efficiency, based on the absolute difference between the value of labour's marginal product and its marginal cost. Following Petrin and Sivadasan (2013) closely, we approximate the marginal cost of a labour

input with the average wage. Hence, the absolute gap between the value of the marginal product of labour and its wage can be written as:

$$(3) \qquad |G_{it}| = |VMPL_{it} - w_{it}|,$$

where $VMPL_{it}$ denotes the value of the marginal product of labour, which we derive from a gross output production function. w_{it} symbolises the average wage. To ensure comparability over time, we deflate $|G_{it}|$ using a GDP deflator.

5.3.3.4 Hsieh-Klenow Indicator (Hsieh and Klenow 2009)

Building on the work of Hsieh and Klenow (2009), we also estimate the dispersion of marginal revenue products, which (under very specific assumptions discussed in Hsieh and Klenow (2009) and Haltiwanger et al. (2018)) provides an additional misallocation measure. To do so, we calculate the unconditional sector, macro-sector, and country level standard deviation of marginal revenue products. Additionally, we also apply a more sophisticated approach following Kehrig (2011).

5.3.3 Distressed Firms

"Distressed firms", sometimes also called "zombie firms", are often described in the literature as firms which, in a perfectly competitive market, would have been forced to exit the market already. There are many ways of defining zombie firms, see for example Caballero et al. (2008) or McGowan et al. (2013). The CompNet dataset includes three different zombie firm dummy specifications to identify distressed firms. These different indicators have different rationales of defining a zombie firm to mirror the variety present in the literature. In the following the three types of zombie firm indicators are discussed: negative profits, "not-high-growth" and interest coverage-based indicators:

Variable:	***Zombie_negprof***
Description:	Dummy equal 1 if firm reports negative profit for three consecutive years and is not considered to be high labour growth firm,[28] and 0 otherwise.

[28] High growth firms are defined as firms with a three-year employment growth rate 20% or more.

Variable:	*Zombie_intcov*
Description:	Dummy equal 1 if firm reports interest payments exceeding operational profit for three consecutive years and is not considered to be high labour growth firm, and 0 otherwise. The profit is assumed to be positive (i.e. only firms with positive profit are taken into account in this case).

Variable:	*Zombie_intcov_np*
Description:	Dummy equal 1 if firm reports interest payments exceeding operational profit for three consecutive years and is not considered to be high labour growth firm, and 0 otherwise. The profit may be also negative.

5.3.4 Indicators of Credit Constraint

For the purpose of the analysis of credit constrained firms and their prevalence, the CompNet dataset contains two indicators, *safe* and *abconstr*. The first indicator takes value 1 when a firm is classified as credit constrained and 0 otherwise. The decision whether the firm is considered credit constrained or not, follows several consecutive steps.

In the first step firms' responses about binding credit constraints from the Survey on Access to Finance of Enterprises (SAFE)[29] are matched with their financial characteristics available in the AMADEUS database from Bureau van Dijk. In the second step the impact of several indicators of the firm's financial position on its probability to be credit constrained is estimated. More concretely, the regression equation is the following:

$$(1) \qquad Prob(credit\_constraint) = \alpha + \beta_1 \cdot finlev + \beta_2 \cdot ifp + \beta_3 \cdot profitmargin + \beta_4 collateral + \beta_5 \cdot cash_{holdings} + \beta_6 \cdot lnTA + \gamma \cdot control\ var + \varepsilon,$$

[29] The SAFE is conducted by the ECB jointly with the European Commission twice per year. The survey intends to assess the financial conditions of firms in the euro area (the survey is also conducted for some non-euro area countries). It defines a firm as credit constrained if: the firm reports loan applications which were rejected; the firm reports loan applications for which only a limited amount was granted; the firm reports loan applications which were rejected by the firms because the borrowing costs were too high; the firm did not apply for a loan for fear of rejection (i.e. discouraged borrowers).

where $finlev_{\mathrm{i}}$ is the financial leverage, ifp_i is the index of financial pressure, pm_i is profit margin, $coll_i$ is collateral, $cashH_i$ is cash holding and TA_i are the total assets for firm i. The control variables are time, sector, firm-size and country-specific effects. For a more detailed explanation of the variables used in the regression, see Ferrando et al. (2015).

The third step is to use the coefficients of the estimated above mentioned probit regression to compute a predicted constrained score for the firms in the CompNet dataset, depending on the value of their financial position indicators. This is what we call the "SAFE score", which is computed as:

$$(2) \qquad SAFE\_score_{\mathrm{i}} = -5.47 + 0.07 \cdot finlev_i + 0.46 \cdot ifp_i - 0.50 \cdot pm_i - 0.09 \cdot coll_i - 1.14\, cashH_i - 0.05 \ln(TA_i)$$

Once the firms are assigned their *safe* scores and ranked according to their values, a threshold value of the SAFE score above which we can define firms in a given level of aggregation as being credit constrained is calculated. The value of the threshold is time-varying and country-specific and is set so that the share of firms above this threshold at the country level is the same as the share of credit constrained firms for a given country-year reported in the SAFE survey. In a final step, the *safe* dummy variable for a given firm is assigned value 1 if the estimated SAFE score index is above the threshold, and 0 otherwise.

The resulting *safe* variable in the dataset reports the share of credit constrained firms in any given level of aggregation. In addition, the variable is also used as a conditional variable for joint distributions, from which we can learn how credit constrained firms differ from unconstrained firms.

The second indicator, *abconstr***,** constructed to detect whether a firm is affected by financial restrictions when planning its investments is closely related to the strand of the economic literature that suggests using "a-priori" classification of being constrained, based on firms' financial conditions.

For the CompNet dataset, the "a-priori" classification proposed by Ferrando and Ruggieri (2015) is applied. The advantage of this classification is that it takes into consideration a set of variables derived from the balance sheet and profit and loss accounts as well as their connection with different investment/financing scenarios. The various scenarios are based on

the interrelation of total investment, financing gap (defined as fixed investment plus the change in the net increase in working capital minus cash flow), financial debt and issuance of new shares in any given year.

Thus, the CompNet dummy variable *absconstr* takes the value 1 when a firm is classified as "absolutely credit constrained" and 0 otherwise. "Absolutely credit constrained" firms are identified as follows:

- firms with positive investment and with total investment higher than the current cash flow as well as a concurrent reduction of debt and capital;
- firms that, although disinvesting, have a positive financing gap.

Similar to the previous credit constrained indicator, the resulting *absconstr* variable in the dataset reports the share of absolutely credit constrained firms in any given level of aggregation. The variable is also used as a conditional variable for joint distributions, from which we can learn how absolutely credit constrained firms differ from unconstrained ones.

5.3.5 Indicators of Market Imperfection

This group of indicators is designed to capture product and labour market imperfections and is based on work by De Loecker and Warzynksi (2012).

Product Markup

CompNet calculates firm and time specific markups based on different gross output production function specifications by using the framework of De Loecker and Warzynski (2012). The associated markup formula writes:

$$(1) \qquad \mu_{it} = \alpha_{it}^{M} * \frac{P_{it}Q_{it}}{P_{it}^{M}M_{it}},$$

where μ_{it} denotes the markup, α_{it}^{M} is the output elasticity of intermediate inputs, and $\frac{P_{it}Q_{it}}{P_{it}^{M}M_{it}}$ is the inverse of the share of intermediate input expenditures in revenues.[30] We recover α_{it}^{M} from estimating a production function based on different aggregation levels, different functional form assumption and different factors of production. In particular, we estimate

[30] We rely on the intermediate input decision of the firms, since we are aware that different degrees of (in)flexibility of labour inputs across different countries might cause biased estimations of the markup parameters (for details please see De Loecker and Warzynski (2012) and De Loecker, Goldberg, Khandelwal, and Pavcnik (2016).

Cobb-Douglas and translog production functions, one time separately for all firms within a two-digit sector and one time separately for all firms within a macro-sector. Arguably, the most sophisticated version of our markup estimates is the one based on the translog production function estimated at the two-digit sector level. However, as in practise we face a trade-off between the number of observations that can be used to estimate consistent parameters and the number of variables or lags included in the regression, we also apply simpler forms of the production functions (i.e. Cobb-Douglas) and also specifications where we pool more firms during our estimation (i.e. at the macro-sector level). When using our markup estimates we also recommend having a look at the non-parametric competition indicators that we provide. These contain price-cost margins, Hirschman-Herfindahl indices, and profit margins.

Labour Market Power

Similar in spirit to Dobbelaere and Mairesse (2013), we measure labour market power, γ, by dividing the markup formulas from De Loecker and Warzynski (2012), based on firms labour input decisions with the corresponding markup formula for firms' intermediate input decision:

$$(2) \quad \mu^M = \theta^M * \frac{P_{it}Q_{it}}{z_{it}M_{it}}$$

$$(3) \quad \mu^L = \theta^L * \frac{P_{it}Q_{it}}{w_{it}L_{it}}$$

$$(4) \quad \frac{\mu^L}{\mu^M} = \gamma$$

where μ^X and θ^X respectively denote the markup based on the input decision of input $X = \{L, M\}$ and the output elasticity of input X. P_{it}, Q_{it}, z_{it}, w_{it}, M_{it}, L_{it} respectively are the output price, output quantity, unit cost for intermediates, wage, intermediate input quantity, labour quantity. For a detailed derivation of this parameter, we refer to the online appendix section of Mertens (2020).

5.3.6 Job Creation Rates (JCR) and Job Destruction Rates (JDR)

To analyse job flows at a given level of aggregation, we follow the seminal paper of Davis et al. (1996). The measures are based on the firm-level growth rate of employment, which is computed in the following way:

$$(5) \qquad X_{it} = 0.5 \cdot (E_{it} + E_{it-1}) \text{ and } g_{it} = \frac{(E_{it} - E_{it-1})}{X_{it}}$$

Where X_{it} is the firm average employment (E_{it} and E_{it-1} are the employment in current and previous time point for a particular firm) and g_{it} is the firm-level growth rate of employment. Since the growths rate incorporates both entry and exit, it also accounts for the creation and destruction respectively.

In particular, in the CompNet dataset, the job creation and destruction rates are estimated at the sector, macro-sector, nuts2, macro-sector-size-class and country levels. For example, at the two-digit sector level the growth rate has to be weighted by a firm weight in the following way:

$$(6) \qquad Firm\ weight = \frac{X_{it}}{X_{st}} \quad \text{and the weighted growth rate is} \quad Firm\ weight \cdot g_{it}$$

where X_{st} is the average employment for a particular sector. Therefore, at the sector level, the growth rate should be adjusted by the firm weight. Finally, the JCR and JDR are the sum of all positive and negative weighted growth rates respectively. We calculate JCR and JDR measures in terms of population equivalents (i.e. weighted versions) and sample data equivalents (i.e. unweighted versions).

5.4 Data Collection and Harmonization

CompNet works bilaterally with national statistical institutes, central banks or ministries in several European countries to create the CompNet dataset. This allows immediate feedback from and to data providers to solve any problem that may arise quickly and efficiently. There are several important concerns regarding firm-level data: confidentiality, the treatment of outliers or comparability of inputs. The following subsections elaborate on the way CompNet deals with these concerns.

5.4.1 Confidentiality

To ensure absolute confidentiality, the code created by the CompNet team is run by the data providers of CompNet themselves. This way, the CompNet team is never directly handling any confidential microdata at the firm-level, but only the anonymized and harmonized output delivered by the individual country teams. The code produces descriptive statistics and regression results at different levels of aggregation (while keeping the rich information of the underlying distributions) and ensures that the user of the final data will not be able to uniquely identify individual firms. The result is the micro-aggregated data provided in the CompNet dataset.

The CompNet team and the individual data providers work intensively together in compiling a high-quality dataset and each member institution is able to individually specify conditions to satisfy any national confidentiality regulations.

