CompNet The Competitiveness Research Network

User Guide for the 8th Vintage of the CompNet Dataset

Last Update: 04.10.2021

When using the data, please cite it as:

CompNet 2021, User Guide for the 8th Vintage CompNet Dataset

This User Guide has been updated by Tommaso Bighelli, Jonathan Deist, Mirja Hälbig, Marcel Lang, Eugenia Macari, Matthias Mertens, Javier Miranda and Sani Wahyudi (all Halle Institute for Economic Research).
The working group on the CompNet 8th Vintage Data Collection has been coordinated by Matthias Mertens and includes: Peter Haug, Jonathan Deist, Tommaso Bighelli, Mirja Hälbig, Verena Plümpe, Marcel Lang, Roman Blyzniuk (all Halle Institute for Economic Research); Roberta Serafini, Philipp Meinen, Marco Matani, Chiara Gilardi (all ECB); Eric Bartelsman, Alessandro Zona Mattioli (both VU Amsterdam); Bernardo Mottironi (London School of Economics); Sergio Inferrera; Tibor Lalinsky (National Bank of Slovakia).

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 of the CompNet-homepage.[1] The CompNet staff will review the request and inform applicants about their decision. The processing time can be reduced if applicants provide sound information about themselves (e.g. CV) and their research project. 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, applicants will receive an email with the necessary credentials to log into the system access to the 8th (sample period up to 2019) data collection rounds for a period of six months. If needed, additional access to the previous 4th (up to 2012), 5th (up to 2013), 6th (up to 2016), and 7th (up to 2017) Vintages can be requested in the application or, if needed, by mail at a later point. Note that the individual vintages differ not only by year coverage, but also by variables included and methods applied to calculate indicators. We generally recommend using the latest vintage. A renewal of the data access 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 8th 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 7th and the 8th Vintage.[2] The above chapters are augmented by an extensive appendix which provides detailed information, lists and tables on:

- List of variables
- Derivation of parametric indicators[3]
- Country specific information
- Covered macro-sectors and two-digit NACE Rev. 2 industries
- CompNet Data Collection
  - Harmonization of input data/data preparation
  - Confidentiality
  - Outlier routine
- Data provider data sources

[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 8th Vintage CompNet Dataset

This chapter introduces the reader to the technical information necessary to use the dataset, including dataset structure, applied naming conventions and information about the content of the different types of sub-datasets.

2.1 Sample, Time Range and Levels of Aggregation

The 8th Vintage of CompNet dataset is an unbalanced panel dataset covering 19 European countries. The dataset includes a rich set of indicators from six roughly defined categories: Productivity, Finance, Labour, Competition, Trade, and Other.

These variables are available for two samples: The "all" sample and the "20e" sample. The all sample includes all firms in the target population, whereas 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] The time spans covered by both versions are identical for each country, but vary between countries. Table 1 shows the samples and time spans available in the 8th Vintage of the CompNet dataset.

Table 1: Countries, Samples and Time Span

Country	All firms	20e	Time Span
Belgium	X	X	2000–2018
Croatia	X	X	2002–2019
Czech Republic	X	X	2005–2019
Denmark	X	X	2001–2018
Finland	X	X	1999–2019
France		X	2004–2018
Germany	X[1]	X	2001–2018[2]
Hungary	X	X	2003–2019
Italy	X	X	2006–2018
Lithuania	X	X	2000–2019
Netherlands	X	X	2007–2018
Poland	X	X	2002–2019

[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.

Portugal	X	X	2004–2018
Romania		X	2007–2019
Slovakia		X	2000–2019
Slovenia	X	X	2002–2019
Spain	X	X	2008–2018
Sweden	X	X	2003–2019
Switzerland	X	X	2009–2018

[1] **Germany:** The macro-sector Manufacturing and Construction includes only 20e firms.
[2] **Germany:** Macro-sector coverage: Manufacturing (2001-2018), Wholesale and Retail Trade and Accommodation and Food Service Activities (2005-2018), other macro-sectors (2003-2018)

Target Population:

The CompNet dataset covers non-financial corporations with at least one employee covering the macroeconomic sectors as in table 24. This definition is consistent with category S.11 in the European System of Accounts (expect for sector 19, which, due to its small number of firms, is not covered by the CompNet dataset).[5] It consists of institutional units which are independent legal entities[6] and market producers, and whose principal activity is the production of goods and non-financial services (excluding sole proprietors). We refer to these independent legal entities as *firms* henceforth. The non-financial corporation sector also includes non-financial quasi-corporations. Detailed information on the sectors covered by the CompNet dataset is provided in Section 5.4.6, in the appendix.

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[7], Macro-Sector-Size-Class, Industry 2-digits[8], and NUTS 2[9].

[5] The small number of firms in sector 19 ("Manufacture of coke and refined petroleum products") makes it incompatible with the confidentiality/disclosure rules that apply to the CompNet dataset.
[6] The Netherlands are an exception because financial data could be provided only at the firm group level instead of the standard firm-level.
[7] Corresponding to NACE Rev. 2 sections
[8] Corresponding to 2-digit NACE Rev. 2 sectors
[9] Corresponding to basic regions for the application of regional policies based on the Nomenclature of Territorial Units for Statistics (NUTS).

The Macro-Sector aggregation is a single-digit industry classifications based on the NACE Rev. 2 sections. Section 5.4.6, in the appendix also contains a detailed definition of the Macro-Sector and industry level aggregation can be found in.

Most countries consistently used the NUTS2 2016 classification for all periods (Table 211 provides the exact reference). However, in some cases, no NUTS-classification could be provided (Netherlands).

Please note that all total factor productivity variables and (non-monetary) marginal productivity variables are available at the Industry 2-digits aggregation level.[10]

The size-class definitions, shown in Table 2, follow the Eurostat classification system:

Table 2 Size-Class Definitions

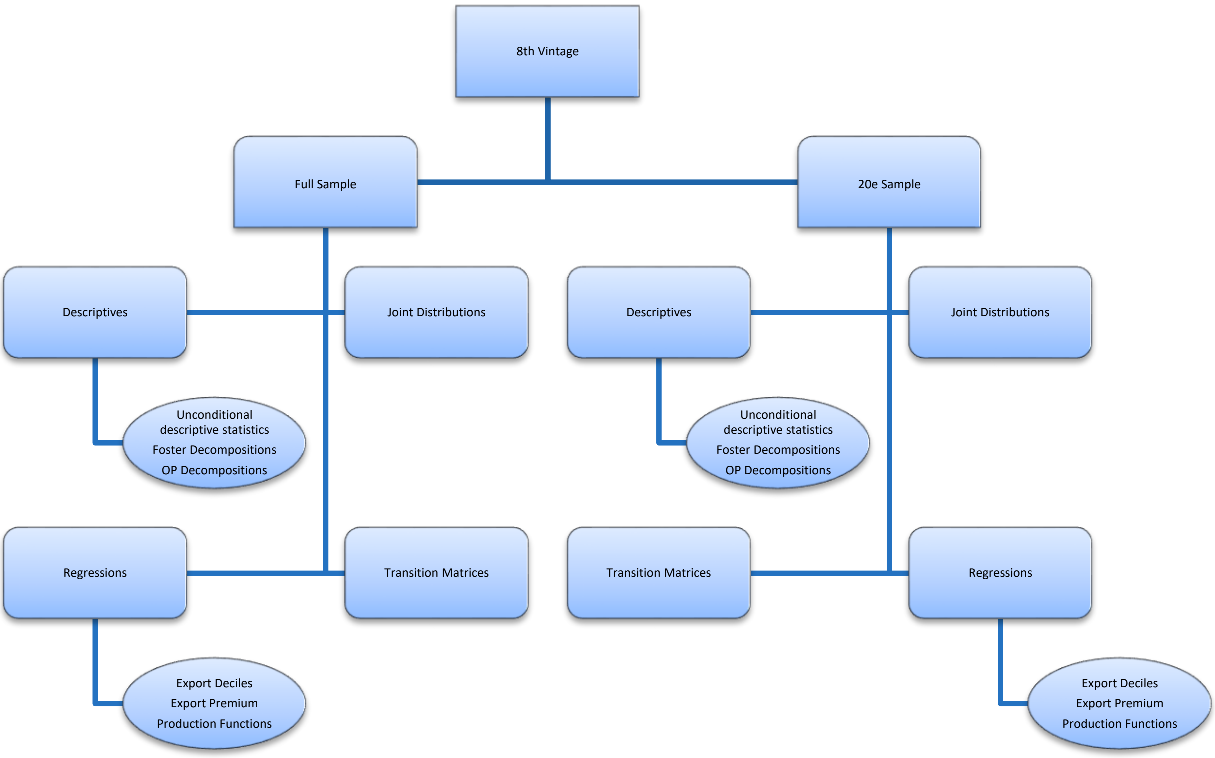
Size Class 1	Size Class 2	Size Class 3	Size Class 4	Size Class 5
1-9 empl.	10-19 empl.	20-49 empl.	50-249 empl.	>249 empl.

2.2 Structure of the 8th Vintage of the CompNet Dataset

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

[10] This concerns the TFP variables with codes PEb0 – PEb5, PEj0 – PEj5, and PEj0G1 (cf. Table 11), and the non-monetary marginal productivities variables with codes PEb7 – PEd4 (cf. Table 15).