The CompNet code includes 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, which might make the identification of individual firms possible, the cell will be dropped. However, this dropped information is still accounted for in the total distribution to maintain a high level of representativeness. 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. These are the parameters which have been used by most of the data providers:

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

It should be noted that the comparability of all data points actually published is **not** affected.

5.4.2 Outlier Treatment

In the 7th Vintage of the CompNet dataset a new outlier routine was introduced. The outlier routine is based on four different procedures. Notably, we do not drop any firm observation, but rather replace outlier values in specific variables by missing values.

In a first step we clean the data from meaningless, mostly negative values in a set of variables (e.g. negative revenue). In the second part of the routine, we eliminate values in the labour variable for firms that exhibit extraordinary growth rates in the labour variable. Here, extraordinary growth is defined as a growth that violates the following condition:

$$\left(\frac{headcount(t)}{headcount(t-1)} - 1\right)\sqrt[3]{(headcount(t-1))} \leq 75$$

$$\left(\frac{headcount(t-1)}{headcount(t)} - 1\right)\sqrt[3]{(headcount(t))} \leq 75$$

In the third part of our routine, we clean trade values. If export values exceed turnover by more than a factor of 1.5, the trade information is replaced with missing values.Such cases likely occur when trade and balance sheet data are collected at different points in time.

Finally, we eliminate capital, turnover, intermediate input expenditure, labour cost, and labour values for the top two and bottom three percent values in the distribution of the ratios of turnover over labour, turnover over labour costs, turnover over capital, turnover over intermediate input expenditures, capital over labour, and intermediates over labour. Notably, we only replace variables by missing if they are flagged as outlier with respect to the ratios involving the respective variable. This means, we replace capital observations only with respect to outliers in turnover over capital and capital over labour ratios by missing, but not with respect to the ratio of turnover to labour.

In addition to this basic variable, we also apply the same outlier cleaning to R&D and energy expenditures for the ratios of turnover to R&D expenditures and turnover to energy expenditures.

5.4.3 Weighting Procedure

The 7th Vintage of the CompNet dataset uses a new weighting procedure which includes population weights derived from business registries to calculate the appropriate weights. Appropriate weighting (in terms of number of firms in a given sector and size class), is a first step to ensure that the data is representative for the underlying firm population. Since not all data providers of CompNet were able to provide population measures at the 2-digit level (sector) according to NACE Rev. 2, the 7th Vintage only uses weights at the 1-digit level (macro sector). However, the census data that have been provided already - are aggregated to the 1-digit level and combined with Eurostat's SBS population measures to provide the best possible coverage.

For the two different samples two slightly different approaches are employed:

a. **Full Sample**: 1-digit macro sector weights from Eurostat are used according to NACE Rev. 2 and make an extrapolation with the available aggregated census data.

b. **20e Sample**: 1-digit macro sector weights based on the aggregated census data are used according to NACE Rev. 2 and make an extrapolation with the Eurostat population measures.

The extrapolation will be done in both cases on the basis of growth rates from Eurostat and census data, given the assumption that, despite differing numbers of firms in the different sources, growth rates should be similar. We employ two different approaches because both have their weaknesses and their parallel use will help us to learn more about their influence.

To illustrate the general weighting procedure[31], let us define *x* as the variable we want to compute a descriptive statistic of and x_i with i=1,2,…*n* as the individual observation on *x* of firm i. The sample number of firms *n* is equal to *variable*_N in the output dataset. Then the individual weight v_i is defined as

$$v_i = \frac{firms\_y\_z}{m\_y\_z} \qquad (1)$$

[31] The Stata command *summarize* with option "aweights" is applied. For further details, please refer to https://www.stata.com/manuals13/rsummarize.pdf. One has to take into account that "aweights" is not strictly speaking correct. However, according to statalist (http://www.stata.com/support/faqs/statistics/weights-and-summary-statistics/),
they produce the correct population variance, mean and percentiles. This allows to use sum though pweights are not available.

with *firms_y_z* as the number of firms (of a certain year) of size class y and macro sector z in the total population and *m_y_z* as the number of firms in the sample with non-missing variables for x of the same size class, macro sector and year. The sum of the weights (= *variable*_sum_weights[32] in the output dataset) is then

$$V = \sum_{i=1}^{n} v_i \,. \tag{2}$$

Then the sum of the individual weights is normalized to *n* so the actual weight w_i is defined as

$$w_i = v_i \frac{n}{V} \,. \tag{3}$$

The weighted sample mean $\bar{x}$ can then be calculated as

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} w_i \, x_i \tag{4}$$

The weighted sample variance s^2 is

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} w_i \, (x_i - \overline{x})^2 \tag{5}$$

With the standard deviation as $\sqrt{s^2}$. The other moments follow the formula

$$m_\tau = \frac{1}{n-1} \sum_{i=1}^{n} w_i \, (x_i - \overline{x})^\tau \tag{6}$$

Consequently, weighted skewness is defined as

$$m_3 / \left(\sqrt[2]{s^2}\right)^3 \tag{7}$$

and the weighted kurtosis as

$$m_4 / (m_2)^2 . \tag{8}$$

Let x_i refer to the *x* in ascending order, and let w_i refer to the corresponding weights of x_i.

To calculate the weighted *pth* percentile x_p, define *P=np/100* and $W_i = \sum_{j=1}^{i} w_j$. Then one has to find the first index *i* for $W_i > P$.

$$x_p = \begin{cases} \frac{x_{i-1} + x_i}{2} & if \; W_{i-1} = P \\ x_i & otherwise \end{cases} \tag{9}$$

[32] By definition *V* should be equal to $\sum firms\_y\_z$ if *m_y_z* >0. However, this is only true at the country-, macro sector- and macro sector size class level. At the NUTS2 level as well as the two-digit sector level this would only hold in case of a perfect random sample, e.g. the firms of a certain macro sector are equally distributed across its two-digit sectors.

5.4.4 Data Sources

Table 16 Country Specific Data Sources

Country	Data Source	Acronym	Institution Responsible for Source	Data Provider	Firms Included in Dataset*	Source Specific Information
Belgium	microBACH (Bank for the Accounts of Companies Harmonized), ECCBSO (European Committee of Central Balance Sheet Data Offices)					
Croatia	Yearly financial statements of firms	FINA	Financial Agency Croatia	Croatian National Bank	NFC drawn from total economy	
Czech Republic	Annual report of economic units in selected production industries P5-01	P501	Statistics Czech Republic	Czech National Bank	NFC drawn from total economy	full coverage for firms with >50 employees, stratified survey for smaller firms
	Extrastat/Intrastat foreign trade transaction data	TRADE	Statistics Czech Republic		NFC drawn from total economy	
	Business Register	RES	Statistics Czech Republic		NFC drawn from total economy	
Denmark	Accounts statistics - non-agricultural industries	Acc. Stat.	Statistics Denmark	Central Bank of Denmark	NFC	
	General enterprise statistics	Gen. Stat.	Statistics Denmark		NFC drawn from total economy	
Finland	Structural business and financial statement statistics data	SBS	Statistics Finland	Statistics Finland	NFC drawn from total economy	

	International trade statistics data	ITS	Finnish Customs		NFC drawn from total economy	
France	Regime of Normal Real Profits	BRN	Statistics France	Statistics France	NFC drawn from total economy	Complementing sources with RSI. BRN covers large firms +788K
	Simplified Regime for the Self-Employed	RSI	Statistics France		NFC drawn from total economy	Complementing sources with BRN. RSI covers small firms below 788K
Germany	Official German Firm Data	AFID	Destatis	Federal Statistical Office of Germany and Federal Statistical Offices of the German Länder	NFC drawn from tot	Manufacturing sectors: only firms with more than 19 employees Other sectors: firms with at least 17,5K revenues
Hungary	Tax registry database of National Tax and Customs Administration	NAV	National Tax and Customs Administration	Central Bank Hungary	NFC drawn from total economy	Non-mandatory variables for tax-records are underreported. E.g. 30% of Firms do not report the number of employees.
	Business Registry	VR	Statistics Hungary and Central Bank of Hungary		NFC drawn from total economy	
	Export-Import data of Hungarian Enterprises	Külker	Statistics Hungary		NFC drawn from total economy	
Italy	microBACH (Bank for the Accounts of Companies Harmonized), ECCBSO (European Committee of Central Balance Sheet Data Offices)					
Lithuania	Statistical Survey on the Business Structure (Annual questionnaire F-01)	F01	Statistics Lithuania	Central Bank Lithuania	NFC drawn from total economy	

	Business Register	BR	Centre of Registers		NFC drawn from total economy	
	Customs, Customs declarations	CU	Customs of the Republic of Lithuania		NFC drawn from total economy	
Netherlands	Statistics finances of non-financial enterprises	SFO	Statistics Netherlands	Statistics Netherlands	NFC drawn from total economy	Full coverage for small firms (< 40 Mln balance sheet total); Large firms (> 40 Mln balance sheet totals) are surveyed
	Business register	ABR	Statistics Netherlands		NFC drawn from total economy	
Poland	Reports on revenues, costs, profit and outlays on fixed assets	F01	Statistics Poland	Central Bank Poland	NFC	Exclusion of firms with less than 10 employees
	Stat. financial report	F02	Statistics Poland		NFC	Exclusion of firms with less than 10 employees
Portugal	Integrated Business Account System	SCIE/IBAS	Statistics Portugal	GEE – Ministry of Economy Portugal	NFC	Full Coverage/Administrative Data
Romania	Balance sheet information on non-financial enterprises	Bal. Sheet	Ministry of Public finances	National Bank Romania	NFC drawn from total economy	
	Exports and imports of goods, firm-level data	TRADE	Statistics Romania		NFC drawn from total economy	
Slovakia	Annual report on production industries	Reports	Statistics Slovakia	National Bank of Slovakia	NFC drawn from total economy	Exclusion of firms with less than 20 employees

	Statistical register of organizations	Register	Statistics Slovakia		NFC drawn from total economy	Exclusion of firms with less than 20 employees
	Foreign trade statistics	Customs	Statistics Slovakia		NFC drawn from total economy	Exclusion of firms with less than 20 employees
Slovenia	Business Register	AJPES	Agency of the Republic of Slovenia for Public Legal Records and Related Services	Institute of Macroeconomic Analysis and Development of the Republic of Slovenia (IMAD)	Non Financial Companies (only companies included) drawn from total economy	full coverage for firms with ≥1 employee
	Trade data					Data only for Companies with ≥1 employee
	Financial data					Data only for Companies with ≥1 employee
Spain	microBACH (Bank for the Accounts of Companies Harmonized), ECCBSO (European Committee of Central Balance Sheet Data Offices)					
Sweden	Structured business statistics	SBS	Statistics Sweden	Statistics Sweden	NFC drawn from total economy	
	International trade in goods	ITG	Statistics Sweden		NFC drawn from total economy	
	Business register	BR	Statistics Sweden		NFC drawn from total economy	

5.4.5 Harmonization of Input Data

For the sake of improving comparability, the CompNet Team introduced a set of definitions for the input variables used for the creation of the CompNet Dataset Table 17.