Figure 1 Structure of the Dataset



The files comprising the dataset are mostly Stata (.dta) files. Only the regression results are provided as Excel files (.xls).[11] 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 following four subsections detail the structure for each of the four themes. In general, file names follow the pattern ***content_dimension_sample_weighting****.dta* (although the order of the single components might change) to specify

1. ***Content*** The dataset's theme, and, if applicable, the main variable of interest (i.e. Descriptive statistics, Joint Distributions, Regressions, Transition Matrices)

[11]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. into R by the R-packages readstata13 or haven or via RStudio. Alternatively, you could use file conversion software like StatTransfer.

2. ***Dimension*** The level of aggregation of the dataset

(i.e. country, macro-sector, macro-sector-size class, NUTS2, and industry 2-digits)

3. ***Sample*** Indicates the sample the dataset is built on:

1. "all" includes all firms with at least 1 employee
2. "20e" includes all firms with at least 20 employees

4. ***Weighting*** "weighted" and "unweighted" for the population weighted and the unweighted sample version, respectively (Section 3.2.1 provides details).

Examples will illustrate the names for each of the four themes.

2.2.1 Descriptives

The descriptive section includes two types of datasets: Unconditional distributions ("Unconditionals") of variables, and aggregate variable decompositions, including aggregate productivity decompositions ("Decompositions").

Unconditionals

The unconditional datasets provide the distributions all indicators available in CompNet.[12] These distributions (and files) are called unconditional because they are given for each indicator and firm sample (as defined by the panel) separately, without encompassing information about other indicators or firm populations. The unconditional distribution of each indicator is described by its percentiles, first four moments (i. e. mean, standard deviation, skew, and kurtosis), and the number of firms in the respective panel subset.

All unconditional datasets can be recognized by the content-prefix "unconditional_" in their file names. They are available for the country, macro-sector, macro-sector-size-class, industry 2-digits and NUTS2 dimension, each for both the all- and the 20e-sample and both as weighted and unweighted versions and identified by the according suffix as follows:

[12] Exception: No unconditional distributions are available for the approximated value-added elasticities with respect to capital and labour (e. g. PEe8_oe_k_0_va and PEg4_oe_l_2_va).

"*unconditional_dimension_sample_[**un]weighted**"*. Table 16, in the Appendix, provides a list of all variables available in the unconditional descriptive files.

Decompositions

The decompositions are datasets which contain the calculations of different aggregate variable decompositions for efficiency and productivity measures as well as other variables. The files can be recognized by the prefix *"fhk_decomp" and "op_decomp"*. These stand for decompositions in line with the methodology in Foster et al. (2006) and in Olley & Pakes (1996), respectively. Both decompositions are available for four different levels of aggregation: country, macro-sector, industry 2-digits, and NUTS2. The Foster decompositions are available only as weighted versions. Among the indicators in this dataset, the user can find the decomposition of sector productivity into the unweighted mean and covariance term between productivity and the applied weight of economic activity (e.g. size). Further information on the computation of these indicators can be found in Section 5.3 of the appendix and the original articles. Table 17 and Table 18, in the Appendix, provide the lists of variables in the decomposition dataset.

Example 1: **File names**

The dataset that contains general unconditional descriptive statistics for the weighted sample including all firms, at country level, is called

unconditional_country_all_weighted.dta

2.2.2 Joint Distributions

Joint distributions give the percentiles and sample moments of a summarized variable under the condition that the respective firm sample is defined by the percentiles of another variable, i.e. that all firms in the sample are similar in terms of the conditioning variable.[13] If the conditional variable is discrete, its levels (instead of percentiles) define the samples for which the conditional distributions of the summarized variable are calculated.

A joint distribution requires that each analysed firm reports both variables. In order to analyse the largest possible number of firms, the 8th Vintage provides, as a novelty, joint distributions

[13] For this reason, the independent distributions of included variables in the descriptive statistics are called *unconditionals.*

for variables from ten separate thematic groups, shown in Table 3. Each group consists of conditional variables and summarized variables. The conditional distributions are provided for both, but the latter are not used as condition*ing* variables. Please note that the files are identified by the abbreviated short group names, given in column two of Table 3.

The unweighted conditional distributions describe the firm sample. The weighted versions use inverse probability weighting within the strata defined in the calculation (not those of the survey data collection) to describe the population of firms.[14] Within each group, all conditional distributions are derived from the same sample of firms and weights (i.e. each joint distribution file contains only firms providing information on all contained variables).[15] Because conditional distributions require all underlying firms to report complete data on all variables described in the respective distribution, all groups use distinct samples, which means that conditional distributions are *not* comparable across group files, because the samples and weights (if applicable) differ. This applies to both the unweighted version and the weighted one.[16]

Table 3 Variables in the Joint Distributions, by group

Group Topic	Short name	Summary	Variables	Description
Input	inp	Labour Labour cost Wages Capital Top10-dummies	LV21_l	Headcount
			LR01_lc_va	Ratio: wageshare: nom. labor cost / nom. value-added
			LV24_rwage	Real wage
			FR35_va_rev	Ratio: nom. value-added / nom. Revenue
			FV18_rva	real value-added, computed as rev – m
			FV17_rrev	real revenue
			FR30_rk_l	Ratio: capital intensity: real capital / labor
			PV03_lnlprod_va	Log labor productivity, real value added based: ln(rva/l)

[14] The weighted joint distributions describe the population of firms reliably only if the sample is a random sample within the respective weight-dimension (i.e. size-class year 2-digit industry).
[15] This ensures within-group consistency while maximizing the number of available observations in each group.
[16] In theory, the weighted conditional distributions would be comparable across groups. In practice, the group samples are not random samples, and can include only those firms that have complete observations.

			FV14_rk	real capital
			PV05_lnsr_cs	Log. Solow residual, weights in CD from cost shares
			PV04_lnsr	Log. Solow residual, weights in CD: labor 2/3, rk 1/3
			LR03_ulc	Ratio: Unit labor costs: nom. labor cost / real value-added
			FD02_t10_rev_C	D = 1, if firm is among Top10 revenue firms, country level
			FD10_t10_rva_C	D = 1, if firm is among Top10 real value added firms, country level
			LD01_t10_l_C	D = 1, if Top10 firm by employee-number, country level
			FD18_t10_rva_2D	D = 1, if firm is among Top10 real value added firms, sector level
			FD04_t10_rev_2D	D = 1, if firm is among Top10 revenue firms, sector level
			LD03_t10_l_2D	D = 1, if Top10 firm by employee-number, sector-level
			LD02_t10_l_M	LD02_t10_l_M
			FD03_t10_rev_M	D = 1, if firm is among Top10 revenue firms, mac-sector level
			FD14_t10_rva_M	D = 1, if firm is among Top10 real value added firms, macsec level
Productivity	prod	TFP Markups MPL/MPK MRPL/MRPK	PEj0_ ln_tfp_1	TFP - Specification 1 (CD, industry-level OLS)
			CE33_dm_1	Firm labor market power - Spec. 1
			CE39_markup_l_1	Firm markup - labor input decision - Spec. 1
			CE45_markup_m_1	Firm markup - intermediate input decision - Spec. 1
			CE51_markup_ml_1	markup - labor + intermed. input decision - Spec. 1
			PEh8_ps_1	MRPL-wage Gap (deflated) - Specification 1

			PEb8_mpk_1	marg. Product of capital - Specification 1
			PEc4_mpl_1	marg. Product of labor - Specification 1
			PEd0_mpm_1	marg. product of intermediates - Specification 1
			PEe2_mrpl_1	marg. revenue product of labor - Specification 1
			PEd6_mrpk_1	marg. revenue product of capital - Specification 1
Finance	fin	Investment Credit access	FD01_safe	D = 1, if firm is financially constrained
			FR38_invest_rev	Ratio of nominal investment to nominal revenue
			FR19_op_inte	Ratio: operating profits / interest payments
			FV31_rinvest_intan	real intangible investment
Trade	trad	Import/Export ratios Categorical variables for large importers/exporters Imports/Exports relative to revenue	**Conditioning variables**	
			TC03_exp_top10	Categorical variable indicating large exporters
			TC06_imp_top10	Categorical variable indicating large importers
			TC01_exp_imp_rel	Categorical variable indicating the direction of trade
			TC00_exp_dest	Categorical variable indicating export destinations
			TC04_imp_dest	Categorical variable indicating the origins of imports
			TD15_exp_adj	D = 1, if exporting, adj.
			TD89_imp_adj	D = 1, if importing, adj.
			Additional summarized variables	
			TR02_exp_adj_rev	Ratio: exports adj. / nom. Revenue
			TR38_imp_adj_rev	Ratio: imports adj. / nom. Revenue
Trade Timing	trca	Variables describing if firms started, stopped, resumed, or paused importing/exporting	**Conditioning variables**	
			TC02_exp_time_3y	Categorical variable indicating the timing of exports
			TC05_imp_time_3y	Categorical variable indicating the timing of imports
			Additional summarized variables	
			TD15_exp_adj	D = 1, if exporting, adj.