Table 17 Raw Input Variables - Definitions

Variable	First Best	Second Best	Third Best
	Finance		
Fixed assets	Sub item of non-current assets (yearly average) consisting of capital (tangible fixed assets) + intangible fixed assets + other fixed assets (mainly financial long-term assets)	Fixed assets at a particular point in time	
Capital	Balance sheet item tangible fixed assets (yearly average), sub-item of fixed assets and non-current assets: only land, machinery, equipment, buildings and other durables; does not include long-term financial assets and intangible assets	Tangible fixed assets at a particular point in time	
Intangible fixed assets	Related balance sheet item intangible fixed assets (yearly average): acquired - not developed in-house - intellectual property (patents, licences, copyrights, trademarks) and goodwill	Intangible fixed assets at a particular point in time	
Other fixed assets	All fixed assets (yearly average), that could not be subsumed under tangible fixed assets or intangible fixed assets: contains mainly long-term financial assets	Other fixed assets at a particular point in time	
Current assets	Current assets (yearly average) are assets according to IAS 1.66: expected to be realised in the entity's normal operating cycle, held primarily for the purpose of trading. Sub-items are: accounts receivable, total inventories and other current assets (including cash and cash equivalents)	Current assets at a particular point in time	cash and cash equivalents + accounts receivable + inventories
Cash and cash equivalents	Balance sheet item cash and cash equivalent (yearly average), it is a sub-item of other current assets: value of a company's assets that are cash or can be converted into cash immediately.	Cash and cash equivalents at a particular point in time	
Total inventories	Inventories (yearly average) according to IAS 2.6: include assets held for sale in the ordinary course of business (finished goods), assets in the production process for sale and raw materials.	Total inventories at a particular point in time	
Accounts receivable	Related balance sheet item: accounts receivable (yearly average), sub-item of current assets	Accounts receivable at a particular point in time.	

Variable	First Best	Second Best	Third Best
Other current assets	All current assets (yearly average) that could not be subsumed under accounts receivables and inventories; contains for example cash and cash equivalent, prepaid expenses and accrued income.	Other current assets at a particular point in time	
Total assets	Total assets refer to the sum of current and fixed assets and match the sum of liabilities (current and non-current) + total shareholder funds (equity).	Total assets at a particular point in time.	
Total shareholder funds (equity)	Balance sheet item total shareholders' funds (yearly average): includes shares issued, retained earnings, additional paid-in capital, reserves, non-controlling interest; should be equal to total assets - liabilities	Shareholder funds (equity) at a particular point in time.	total shareholder funds
Non-current liabilities	Also called long-term liabilities (yearly average) in the balance sheet; includes all liabilities that are not due within the next 12 months. See also definition of current liabilities.	Non-current liabilities at a particular point in time	long term debt + provisions
Long-term debt	Sub-item of non-current liabilities: 1) loans (yearly average) due in more than 12 months. Includes bank loans, loans from affiliated companies, shareholder loans or loans from anyone else; 2) Bonds beyond 12 months + Convertible bonds beyond 12 months	Long term debt at a particular point in time	
Other non-current liabilities	All non-current liabilities (yearly average) that could not be classified as long-term debt: deferred income tax, provisions for pension plans etc. Should be equal to non-current liabilities minus long-term debt	Other noncurrent liabilities at a particular point in time	
Current liabilities	Current liabilities (yearly average) According to IAS 1.60: All other liabilities that not meet this definition are classified as **non-current.** Should be equal to short-term debt + accounts payable + other current liabilities.	Current liabilities at a particular point in time	short-term debt + accounts payable
Short-term debt	Sub-item of current liabilities: 1) loans (yearly average) to banks and other lenders due within less than 12 months; 2) Bonds + Convertible bonds	Short-term debt at a particular point in time	
Accounts payable	Related balance sheet item: accounts payable (yearly average), sub-item of current liabilities; accounts payable are a business to business agreement in which a customer can purchase goods on account (without paying cash up front), paying the supplier at a later date.	Accounts payable at a particular point in time.	
Other current liabilities	All current liabilities (yearly average) that could not be subsumed under short-term debt and accounts payable	Other current liabilities at a particular point in time	

Variable	First Best	Second Best	Third Best
Turnover	Turnover has to be valued at factor cost: gross sales revenues minus customer discounts, returns and allowances; excluding indirect taxes but including subsidies on products and production	Valued at market prices	
Labour cost	Gross wages and salaries paid to employees, other monetary or non-monetary expenses for employee benefits that could be attributed to the current accounting period	Total employee benefits expense (including pension payments to retired staff)	
Intermediate inputs	All expenses of the firm for products and services acquired valued at basic prices, i.e. excluding Non-VAT taxes on products but including subsidies on products.	Intermediate inputs valued at market prices	
R&D expenditures (New in CompNet)	Research and development (R&D) refers to the work a business conducts for the innovation, introduction and improvement of its products and procedures. R&D expenditures are operating expenses (not expenditures for purchasing R&D-related fixed assets like laboratory equipment) related to the firm's research and development.		
Operating profit/loss (EBIT)	Either conducted by IAS 1.92 EBIT (Earnings Before Interest and Taxes) according to the "cost of goods sold approach" or by IAS 1.91 EBIT according to the "nature of expense method"	revenues (turnover) - intermediate inputs - labour cost - depreciation	
Interest paid and financial charges	All interest payable on any borrowings, i.e. bonds, loans, convertible debt or lines of credit		
Depreciation	Includes only depreciation (ordinary or extraordinary) of fixed tangible assets, not depreciation/impairment of financial (non-current) assets or amortization of intangible fixed assets.	total depr. of fixed tangible assets + depr. on/amortization of intangible fixed assets + depr. of financial fixed assets	
Profits and losses before taxes	Earnings [from continuing operations] before [income] Taxes (EBT) = EBIT (see operating profit/loss) + financial revenue	Operating profit/loss- interest paid +	

Variable	First Best	Second Best	Third Best
	[e.g. interest received] - financial costs [e.g. interest paid] +/- equity in earnings of subsidiaries	interest received - interest paid and financial charges	
Cash flow (from profit/loss statement)	Cash flow from operating activities according to IAS 7 (before taxes and interest paid), indirect method: Profit/loss before interest and income taxes (EBIT) + depreciation + impairment of inventories and receivables - increase in inventories, receivables + increase in liabilities - decrease in liabilities	Complete (gross) cash flow from operating activities before interests and taxes	Operating profit + depreciation
Dividends	Dividend payments to shareholders as reported in the statement of changes in equity or the statement of cash flows according to IAS 1.137		
Trade			
Export value	Exports valued at factor cost: Nominal export turnover (see definition of turnover) excluding indirect taxes, tariffs etc., but including subsidies on products and production	Valued at market prices: including	
Exports to extra-EU	Valued at factor costs: Nominal export turnover (adjusted or unadjusted exports) outside EU (see definition of exports and turnover) excluding indirect taxes, tariffs etc., but including subsidies on products and production.	Valued at market prices	
Exports to intra-EU (new in CompNet)	Valued at factor costs: Nominal export turnover (adjusted or unadjusted exports) within EU (see definition of exports and turnover) excluding indirect taxes, tariffs etc., but including subsidies on products and production.	Valued at market prices	
Import value	Expenses for imported products and services acquired valued at basic prices, i.e. excluding Non-VAT taxes or tariffs on products but including subsidies on products.	Imports valued at market prices	
Imports from extra-EU (new in CompNet)	Expenses for imported products and services acquired from outside the EU valued at basic prices, i.e. excluding Non-VAT taxes or tariffs on products but including subsidies on products.	Valued at market prices	
Imports from intra-EU (new in CompNet)	Expenses for imported products and services acquired from the EU valued at basic prices, i.e. excluding Non-VAT taxes or tariffs on products but including subsidies on products.	Valued at market prices	
Other			
Sector	Two-digit division number according to NACE Rev. 2		
nuts2	Four-digit code (combination of country and region) according to REGULATION (EC) No 1059/2003, Annex 1		

Variable	First Best	Second Best	Third Best
Number of firms in the population in a given sector and size-class	Number of firms in the total population in a given NACE 2 2-digit sector and size class; size classes according to the number of employees		
Firm's birth year	The year of the creation of the legal unit		
Firm's exit year	The year when the firm has been deleted from the business register.		
Foreign ownership	Dummy that equals one if more than 50% of the firm's shares are controlled by foreign owners and 0 otherwise.		
Labour	Headcounts of the number of employees (yearly average) with employed shareholders/owners excluded	Headcounts at a certain date	Full time equivalent
Legal form	Categorical variable taking the values: 1 = limited liability companies and limited liability partnerships; 2 = Sole proprietorship; 3 = unlimited liability partnerships; 4 = Co-operative societies; 5 = Non-profit making bodies; 6 = other legal forms		
Public or non-profit enterprise	Categorical variable taking the values: 1 = more than 50% of the firm`s shares are held by the government directly or indirectly by firms/associations controlled by the government; 2 = more than 50% of the firm`s shares are held by non-profit organization(s) or indirectly by firms/associations controlled by non-profit organizations; 3 = government and non-profit organization(s) hold together more than 50% of the shares of the firm directly or indirectly; 4 = otherwise (private firm)		
Share of skilled labour	Share of employees having post-secondary (tertiary) education. Tertiary education is the educational level following the completion of a school providing a secondary education. It includes universities as well as trade schools, colleges and vocational training.		

Table 18 Country specific Definitions of Input Variables

Variable / Country	BE	HR	CZ	D	DK	FI	FR	HU	IT	LT	NL	PL	PT	RO	SK	ES	SE	CH
Fixed assets	2	2	1	2	2	2	2	2	2	2	Unknown	1	2	2	2	2	2	2
Capital	2	2	1	2	2	2	2	2	2	2	Unknown	1	2	2	2	2	2	2
Intangible fixed assets	2	2	1	2	2	2	2	2	2	2	Unknown	1	2	3	2	2	2	2
Other fixed assets	2	2	1	2	2	2	2	2	2	2	Unknown	1	2	2	0	2	2	2
Current assets	Other	2	1	2	2	2	2	2	Other	2	Unknown	1	2	2	2	Other	2	2
Cash and cash equivalents	Other	2	Other	2	2	2	2	2	Other	2	Unknown	1	2	2	0	Other	2	2
Total inventories	2	2	1	0	2	2	2	2	2	2	Unknown	1	2	2	2	2	2	2
Accounts receivable	Other	2	1	0	2	2	2	2	Other	2	Unknown	1	2	2	2	Other	2	2
Other current assets	Other	2	1	2	2	2	2	2	Other	2	Unknown	1	2	3	0	Other	2	2
Total assets	2	2	1	2	2	2	2	2	2	2	Unknown	1	2	2	2	2	2	2
Total shareholder funds (equity)	Other	2	1	2	2	2	2	2	Other	2	Unknown	1	2	2	2	Other	2	2
Non-current liabilities	Other	2	1	2	2	2	2	2	Other	2	Unknown	1	2	3	0	Other	2	2
Long-term debt	Other	2	1	2	2	0	2	2	Other	2	Unknown	1	2	3	2	Other	2	2
Other non-current liabilities	Other	2	1	2	2	0	2	2	Other	2	Unknown	Other	2	3	0	Other	2	2
Current liabilities	Other	2	1	2	2	2	0	2	Other	2	Unknown	1	2	2	2	Other	2	2
Short-term debt	Other	2	1	0	2	0	2	2	Other	2	Unknown	Other	2	3	2	Other	2	2
Accounts payable	Other	2	1	0	2	2	2	2	Other	2	Unknown	1	2	2	2	Other	2	2
Other current liabilities	Other	2	1	2	0	0	2	2	Other	2	Unknown	Other	2	0	0	Other	2	2
Turnover	Other	2	1	1	Other	1	1	1	Other	1	Unknown	1	Other	1	1	Other	1	2
Labour cost	Other	2	1	2	Other	1	1	1	Other	1	Unknown	1	1	1	1	Other	1	2
Intermediate inputs	Other	1	1		Other	1	1	1	Other	2	Unknown	Other	1	2	1	Other	1	1
R&D expenditures	0	1	0		0	0	0	0	Other	0	0	Other	Unknown	1	0	0	0	0
Operating profit/loss (EBIT)	Other	2	2	1	Other	2	2	1	1	1	Unknown	Other	1	1	1	Other	1	1