			TD89_imp_adj	D = 1, if importing , adj.
			TR02_exp_adj_rev	Ratio: Export Ratio: exports adj. / nom. Revenue
			TR38_imp_adj_rev	Ratio: import Ratio: imports adj. / nom. Revenue
Growth Rates	grow	Indicators for growth of productivity, firm size, and exports	PV03G1_lnlprod_va	Growth rate (from t-1): log. labor prod., real value-added based
			TV02G1_exp_val_adj	Growth rate (from t-1): adjusted exports
			PEj0G1_ln_tfp_1	Growth rate (from t-1): ln_tfp_1
			LV21GH_firm	Haltiwanger-Davis growth rate (from t-1): labor.
Firm Demography	demo	Firm age	OC00_firm_age	`1 "0-2 years" 2 "3-5 years" 3 "6-25 years" 4 "more than 25 years"'
Zombie	zomb	Status of zombie firms	FC07_y_zombie_intcov_pos	Categorical: Duration of current spell as zombie
Intangible Fixed Assets	ifa	Intangible capital to nominal revenue	FR36_ifa_rev	Share of intangible k to rev_nom
Tax	tax	Effective tax rate	FV24_etr	Effective tax rate

To enable cross-group comparisons based on consistent samples without compromising the sample sizes, additional conditional distributions are available for each pairwise combination of groups (e.g. Finance-Trade). These pairs, shown in Table 4, combine the conditional and summarized variables of the two respective groups and report the resulting joint distributions for a homogenous sample using the same weights. Please note that each of these paired groups, too, is based on a unique sample. In other words, all joint distributions inside a given combined group file (e.g. Finance-Trade) are comparable to one another. Comparison between different groups (e.g. Finance-Growth) is possible for the weighted versions and assuming no sample-selection in the weighting.

The file names follow the same general naming scheme introduced above. Additionally, the file names identify the respective thematic group that are based on using the short names given in Table 3. Files are based on a pair of groups as described below (e.g. inp-tax, see Table

4). The following examples illustrate the logic. The joint distribution files are available at the country, macro-sector, industry 2-digits, and NUTS2 dimensions for both the all-firms and the 20e sample in weighted and unweighted versions.

Example 2: **Joint Distribution File name**

The joint distribution file containing the real capital variable –which is part of the input-group (inp, for short, please see Table 4 for the short names) and called FV14_rk (the identification code is explained in Section 2.3) – at the macro-sector for the weighted 20e sample has the name:

jd_*inp*_*mac_sector* _*20e*_*weighted*.*dta*

Example 3: **Joint Distribution File name**

The file containing the joint distribution of the real capital (group: inp [inputs]) conditional on firm age (group: demo [demography]) variable OC00_firm_age, at the two-digit industry level for the unweighted all-firms sample is called

jd_*inp*_*demo*_*industry2d* _*all*_*unweighted*.*dta*

The ten thematic groups result in a total of 45 *group* pairs, listed in Table 4. Each pair exists only in one of two permutations, meaning that while e.g. the *group* Input-Finance (called inp_fin, in short) exists, no files exist for the *group* Finance-Input: The latter pair's content would be identical to that of the *group* pair Input-Finance and would differ in name only.[17] This naming scheme and the use of a continuous joint distribution are illustrated in the box Example 4, below.

Table 4: Overview of the Joint Distribution group combinations, using the short names (see Table 3)

Group	inp	prod	fin	trad	Trca	Grow	demo	zomb	ifa	tax
inp	inp	inp_prod	inp_fin	inp_trad	inp_trca	inp_grow	inp_demo	inp_zomb	inp_ifa	inp_tax
prod		prod	prod_fin	prod_trad	prod_trca	prod_grow	prod_demo	prod_zomb	prod_ifa	prod_tax
fin			fin	fin_trad	fin_trca	fin_grow	fin_demo	fin_zomb	fin_ifa	fin_tax
trad				trad	trad_trca	trad_grow	trad_demo	trad_zomb	trad_ifa	trad_tax
trca					trca	trca_grow	trca_demo	trca_zomb	trca_ifa	trca_tax
grow						grow	grow_demo	grow_zomb	grow_ifa	grow_tax
demo							demo	demo_zomb	demo_ifa	demo_tax

[17] Group "pairs" like input-input would be similarly redundant and thus do not exit.