Variable / Country	BE	HR	CZ	D	DK	FI	FR	HU	IT	LT	NL	PL	PT	RO	SK	ES	SE	CH
Interest paid and financial charges	1	1	1		Other	1	1	1	Other	1	Unknown	Other	1	1	1	1	1	1
Depreciation	Other	1	2	1[a]	Other	2	1	1	Other	1	Unknown	Other	1	3	1	2	1	1
Profits and losses before taxes	Other	2	1	1	2	2	1	1	Other	2	Unknown	Other	1	2	2	Other	1	2
Cash flow (from profit/loss statement)	Other	3	3	1	2	2	3	1	0	1	Unknown	3	1	3	2	Other	1	3
Dividends	0	0	1	0	Other	1	1	1	0	1	Unknown	0	Unknown	0	1	0	1	1
Export value	0	2	1	1	2	1	1	1	0	1	Unknown	Other	Other	3	1	0	1	0
Exports to extra-EU	0	0	1	0	0	1	1	0	0	1	Unknown	0	Other	3	2	0	1	0
Exports to intra-EU	0	0	1	0	0	1	1	0	0	1	Unknown	0	Other	3	2	0	1	0
Import value	0	2	1	0	2	1	1	0	0	1	Unknown	Other	Other	3	2	0	1	0
Imports from extra-EU	0	0	1	0	0	1	1	0	0	1	Unknown	0	Other	3	2	0	1	0
Imports from intra-EU	0	0	1	0	0	1	1	0	0	1	Unknown	0	Other	3	2	0	1	0
Sector	1	1	1	1	1	1	1	1	1	1	Unknown	Other	Other	1	1	1	1	1
nuts2	1	1	1	Other: NUTS1	1	1	1	1	1	1	Unknown	Other	Other	1	1	1	1	1
Firm's birth year	Other	1	1	1	1	0	1	1	Other	1	Unknown	0	Unknown	1	1	Other	1	0
Firm's exit year	1	1	0	0	1	0	0	1	1	0	Unknown	0	Unknown	0	1	1	1	0
Foreign ownership	0	1	1	0	0	0	0	1	0	0	Unknown	Other	Other	0	1	0	1	0
Labour	1	2	1	2	Other	1	1	1	Other	1	3	Other	Other	2	1	1	1	2
Legal form	1	1	1	0	1	1	0	1	1	1	Unknown	0	Other	0	1	1	1	0
Public ownership	0	1	1	0	0	1	0	2	0	0	Unknown	Other	Other	0	1	0	Other	0
Share of skilled labour	0	0	0	0	0	1	0	0	0	0	Unknown	0	Other	0	0	0	1	0

Notes: 0 "not available" – 1 "first-best definition"according to table 17. - 2 "second-best definition" – 3 "third-best definition"[33]

[a] Second best definition (2) for manufacturing sectors

5.4.6 List of Sectors

Table 19 List of Sectors Included in the 7th Vintage

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
C	1	Manufacturing	10	Manufacture of food products
			11	Manufacture of beverages
			12	Manufacture of tobacco products
			13	Manufacture of textiles
			14	Manufacture of wearing apparel
			15	Manufacture of leather and related products
			16	Manufacture of wood and of products of wood and cork, except furniture
			17	Manufacture of paper and paper products
			18	Printing and reproduction of recorded media
			19	Manufacture of coke and refined petroleum products
			20	Manufacture of chemicals and chemical products
			21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
			22	Manufacture of rubber and plastic products
			23	Manufacture of other non-metallic mineral products
			24	Manufacture of basic metals

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
C	1	Manufacturing	25	Manufacture of fabricated metal products, except machinery and equipment
			26	Manufacture of computer, electronic and optical products
			27	Manufacture of electrical equipment
			28	Manufacture of machinery and equipment n
			29	Manufacture of motor vehicles, trailers and semitrailers
			30	Manufacture of other transport equipment
			31	Manufacture of furniture
			32	Other manufacturing
			33	Repair and installation of machinery and equipment
F	2	Construction	41	Construction of buildings
			42	Civil engineering
			43	Specialised construction activities
G	3	Wholesale and retail trade; repair of motor vehicles and motorcycles	45	Wholesale and retail trade and repair of motor vehicles and motorcycles
			46	Wholesale trade, except of motor vehicles and motorcycles
			47	Retail trade, except of motor vehicles and motorcycles
H	4	Transportation and storage	49	Land transport and transport via pipelines
			50	Water transport
			51	Air transport

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
			52	Warehousing and support activities for transportation
			53	Postal and courier activities
I	5	Accommodation and food service activities	55	Accommodation
			56	Food and beverage service activities
J	6	Information and communication	58	Publishing activities
			59	Motion picture, video and television program production, sound recording and music publishing
			60	Programming and broadcasting activities
			61	Telecommunications
			62	Computer programming, consultancy and related activities
			63	Information service activities
L	7	Real Estate activities	68	Real estate activities
M	8	Professional scientific and technical activities	69	Legal and accounting activities
			70	Activities of head offices; management consultancy activities
			71	Architectural and engineering activities; technical testing and analysis
			72	Scientific research and development
			73	Advertising and market research
			74	Other professional, scientific and technical activities
			75	Veterinary activities

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
N	9	Administrative and support service activities	77	Rental and leasing activities
			78	Employment activities
			79	Travel agency, tour operator and other reservation service and related activities
N			80	Security and investigation activities
			81	Services to buildings and landscape activities
			82	Office administrative, office support and other business support activities

5.4.7 Coverage Ratios

Table 20 Coverage Ratios Country Preclean Sample vs Eurostat[34]

		All sample				20e sample		
Country	**Year**	Number Of firms	Employment	Revenue	Value added	Number Of firms	Revenue	Value added
Belgium	2010	18.46	73.08	49.36	83.59	73.15	62.35	108.19
	2011	19.47	75.98	50.87	84.60	74.82	63.31	111.07
	2012	19.34	78.20	51.51	84.82	79.15	64.36	113.29
	2013	19.87	77.64	50.18	81.90	81.06	62.47	108.38
	2014	19.15	85.06	51.03	83.82	83.81	64.72	112.18
	2015	18.72	81.37	51.66	87.14	83.39	65.86	117.32
	2016	18.66	80.59	51.69	83.09	84.17	65.82	112.94
	2017	17.64	82.04	53.49	.	84.09	66.34	.
Croatia	2010	35.55	78.60	124.06	201.96	85.39	128.55	201.65
	2011	38.36	80.41	126.77	205.30	86.52	130.95	204.07
	2012	39.14	80.27	126.30	212.12	86.20	129.75	212.92
	2013	41.25	80.09	124.48	198.40	86.27	129.01	201.22
	2014	42.52	80.17	125.43	201.65	87.26	133.88	208.21
	2015	43.84	81.80	127.66	208.86	87.14	135.70	219.68
	2016	47.73	86.04	129.54	212.43	89.89	133.37	.
	2017	50.28	86.12	127.35	.	90.66	.	.
Czech R.	2010	2.62	63.30	98.05	92.04	66.70	121.99	115.35
	2011	2.60	65.12	100.35	93.36	69.25	124.13	117.83
	2012	1.64	57.18	94.76	85.77	44.77	117.25	110.61
	2013	1.57	55.48	91.57	84.24	42.83	113.06	107.88
	2014	1.71	59.21	98.47	88.76	46.90	120.73	113.60
	2015	1.74	60.00	100.78	91.67	47.25	123.12	117.63
	2016	1.74	60.04	100.96	92.61	47.49	125.71	118.59
	2017	1.79	1.79	60.94	101.21	47.23	124.91	.
Denmark	2010	85.89	84.58	56.35	171.66	.	.	.

[34] Missing values due to unavailability of the Eurostat data.

		All sample				20e sample		
Country	**Year**	Number Of firms	Employment	Revenue	Value added	Number Of firms	Revenue	Value added
	2011	85.62	80.03	60.02	177.71	.	.	.
	2012	85.58	80.02	59.43	181.43	69.83	60.79	193.55
	2013	85.36	79.78	59.05	176.04	70.03	.	.
	2014	85.11	80.66	60.46	179.90	71.17	.	.
	2015	84.78	80.19	61.48	175.38	71.54		
	2016	84.46	80.39	61.85	.	70.08	.	.
France	2010	41.69	86.50	74.46	156.76	112.48	79.83	176.87
	2011	40.69	86.21	74.79	159.04	110.79	80.32	180.21
	2012	36.25	88.79	74.24	162.02	.	.	.
	2013	34.57	90.33	74.68	162.51	.	.	.
	2014	32.08	88.87	75.77	162.57	.	.	.
	2015	35.33	93.07	76.44	162.28	115.35	81.17	177.21
	2016	33.22	86.67	77.29	.	112.13	82.27	.
Finland	2010	55.98	97.15	73.23	65.70	.	.	.
	2011	53.25	97.18	72.87	64.64	93.30	74.82	67.10
	2012	56.38	97.11	73.02	66.68	93.86	74.79	68.37
	2013	53.62	97.53	73.03	67.07	110.37	76.04	71.52
	2014	53.92	97.49	73.45	64.77	93.67	75.16	67.86
	2015	56.63	97.57	74.01	65.74	93.82	76.28	69.23
	2016	56.78	97.81	74.10	63.90	94.35	76.22	66.97
Germany	2010							
	2011							
	2012							
	2013							
	2014							
	2015							
	2016							
Hungary	2010							

		All sample				20e sample		
Country	**Year**	Number Of firms	Employment	Revenue	Value added	Number Of firms	Revenue	Value added
	2011							
	2012							
	2013							
	2014							
	2015							
	2016							
Italy	2010	11.31	72.85	62.23	121.76	.	.	.
	2011	11.66	76.13	65.33	126.52	.	.	.
	2012	11.71	75.09	61.82	128.08	.	.	.
	2013	11.91	77.01	63.14	130.32	84.73	73.53	163.95
	2014	11.77	77.35	63.46	129.73	.	.	.
	2015	11.70	76.38	63.53	128.48	83.82	73.65	160.26
	2016	11.69	76.48	65.57	.	84.21	76.67	.
Poland	2010	.	.	.	.	69.13	131.70	127.78
	2011	.	.	.	.	67.53	131.71	127.74
	2012	.	.	.	.	67.85	130.31	126.59
	2013	.	.	.	.	68.36	131.21	135.07
	2014	.	.	.	.	68.99	134.93	140.25
	2015	.	.	.	.	69.59	135.63	139.70
	2016	.	.	.	.	73.03	146.38	148.89
	2017	.	.	.	.	70.93	144.92	.
Portugal	2010	36.32	95.08	104.53	260.44	96.91	.	.
	2011	37.42	95.49	103.51	264.28	96.91	.	.
	2012	38.33	95.94	102.33	266.05	97.02	.	.
	2013	38.96	96.43	102.15	268.12	97.09	.	.
	2014	39.25	96.83	102.96	265.95	97.21	.	.
	2015	38.85	97.14	103.99	263.14	97.32	.	.
	2016	38.45	97.29	105.07	260.40	97.46	.	.
	2017	38.18	97.62	105.31	.	97.48	.	.

		All sample				20e sample		
Country	**Year**	Number Of firms	Employment	Revenue	Value added	Number Of firms	Revenue	Value added
Romania	2010	72.31	95.85	167.25	90.91	92.53	169.29	99.60
	2011	75.64	128.14	167.70	92.28	92.44	169.11	102.43
	2012	70.82	93.72	165.11	92.22	91.22	167.92	121.53
	2013	68.63	88.95	166.89	91.27	92.11	170.07	.
	2014	65.73	87.35	165.00	94.26	90.93	169.25	.
	2015	67.68	88.31	168.58	110.66	91.05	172.40	.
	2016	68.65	87.42	170.69	.	90.46	174.25	.
Slovakia	2010	1.18	59.76	108.24	94.96	71.61	150.44	148.44
	2011	1.27	61.75	110.32	96.75	69.97	148.21	146.06
	2012	1.22	63.24	110.88	97.84	75.20	149.37	151.73
	2013	1.36	62.20	114.03	109.44	83.67	152.62	162.34
	2014	1.29	62.09	113.49	115.72	88.38	157.85	161.34
	2015	1.31	60.92	115.84	117.35	88.86	160.32	159.47
	2016	1.41	63.34	115.09	118.11	88.04	160.70	163.74
	2017	1.40	62.72	114.61	.	90.33	160.15	.
Slovenia	2010	26.49	78.87	107.44	101.60	82.59	114.78	111.10
	2011	25.88	79.46	106.21	102.73	82.28	113.13	112.79
	2012	23.32	78.85	103.80	105.71	83.07	110.80	116.94
	2013	22.54	78.24	104.64	105.75	80.98	111.41	116.93
	2014	23.62	79.03	105.70	105.40	82.69	112.92	116.14
	2015	24.03	79.92	106.26	105.15	83.78	112.67	116.64
	2016	24.34	81.09	107.79	104.48	85.50	114.48	115.85
	2017	24.48	81.89	106.18	103.14	85.58	112.30	114.86
Spain	2010	14.89	46.86	47.03	81.82	41.42	55.11	99.23
	2011	15.66	49.72	48.53	84.93	44.74	55.91	100.95
	2012	15.44	50.18	48.68	86.71	45.29	55.82	102.95
	2013	14.97	52.91	51.10	91.18	47.50	58.48	108.45
	2014	16.62	56.67	54.94	97.43	52.07	62.07	113.02
	2015	15.47	53.83	53.21	93.63	50.14	60.84	110.09

		All sample				20e sample		
Country	**Year**	Number Of firms	Employment	Revenue	Value added	Number Of firms	Revenue	Value added
	2016	12.97	47.87	48.92	86.76	41.95	56.36	100.91
	2017	13.90	49.83	50.86	.	44.44	57.65	.
Sweden	2010	23.50	85.01	65.57	184.30	88.50	69.70	197.79
	2011	22.83	84.40	65.58	186.93	88.15	70.30	205.02
	2012	22.54	83.62	62.49	185.87	87.16	65.55	199.26
	2013	22.76	83.65	62.98	185.20	87.66	67.49	.
	2014	22.77	83.98	62.43	181.16	88.07	.	.
	2015	22.38	83.96	53.27	178.66	87.79	.	.
	2016	21.99	83.48	63.44		86.70	.	.
Switzerland	2010	5.85	.	.	49.27	.	.	.
	2011	6.43	.	20.11	46.25	.	.	.
	2012	6.71	.	19.65	46.67	.	.	.
	2013	6.30	.	18.24	46.02	35.04	.	.
	2014	6.25	.	19.63	46.83	34.61	.	.
	2015	6.59	.	22.34	49.13	35.45	.	.
	2016	6.49	.	25.99	49.34	.	.	.
	2017	5.82	.	20.54	.	.	.	.

5.5 Validation

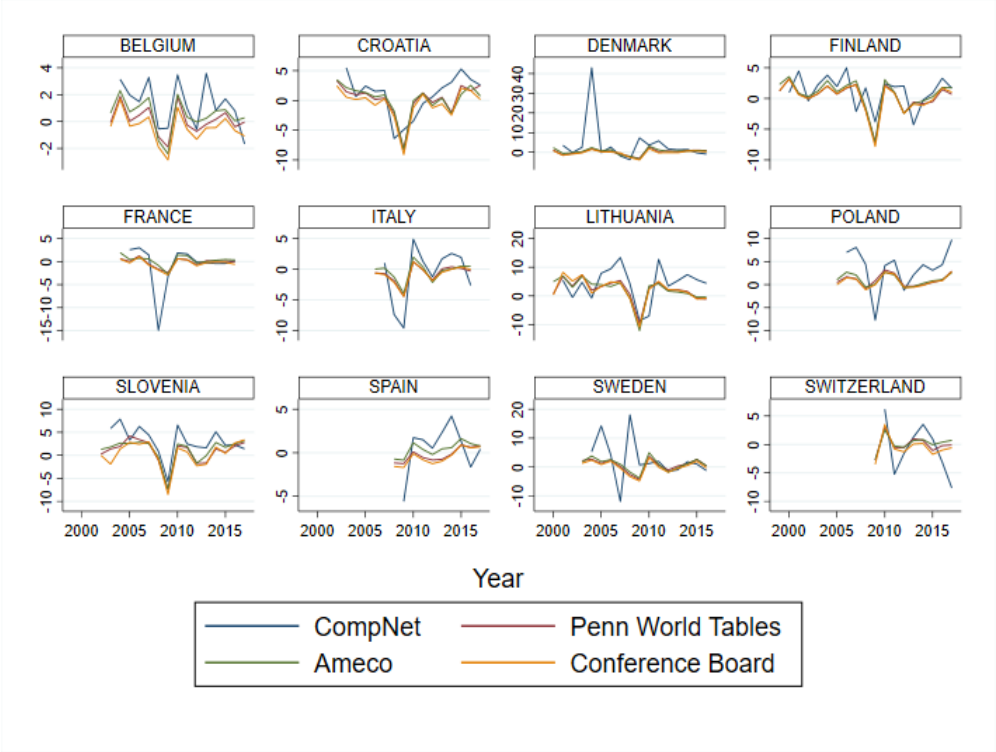
In this section we provide a corroboration exercise validating selected indicators that have been used in the CompNet Firm Productivity Report (Compnet (2020)). In order to do so, we compare the evolution of some indicators from our dataset with official sources, such as Eurostat, AMECO or the Survey on the Access to Finance of Enterprises. This is not the only validation exercise that we perform. We developed a Stata routine that systematically detects outliers and impossible values - e.g. ratios that exceeds 1 or negative values for stock values that should be positive – which are subsequently verified and contextualized manually. This routine is ran on all the indicators we have in our dataset. Therefore, the type of exercise we present here is deeper and to be intended as a validation background for the CompNet Firm Productivity Report.

5.5.1 Productivity Measures

Productivity is the key determinant of welfare in most economic models and the core driver of competitiveness. Over the last decades, however, aggregate productivity growth has slowed down in developed economies[35]. On the micro-level, studies document pronounced and persistent differences in firm productivity even within relatively narrowly defined industries (Syverson 2004, Hsieh and Klenow 2009). To deepen our understanding of the micro-level determinants of productivity that generate the patterns we observe on the aggregate level, CompNet offers a wide variety of micro-aggregated productivity estimators and indicators. We provide a non-parametric labour productivity and Solow residual using fixed weights. In addition to such productivity indicator derived from macroeconomic concepts, we also estimate one and two-digit sectoral level production functions to determine firm-level TFP. Specifically, we follow the Wooldridge (2009) methodology to control for the well-known simultaneity bias, which combines computational simplicity and wide adoption. Using this technique, we estimate value added and gross output production functions at the sector and macro-sector level and store all resulting firm-level TFP estimates.

[35] A more detailed discussion can be found in the CompNet Firm Productivity Report.

Figure 3 Productivity growth, CompNet versus Conference Board, Ameco AMECO and Penn World Tables

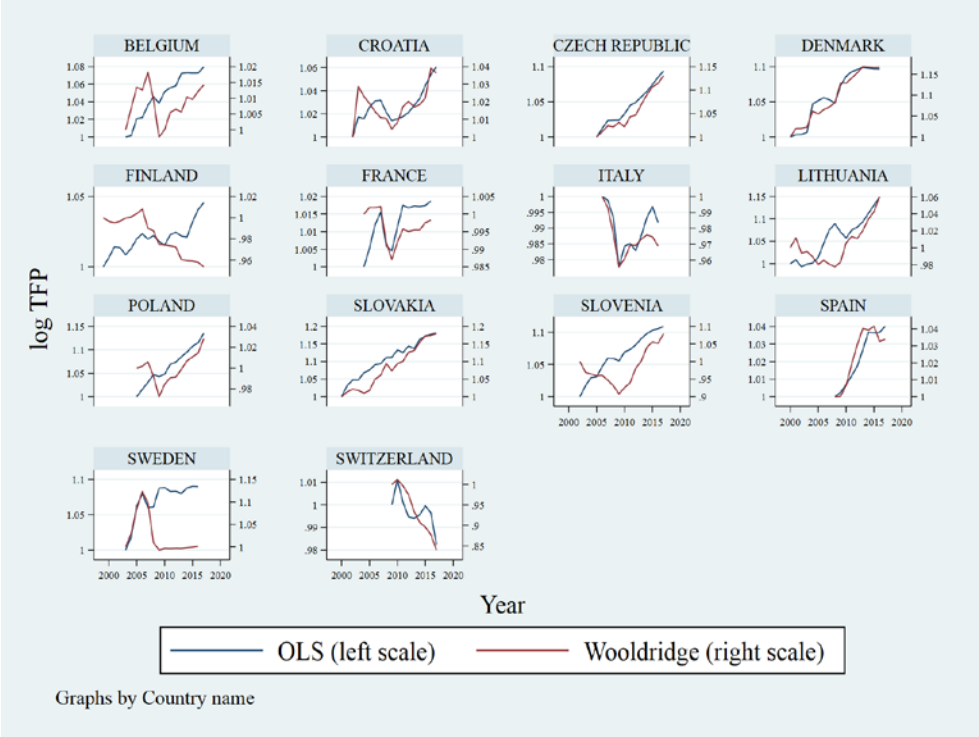


Sources: 7th vintage of CompNet, 20e sample, AMECO, Conference Board and Penn World Tables.

Note: CompNet TFP data refer to firms with employees operating in the non-financial business sector. TFP in CompNet is calculated as the ratio of a firm's real value added to its inputs labour and capital, weighted 2/3 and 1/3 respectively. We show the growth rate of the mean in the business sector. Conference Board and AMECO refer to the whole economy and calculate TFP as difference between aggregate real output growth and aggregate input growth, as derived from the national accounts (Solow residual).

The figure above depicts productivity growth by country over the span from 2000 to 2016. Despite differences in the concept and measurability of indicators, and being more volatile, the CompNet indicator displays the same trends as the indicators derived from aggregate data, as well as common reactions to macroeconomic events (e.g. the financial crisis in 2007-08).

Figure 4 Comparison of Log Total Factor Productivity (OLS/Wooldridge)



Note: CompNet TFP data refer to firms with employees operating in the non-financial business sector. TFP in CompNet is calculated based on a firm's real revenue using (i) the Ordinary Least Squares method (OLS) and (ii) the Wooldridge approach.

The figure above compares TFP derived from the OLS and the Wooldridge approach, respectively, where the latter adjusts for endogeneity of firm input choices. For most countries, the TFP measures follow the same trend while opposite trends are found for Finland and, preceding 2009, for Lithuania, Slovenia and Poland.

5.5.2 Approaches to measure allocative efficiency in the CompNet database

Dispersion of marginal products (Hsieh & Klenow 2009))

Perhaps the most influential work of the recent stream of research in the field of allocative efficiency constitutes the article by Hsieh & Klenow (2009). Those authors provide a simple model showing how dispersion in marginal products of production factors have a negative impact on aggregate productivity. The basic intuition behind this approach is that an aggregate output-maximizing planner would allocate production factors to firms where they generate the highest marginal product, leading in the long-run to an equalization of marginal products across firms. Any deviation from this first-best allocative efficient scenario indicates the existing of allocative inefficiencies.[36]

[36] While the approach is widely applied in the literature, it relies on a set of strong assumptions, particularly when it comes to quantifying the exact aggregate productivity losses from dispersion in marginal products of input factors (for discussion see Haltiwanger et al. (2018)).