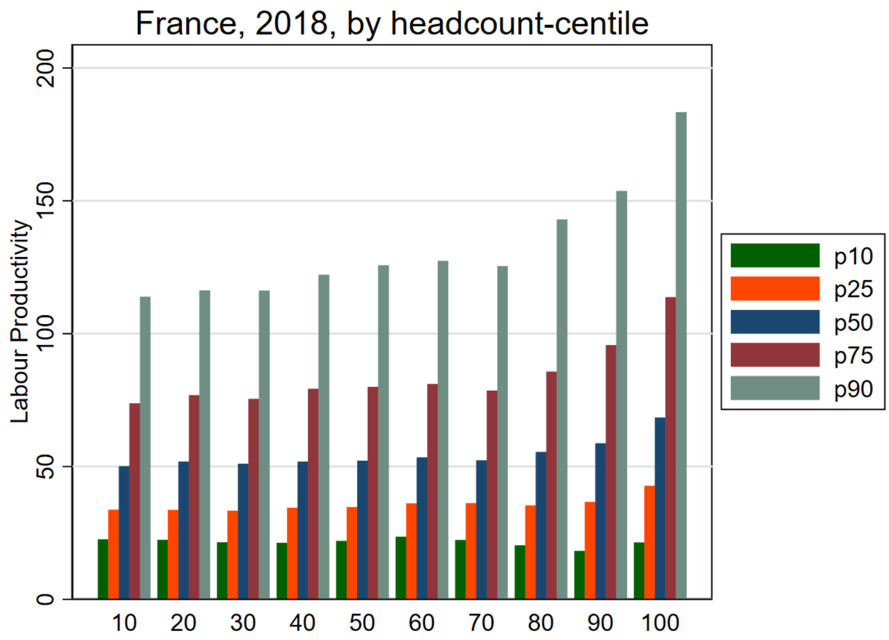
zomb								zomb	zomb_ifa	zomb_tax
ifa									ifa	ifa_tax
tax										tax

A discrete conditional variable could be e.g. a dummy variable that takes the value one for zombie 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 being a zombie or not a zombie. To condition on a continuous variable, for example labour productivity, the dataset uses deciles of the variable (at the industry 2-digit aggregation level, quintiles are used instead of deciles). The distributions of the summarized indicators are provided given the deciles of firms' labour productivity within the respective dimension.

Example 4: Working with Joint Distributions

Figure 2 shows an example of a (continuous) joint distribution.[18] In particular, it uses the variable *PV03_lnlprod_va* to show the labour productivity distribution of firms for different deciles of the firm size (*LV21_l*; taken from the variable/column *by_var*) in France in 2018. The conditional distribution reveals that there may be a discontinuity in the productivity of firms by size: The largest 10% of firms in the 20e sample appear to be on average more productive than smaller ones, except for the 10% least productive firms (represented by the green bars) in each size decile.

Figure 2: Labour Productivity Distribution by Firm Size Centiles in France



2.2.3 Transition Matrices

The CompNet transition matrices track the evolution of firms over a three-year window and thereby allow researchers to study firm size dynamics as well as the characteristics of firms with different growth performances over the respective period. Conditional on surviving the three-year period under study, firms are classified into quintiles in a given macro-sector based on firm number of employees for year *t* and *t-3*, respectively. For example, one can analyze the movement of firms across these quintiles, moving from size quintile 1 in *t-3* to size quintile 5 in *t*. In addition, the transition matrices provide statistics on selected characteristics of firms

[18] The file containing this joint distribution is *jd_inp_country_20e_weighted.dta*.

at time *t-3* and *t*, so that it is possible to analyze firms' features before and after their growth or shrinking.

The transition matrix datasets are indicated with the prefix "*tm_tm"* and are available at the macro-sector level of aggregation. Size quintile in t-3 is indicated by *qt_l3_l* and size quintile in *t* is indicated by *qt_l*. The prefix *l3_* indicates firm characteristics as of t-3, while the prefix g3_ indicates firms' annualized growth rates from *t-3* to *t. Additionally, the files with the prefix "tm_unc" provide summary statistics for the firm-sample that was used to compute the transition matrices, but without conditioning on the employment quintile transitions.*

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. Additionally, for countries with trade data, regression results for export probability or export premiums are available. The following subsections describe the three types of regressions.

Production Functions:

Production functions are estimated by pooling all firms operating in a given 2-digit NACE Rev. 2 sector. Each production function estimation uses the log of deflated turnover as dependent variable, while the regressors are the log of deflated capital stock, deflated cost of intermediate inputs and labour (headcounts). Apart from the regression output files in Excel format, the coefficients and standard errors of the different specifications are also provided collectively in Stata files, one for each sample and weighting. In the 8^{th} Vintage, the unconditional descriptive statistics provide information on standard deviations for all variables, and the previously reported Kehrig's dispersion measures have been discontinued. Specifications including intangible capital as separate production input were discontinued, too, as we now include intangible capital into the capital input variable.