While we do not provide the exact quantification of aggregate productivity losses due to misallocation using the Hsieh & Klenow (2009) approach, the CompNet database provides readily available statistics on the dispersion (i.e. standard deviation) of marginal products in various aggregation levels (e.g. sector-level, country-level, etc.). Notably, in contrast to the original framework of Hsieh & Klenow (2009), the marginal products in the CompNet data are based on a (semi-)parametric estimation of firms' production function. For details on the associated estimation methodology, we refer to the User-Guide for the CompNet database.

Lost-output gaps (Petrin & Sivadasan (2013))

Similar to Hsieh & Klenow (2009), Petrin & Sivadasan (2013) calculate lost-output gaps as a measure of allocative inefficiency based on wedges between the value of the marginal products of input factors and observed unit input costs. Formally, for the labour input this implies:

$$GAPL_{it} = w_{it} - VMPL_{it}, \tag{4.1}$$

where w_{it} denotes the wage and $VMPL_{it}$ is the value of the marginal products of labour (i.e. the marginal products of labour evaluated at output prices). Holding input costs constant, these wedges give a monetary measure of the firm-level output change. As Petrin & Sivadasan (2013) show, in a basic growth accounting framework, the average across all *absolute values* of these firm-level wedges equals the counterfactual aggregate output gain from reallocating one unit of the input factor (here labour) across all firms in the optimal direction (i.e. the direction that increases aggregate output).

In the CompNet database, we follow Petrin & Sivadasan (2013) and calculate firm-level absolute wedges between wages and the value of marginal products of labour. For other input factors, we do not apply this procedure, as we do not have information on unit input costs for other input factors. Again, marginal products are calculated from a (semi-)parametric estimation of firms' production function.

Covariance between size and productivity (Olley & Pakes (1996))

An alternative measure of allocative efficiency applied in the literature is based on the covariance between size and productivity. Olley & Pakes (1996) show that aggregate productivity, which is a weighted average, of firm-level productivity can be decomposed into the unweighted average of firm-level productivity and the covariance between the weight of economic productivity (i.e. size or economic importance of the firm) and firm-level productivity:

$$\Omega_{jt} = \sum s_{it}\omega_{it} = \bar{\omega}_{jt} + cov_{jt}(s_{it}, \omega_{it}), \tag{4.2}$$

where s_{it} is the share of economic activity. Ω_{ij}, $\bar{\omega}_{jt}$, and cov_{jt} denote aggregate productivity, average firm productivity and the covariance between firm-level size and productivity for aggregation level j, respectively. ω_{it} is firm-level productivity. $cov_{jt}(.)$ measures the extent to which more productive firms are larger. Under the premise that it is desirable that firms that are more productive should possess larger market shares (i.e. should be larger and should employ a larger share of the available input factors), higher values of $cov_{jt}(.)$ indicate a higher level of allocative efficiency. Under this logic, changes in $cov_{jt}(.)$ reflect changes in the allocative efficiency or between-firm productivity within aggregation level j. In constrast, changes in $\bar{\omega}_{jt}$ reflect changes in within-firm productivity.[37]

In the CompNet database, we provide this decomposition for various variables, including productivity measures. When using total factor productivity measures, we apply output weights as weights of firms' economic activity, while we rely on headcount-weights when using labour productivity.

5.5.3 Labour indicators

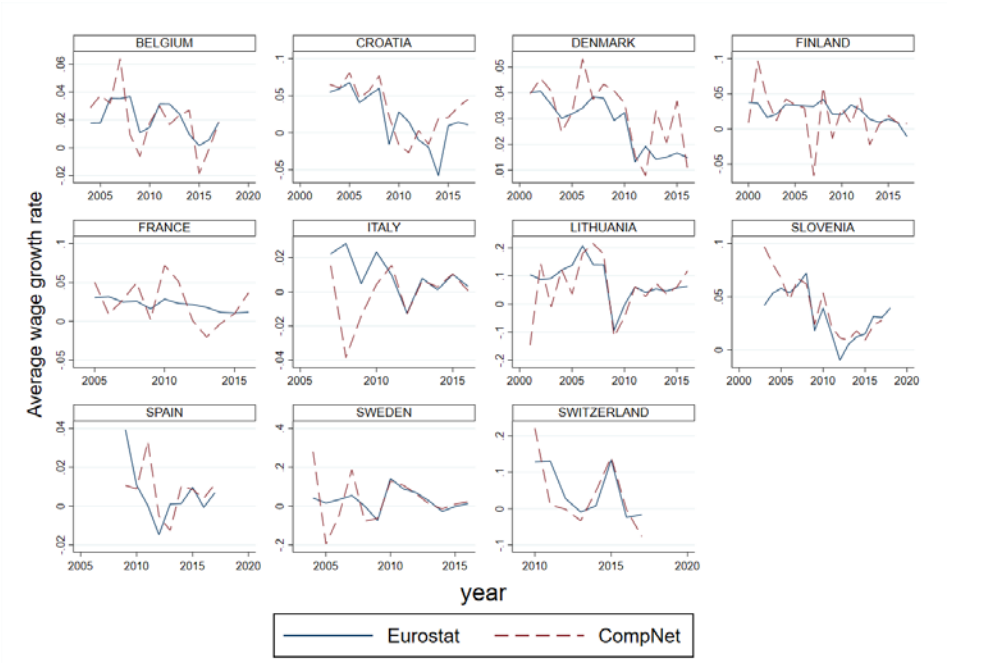
Definition, validation and granularity

CompNet collects data on firm's average nominal labour cost per employee (avg_wage). This indicator is computed as gross wages and salaries plus overtime payments and employers' social security contributions, divided by the total number of employees in the firm (average of the year, normally in FTE). It is also possible to compute the average real labour cost per employee from the CompNet dataset.

To get a sense of the accuracy of the data, Figure 5 shows the nominal average wage growth rate for all available countries in CompNet versus the annual growth rate of the average compensation per employee provided by Eurostat. Generally, the wage growth rates over follow the same time trend for the CompNet and Eurostat data. However, the CompNet growth rates are more volatile than the Eurostat data.

[37] This basic decomposition abstracts from exit and entry dynamics. For an extension, properly separating changes in aggregate productivity due to firm entry and exit, please see Melitz & Polanec (2015).

Figure 5 Annual growth rate of nominal average wage rates by country

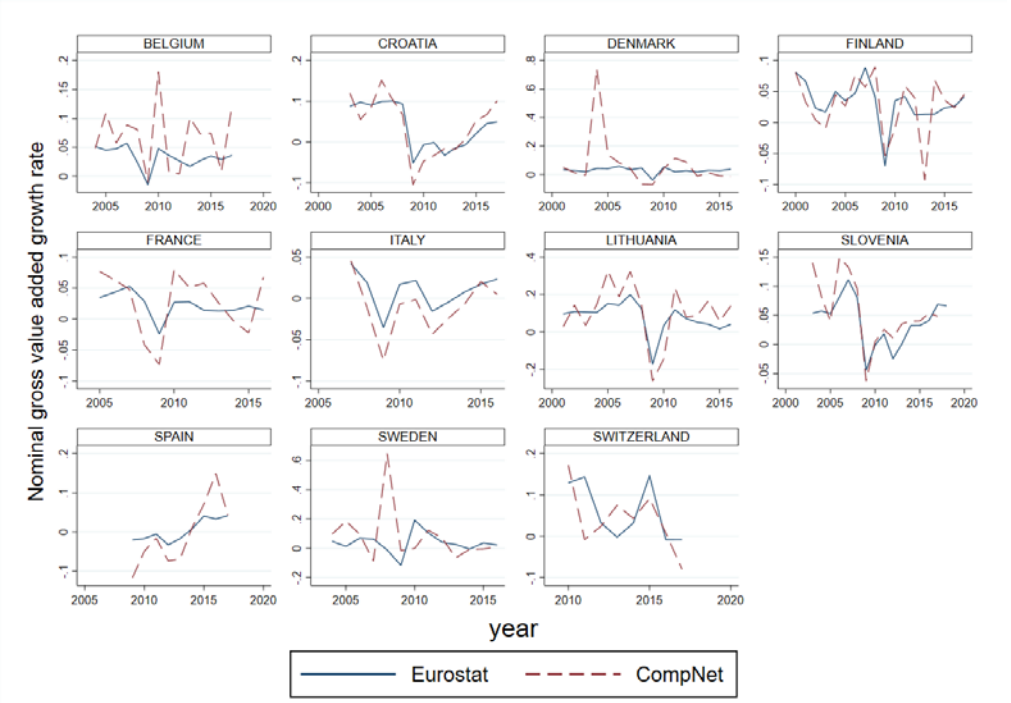


Sources: CompNet 7th Vintage database (full sample, country level) and Eurostat nama_10_lp_ulc (compensation per employee)

Note: Average wage in Compnet = (mean labour cost * sum weights labour cost)/(mean labour * sum weights labour)

Similar results can be found for the development of nominal value added over time (Figure 6). However, one should keep in mind that the average aggregated wage as well as the aggregated CompNet output at country level will never perfectly coincide with its Eurostat counterparts. This is mainly due to the fact that CompNet does not include all macro-sectors in its dataset (e.g. no financial corporations, no agricultural sector, only restricted number of public service sectors). Furthermore, there might be slight differences between CompNet and Eurostat definitions of value added as well as wages. Consequently, for example the total average wage as well as the total aggregated value added are lower than the Eurostat country figures.

Figure 6 Annual growth rate of nominal value added by country



Sources: CompNet 7th Vintage database, "all" sample, and Eurostat ama_10_gdp (gross value added current prices)

Notes: total nominal gross value added in Compnet = mean nominal value added * sum weights nominal value added

Besides average wage and output growth in a given country, CompNet collects information on the distribution of wages by sector. This granularity helps understanding aggregate trends like the popular "decline and fall of the labour share" theory, as will be shown in the next section.

5.5.4 Trade indicators

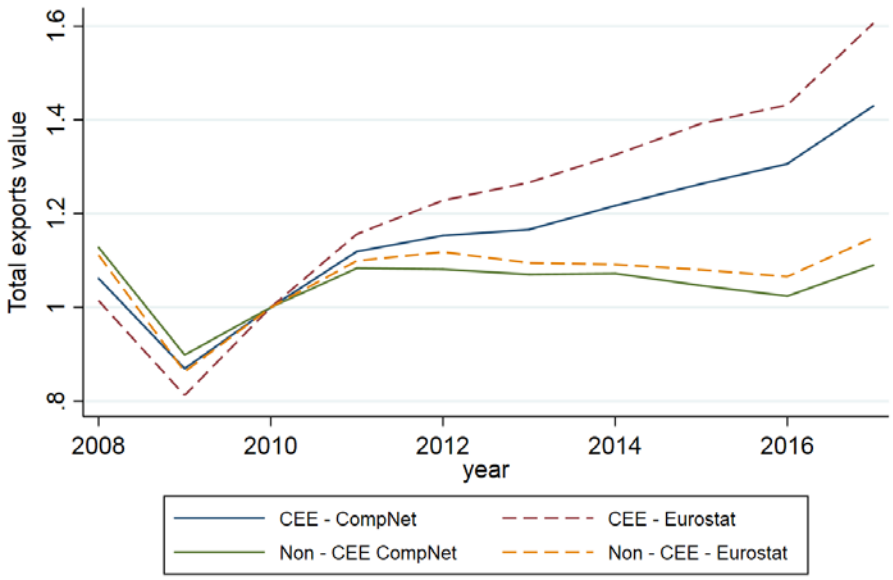
CompNet collects annual data on both exports and imports of goods. The novelty of the 7th vintage of the CompNet dataset is at least partial information on trade by destination. For several countries we may distinguish between the intra-EU and extra-EU exports or imports.

The CompNet dataset provides a unique source of trade characteristics of manufacturing firms. Besides the detailed information for different levels of aggregation countries, regions, sectors, macro-sectors and macro-sector-size classes, the trade module consists of a number of trade related joint distributions and at the same time trade variables are part of descriptive variables available for other joint distributions based on productivity, labour, or financial indicators.

Figure 7 aims at comparing the overall developments in exports from CEE and non-CEE countries. The presented indices confirm that CompNet dataset matches the overall trends very well. Although, the Eurostat data suggests somewhat stronger aggregate dynamics for both series, both datasets indicate more robust increase in exports from CEE countries and

much lower dynamics or stagnation in exports from non-CEE countries (especially since year 2011).

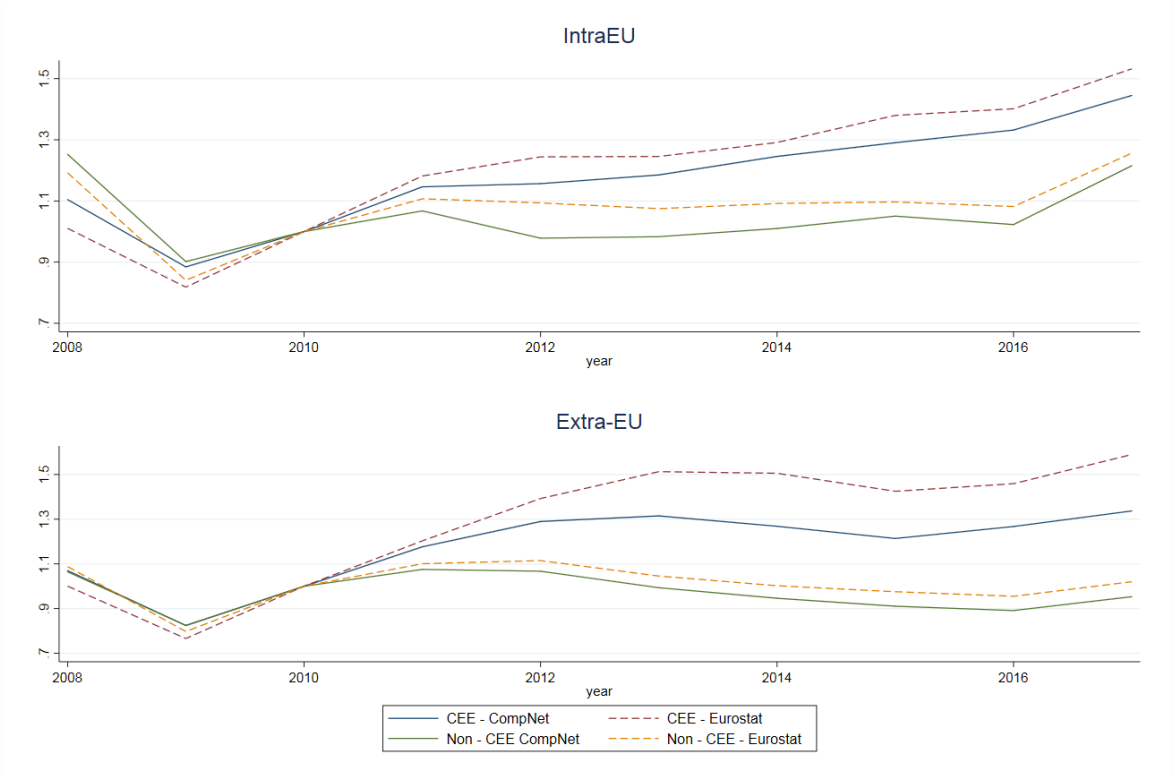
Figure 7 Overall export dynamics in CompNet and Eurostat (2010 =1)



Source: Eurostat, CompNet and authors' calculations.

Note: Based on the 20e population weighted sample. CEE countries: HR, CZ, LT, PL, SK, SI; Non-CEE countries: DK, FI, FR, SE. Country group aggregates are obtained by computing the simple average across country-specific indexes. DK only up to 2014.

Figure 8 Intra-EU and extra-EU export dynamics in CompNet and Eurostat (2010=1)



Source: Eurostat, CompNet and authors' calculations.

Note: Based on the 20e population weighted sample. CEE countries: CZ, LT, SK, SI; Non-CEE countries: FI, SE. Country group aggregates are obtained by computing the simple average across country-specific indexes.

Overall trade dynamics according to Eurostat are captured by CompNet data. It should be noted that observed discrepancies in CompNet and Eurostat data series may arise from various sources. The most dominant factors influencing the overall figures are cross-country differences in sources of original trade data (with different thresholds applied), uneven sectoral coverage (only export of manufacturing firms present in CompNet dataset) and country-specific differences in exporting firm coverage (departing from overall firm coverage).

The share of exporting firms recorded in CompNet varies between [68% in Lithuania] to [97% in Slovenia] and in most cases exceeds the share of exporting firms in Eurostat.

Table 21 Percentage of exporting firms in CompNet and Eurostat

Country	CompNet	Eurostat
Croatia	73%	39%
Czech Republic	78%	49%
Lithuania	68%	72%
Poland	72%	81%
Slovakia	83%	72%
Finland	82%	64%
France	71%	51%
Sweden	97%	75%
Slovenia	93%	79%
Denmark	89%	77%

Source: CompNet and Eurostat (TEC).

Note: 2016-2017 average. For Denmark CompNet data the share is computed for 2014 only, and compared with 2016 TEC values. These numbers refers to the manufacture sector, as the trade data availability in CompNet.

5.5.5 Financial Indicators

Data validation and stylized facts

Comparing a given dataset with official figures is a crucial task for researchers: firstly, it makes it possible to assess the validity of the created weighting scheme; then, it makes the researcher able to reach sound conclusion, not biased by the particular composition of the sample. With this spirit, we compare one of our indicators of financial constraints with the one published by the ECB based on the replies from the Survey on the access to finance of enterprises (SAFE)[38].

As explained in detail earlier in this methodological report, the CompNet SAFE indicator of credit constraints is derived in two steps. First, the SAFE score is a linear combination of some financial ratios as the following:

$$SAFE\_score_{\text{it}} = -5.47 + 0.07 \cdot finlev_i\text{t} - 1 + 0.46 \cdot ifp_{it-1} - 0.50 \cdot pm_{it-1} - 0.09 \cdot coll_{it-1} - 1.14\, cashH_i t - 1 - 0.05 \ln(TA_{it-1})$$

Second, firms are defined credit constrained if their SAFE score is above a certain threshold, which is directly provided by the survey and which varies across countries over time. The firm level indicator of credit constraints is then aggregated at different levels (e.g.: country, sector, nuts2), in order to show the (weighted) percentage of constrained firms in a given year at a specific aggregation level. This is important to recall, because the threshold is a weighted percentage where the weights are slightly different from those used in the CompNet dataset. In SAFE the country weights are derived from the Eurostat's structural business statistics to express the economic weight in terms of number of employees for four size classes (1 to 9, 10 to 49, 50 to 249 and 250 employees or more) as in CompNet, but the four macroeconomic sectors (industry, wholesale and retail trade, transport and other services to business or persons) are different between SAFE and CompNet.

In Figure 9 through Figure 12 we compare the evolution over time of the CompNet indicator of credit constraints with the share of constrained firms as derived directly from SAFE. This is done along different levels of aggregation, that being country, industry and country-firm size. All the comparisons refer to the all-firms sample, therefore eliminating some countries (e.g.: Slovakia and Poland) from the analysis. This choice is explained by the fact that the share of financially constrained firms whose number of employees is between 1 and 19 is non-negligible. Therefore, at the expense of not analyzing some countries due to lack of data for this particular part of the population of firms, we decided to focus on the sample including all sizes of firms.

[38] The SAFE is conducted by the ECB jointly with the European Commission twice per year. The survey intends to assess the financial conditions of firms in the euro area (the survey is also conducted for some non-euro area countries). It defines a firm as credit constrained if: the firm reports loan applications which were rejected; the firm reports loan applications for which only a limited amount was granted; the firm reports loan applications which were rejected by the firms because the borrowing costs were too high; the firm did not apply for a loan for fear of rejection (i.e. discouraged borrowers).
https://www.ecb.europa.eu/stats/ecb_surveys/safe/html/index.en.html

Figure 9 Share of credit constrained firms by country: SAFE data vs CompNet data

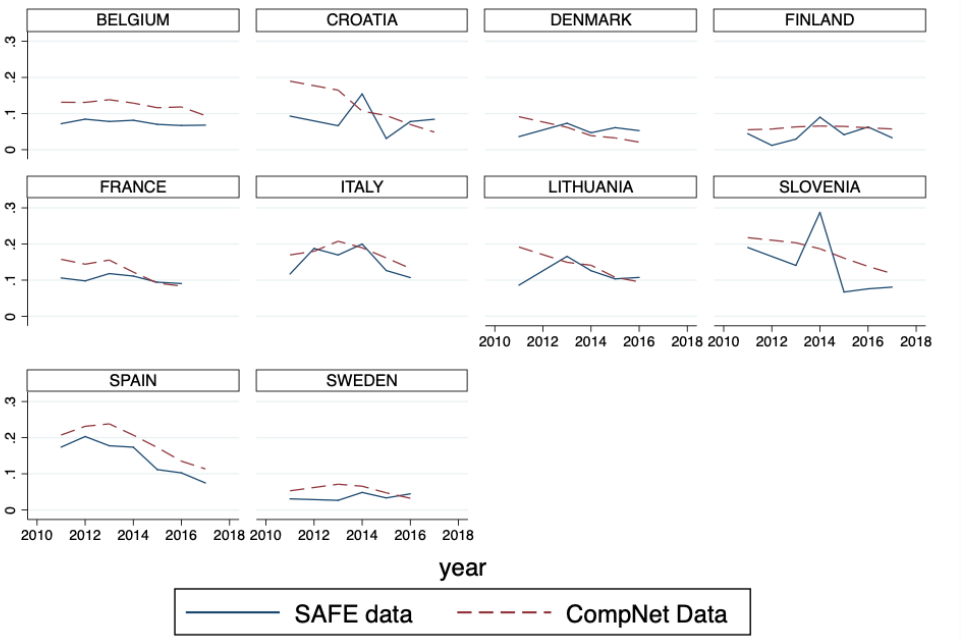


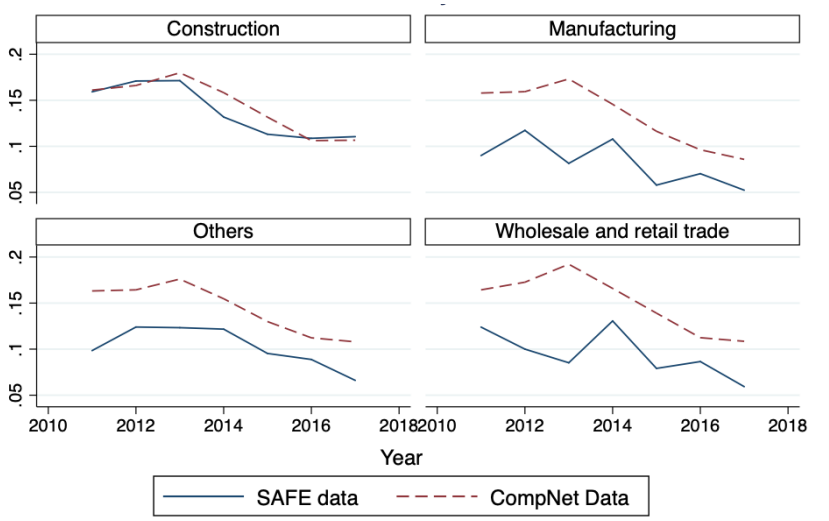
Figure 9 plots the share of financially constrained firms through several years according to CompNet and SAFE data. In the larger economies of the EU, like Italy, France, Spain or Belgium, the trend established by CompNet data is quite similar as it should be almost by construction. For the other countries, the trend seems quite similar as well, with the exception of Croatia and Slovenia, that exhibit both an upward spike in 2014. If for Croatia this might be due to the shock of an enlargement in the market that firms want to serve[39], the same rationale might not be applied to Slovenia, whose sudden increase in financially constrained firms might be attributed to some idiosyncratic factors. However, the comparison is still reasonable as, except for 2014, CompNet and SAFE indicator move on average in the same direction.