The following parametric specifications replace the formerly used methodologies:

1. OLS estimation assuming a Cobb-Douglas (CD) production function.
2. OLS estimation assuming a Translog (TL) production function.
3. OLS estimation, assuming CD production function with time-varying output elasticities.

4. GMM estimation following Ackerberg, Caves and Frazer (2015) assuming a CD production function.
5. GMM estimation based on Ackerberg, Caves and Frazer (2015) assuming a TL production function.

Please refer to Section 5.3.1, in the appendix, for more details on the estimation procedure; also, a detailed description on how to compare productivity across industries, sectors, regions, and countries can be found in Section 5.3.4.

Export-Probability:

The first set of export-related regression output files give the user the results of probit estimations of the probability to export conditional on firms' productivity and size. 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:

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

with **ct_x**[19] 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.

Export Premium:

The second set 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 and size class dummy variables, a crisis dummy, and interaction terms. Just like in the probit estimation, the observations are weighted by their inverse sampling probability defined as the

[19] 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.

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:

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

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

Table 5 shows the dependent and independent variables:

Table 5: Export Premium (weighted OLS regressions)

Dependent variables Identification Code	Variable Name	Independent variables Variable name
PV01	lnkprod_va	d_exp_val_adj
PV03	lnlprod_va	Sector
PV02	lnlprod_rev	Crisis
PEi9	ln_tfp_0	Szclass
PEj0	ln_tfp_1	crisis#d_exp_val_adj
PEj1	ln_tfp_2	
PEj2	ln_tfp_3	
PEj3	ln_tfp_4	
PEj4	ln_tfp_5	
PV05	lnsr_cs	
PV04	lnsr	

2.3 Naming Convention of Variables

The naming convention for variables in the 8th Vintage of the CompNet dataset continues the scheme introduced in the 7th Vintage and enhances the naming of the growth rates of selected variables. In addition, the code system from the last vintage is maintained and uniquely identifies variables by a combination of 4 characters, also across both vintages (with a few exceptions). This "identification code" can always be found at the beginning of each variable. The data section of the CompNet website provides a short Stata script which can be used to remove the CompNet identification code, if desired. This can be useful for individual analysis on a standard set of indicators. However, we recommend using the code system as it allows for quick identification of needed indicators.

8th Vintage Variable Naming Scheme:

Every variable included in the 8th Vintage of the CompNet dataset follows the pattern:

IdentificationCode_IndividualName_[Weightedby]_Suffix

The (4-digit) **identification code** is built with three elements: the thematic category, the variable type, and the 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. 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 to distinguish different indicators within each category and variable group. For example, in the category “competition” the dataset includes 3 “ratio” type variables with the codes 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 100 variables for a given category and variable group. In such cases, the previous numerical coding is extended by the following alpha-numeric sequence: *{a0, a1,…, a9, b0, b1,…, b9, c0, c1…}*. Table 6 summarizes the details of the identification codes. Please be especially aware of the fact, that the identification codes are unique across the 7th and 8th vintages. Variables that have an equivalent in the last vintage will have the same code as before. At the same time variables and their codes from the 7th vintage which are not part of the current vintage have disappeared. The latter also applies to cases where the definition of a variable has significantly changed. Details about these cases can be seen in the “Detailed Variables Overview” (Appendix Section 5.2).

Table 6: CompNet Identification Code

Categories:	Variable Types:	Alpha-Numeric Code System (read vertically)						(optional) Growth Rates
C – Competitiveness	C – Categorical	00	06	a0	a6	b2	b8	GH
F – Financial	D – Dummies	01		a1	a7	b3	b9	G1
L – Labour	E – Estimates[3]	02	…	a2	a8	b4	c0	G3
P – Productivity	R – Ratios	03	09	a3	a9	b5	c1	
T – Trade	V – Values[4]	04	10	a4	b0	b6	c2	
O – Other		05	…	a5	b1	b7	…	
			99					

[3]Estimates are defined as any variable which is based on a production function
[4]Defined as a number that represents an amount

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 "top_rev_sam_C" in the variable name CR00_top_rev_sam_C_* stands for the top10 firms' share in the aggregated sample revenues at the country level. The overview including the identification code, the individual variable name and a definition of all published output variables can be found in Section 5.2 in the appendix.

Growth rate variables (calculated for selected variables) provide information about the distribution of a variable's growth per quantile. They are identified by the additional two-digit suffix **G#** following their underlying variable's four-digit code. The # identifies the specific type of growth rate: For example, the variable *real capital* has the code FV14_rk, and the variables containing its growth rates from one year and three years ago are called FV13**G1**_rk and FV13**G3**_rk, respectively, where 1 and 3 denote the time dimension.[20] Growth rates calculated according to Davis, Haltiwanger and Schuh (1996) are denoted by the third alternative suffix **GH** (these are one-year growth rates with a definition accounting for entry and exit).