We did the same comparison at the industry level to check the reliability of our data. Here, the challenge is that CompNet macro-sectors do not coincide: in order to make an accurate comparison, we had to carefully select those sectors that are in common between the two datasets. Then, after having matched the single industries, we gathered data at the higher level of aggregation. In this way, we obtained a reliable comparison. The results for the four main sectors are presented in Figure 10.

It can be seen that overall trends are quite similar. However, while the shares of constrained firms are quite similar for the construction sector, the percentages are quite different and usually higher for the CompNet database. This might be due mainly to the fact that the sectoral composition of the macro-sectors can be different between CompNet and SAFE. More importantly, the sectoral developments show a similar peak in the share of financial constrained firm around 2013.

[39] Croatia joined the EU in July 2013.

Figure 10 Share of credit constrained firms over time by industry: CompNet vs SAFE data



Then, we applied the same calculations to firms of different sizes. The results are available in the following Figures. We differentiated for small enterprises (up to 50 employees) and medium-large ones. The results indicate the same spikes in Croatia among medium and large firms, and in Slovenia among small enterprises, clearing doubts about whose type of firms are responsible for the spikes. Generally speaking, the two series of data behave quite similarly not only at the country macro-level of aggregation, the highest one, but even when we decide to inspect micro level of aggregation, such as the industry or the country-size one.

Figure 11 Share of constrained firms by year. Small enterprises (1-49 employees)

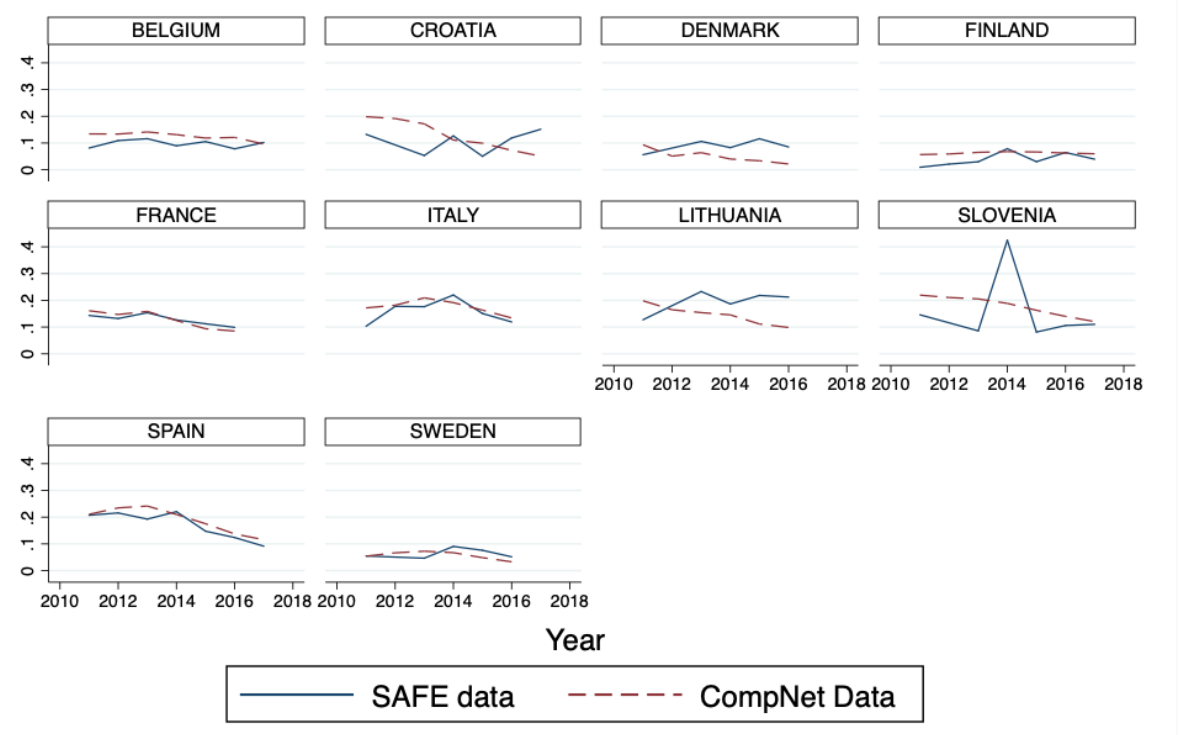
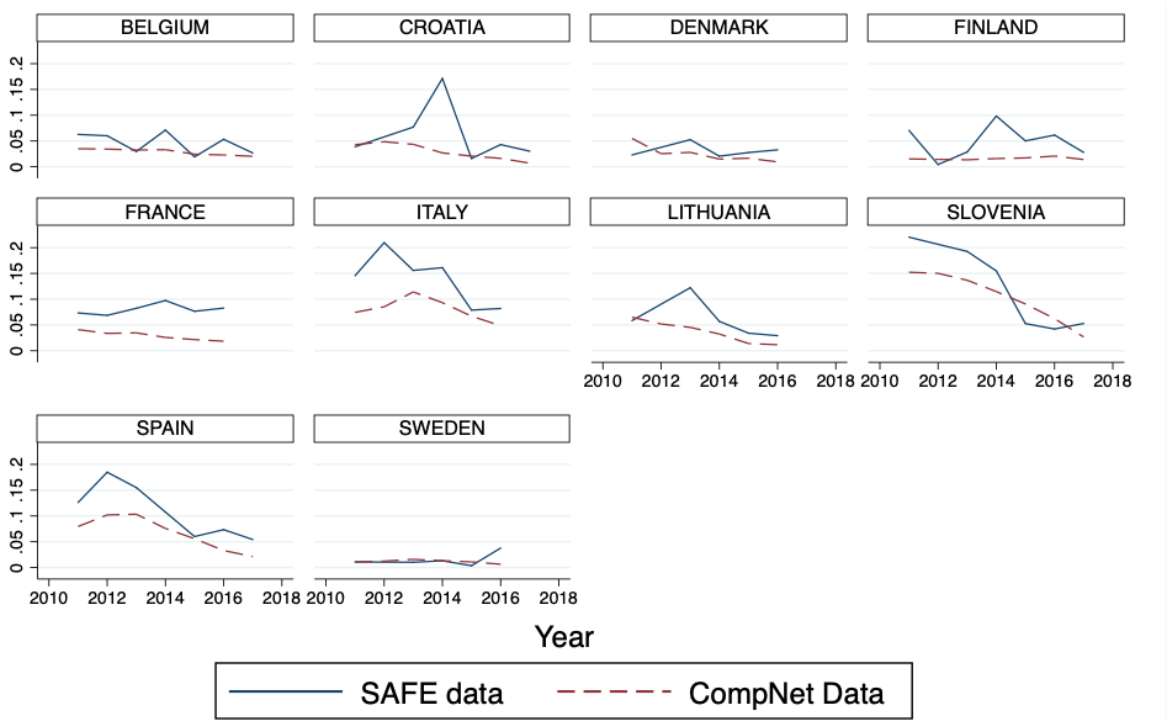


Figure 12 Share of credit constrained firms by year. Large firms only (50+ number of employees)



6. References

Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap. "Zombie lending and depressed restructuring in Japan." *American Economic Review* 98.5 (2008): 1943-77.

CompNet, "Firm Productivity Report" (2020).

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

Davis, Steven J., John Haltiwanger, and Scott Schuh. "Small business and job creation: Dissecting the myth and reassessing the facts." *Small Business Economics* 8.4 (1996): 297-315.

De Loecker, Jan, and Frederic Warzynski. "Markups and firm-level export status." *American Aconomic Review* 102.6 (2012): 2437-71.

De Loecker, Jan, Pinelopi Goldberg, Amit Khandelwal, and Nina Pavnik. "Prices, markups, and trade reform." *Econometrica* 84.2 (2016): 445-510.

Dobbelaere, Sabien, and Jacques Mairesse. "Panel data estimates of the production function and product and labour market imperfections." *Journal of Applied Econometrics* 28.1 (2013): 1-46.

Ferrando, Annalisa, and Alessandro Ruggieri. "Financial constraints and productivity: evidence from euro area companies." *International Journal of Finance & Economics* 23.3 (2018): 257-282.

Ferrando, Annalisa, Matteo Iudice, Carlo Altomonte, Sven Blank, Marie-Hélène Felt, Philipp Meinen, Katja Neugebauer, and Iulia Siedschlag. "Assessing the financial and financing conditions of firms in Europe: the financial module in CompNet." (2015). *ECB Working Paper Series No 1836.*

Gal, Peter. "Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS", *OECD Economics Department Working Papers*, No. 1049, OECD Publishing, Paris (2013), https://doi.org/10.1787/5k46dsb25ls6-en.

Haltiwanger, John, Robert Kulick, and Chad Syverson. "Misallocation measures: The distortion that ate the residual." *No. w24199. National Bureau of Economic Research* (2018).

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda. "Who creates jobs? Small versus large versus young." *Review of Economics and Statistics* 95.2 (2013): 347-361.

Hirschman, Albert O. "The Paternity of an Index" *The American Economic Review* Vol. 54, No. 5 (Sep., 1964), p. 761

Hsieh, Chang-Tai, and Peter J. Klenow. "Misallocation and manufacturing TFP in China and India." *The Quarterly journal of economics* 124.4 (2009): 1403-1448.

Foster, Lucia, John Haltiwanger, and Cornell J. Krizan. "Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s". *The Review of Economics and Statistics* 88.4 (2006): 748-758.

Kehrig, Matthias. "The cyclicality of productivity dispersion". *US Census Bureau Centre for Economic Studies Paper No. CES-WP-11-15* (2011).

López-Garcia, Paloma, Daniele Aglio, Richard Bräuer, Peter Haug, Jan Paul van der Kerke, Matthias Mertens, Matteo Sartori, Roberta Serafini, Ana Cristina Soares and Alessandro Zona Mattioli. "CompNet's 6th vintage of data: Novelties and main stylised facts". *The Competitiveness Research Network,* Available, 8 October 2018, www.comp-net.org. Accessed 16 November 2018.

McGowan, Muge Adalet, Dan Andrews, and Valentine Millot. "The Walking Dead? : Zombie Firms and Productivity Performance in OECD Countries". (2017). *OECD Economics Department Working Papers*, No. 1372, OECD Publishing, Paris, https://doi.org/10.1787/180d80ad-en.

Mertens, M. (2020). *Micro-mechanisms behind declining labour shares: Market power, production processes, and global competition*. Halle Institute for Economic Research and the Competitiveness Research Network, mimeo.

Olley, G. Steven, and Ariel Pakes. "The dynamics of productivity in the telecommunications equipment industry." *National Bureau of Economic Research No. w3977.* (1992).

Petrin, Amil, and Jagadeesh Sivadasan. "Estimating lost output from allocative inefficiency, with an application to Chile and firing costs." *Review of Economics and Statistics* 95.1 (2013): 286-301.

Wooldridge, Jeffrey M. "On estimating firm-level production functions using proxy variables to control for unobservables." *Economics Letters* 104.3 (2009): 112-114.