Weightedby is used only for the aggregate variable decompositions (which are part of the unconditional distribution files) and indicates the input-weighting method. The input-weighting used here is not to be confused with weighting regarding the sample/population used elsewhere in this user guide. The labels indicating the weightings always start with a capital W and end with the abbreviation of the weighting method, e.g. Wrrv (which stands for

[20] In the 7th Vintage, the same growth rates were coded F**G**01_rk_y1 and FG01_rk_y3, respectively – which masked the 'value' type of the underlying variable rk, and still required the _y# suffix in the variable names.

"weighted by real revenue"). If no input-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 7 reports the different input-weighting methods used in the 8th Vintage of the CompNet dataset for the decomposition variables[21] including the applied abbreviation.

Table 7 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 provides a rich set of information for every variable included in the dataset. This includes descriptive statistics like the mean, percentiles or the number of firms with non-missing values for the respective indicator. These statistics are identified by **suffixes** (Table 8) in the variable names.

Table 8: Suffixes

Suffix	Meaning
p1, p5, p10, p25, p50, p75, p90, p95, p99	Percentiles
mn	Mean
sd	Standard deviation
skew	Skewness
kurt	Kurtosis
n	Number of observations
sw	Summed weights (= population number of firms)
umn	The input-unweighted mean in the OP decomposition
usw	Input-Unweighted summed (population-)weights in the OP decomposition
wmn	Input-weighted mean in the OP decomposition
cov	The covariance term in the OP decomposition

[21] This should not be confused with the general weighting procedure that is applied to all indicators. See 5.4.3 in the appendix.

wth	Indicates the *within*-component in the FHK decomposition
btw	Indicates *between* component in the FHK decomposition
agg	Indicates the *aggregate* term in the FHK decomposition

3. Important Notes on Using the Dataset

This chapter highlights a few important features of the 8th 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 8th Vintage. Topics range from technical intricacies to correctly interpreting and combining provided information. The purpose is to help the user avoid "technical" mistakes in using the dataset and enable him or her to conduct sound research with the data.

It is important to stress that the 8th Vintage of the CompNet dataset addresses a multitude of caveats existing in the 7th Vintage of the CompNet dataset, improving the accuracy and comparability of many variables. An overview of these improvements is included in Section 4. More information about the comparability 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.[22] 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, not

[22] For more details see Section 2.1

the underlying sample.[23] Consequently, one frequent question is how to compute aggregate statistics with the variables provided in the weighted datasets in 8th 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 non-ratios variables. Variables with the suffix "N" show the relative number of firms in the cell with available information to construct the given variable.

If the researcher wants to collapse the dataset to a higher level of aggregation (for example, from the industry 2-digits to the macro-sector level), one needs to use the "_sw" variable to have population-representative weights (it applies only to non-ratio variable). This is an important difference with respect to CompNet vintages **preceding** the 7th 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 in all categories. Dummy variables are identified in their naming code by the variable type ***D*** (for example F**D**01_safe). Due to the binary nature of dummies, no percentiles or moments of their distribution are provided as descriptive statistics. The mean, however, does provide useful information about a dummy – namely the percentage of observations (firms) for which the variable is equal to 1 – and is included in the dataset. For example, the mean of "TD14_exp" gives the share of exporters in the given cell and, therefore, takes values between 0 and 1.

3.2.3 Categorical Variables

The conditional distributions in the 8th Vintage leverage new categorical variables, thereby making the information formerly contained in numerous binary dummy variables more accessible. All categorical variables are identified in their naming code by the type ***C*** (for

[23] 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.

example O**C**00_*firm_age*). The conditional distributions contain moments for each level of the categorical variables.

3.3 Comparability

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

(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, changes in sector classifications or improvements of the underlying methodologies.

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.[24] 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.[25] 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 goal of the CompNet dataset is to provide the distributions of variables rather than their total values. This goal is less sensitive to varying coverage ratios, and the overall assessment of the sample representativeness is very positive.[26] For details, please see Section 5.4.7.

[24] "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."
[25] See appendix Section 5.4.4
[26] For an assessment see the CompNet cross-country comparability report (2018)

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 differences in the output of the code. All data providers use a set of harmonized definitions, including 1st, 2nd and 3rd best variable definitions. Section 5.4.5, in the appendix, contains detailed overviews:

- Table 21 includes information on all raw variables and their possible definitions
- Table 22 highlights the used definitions for each country included in the dataset.

4. Differences 7th Vintage vs 8th Vintage of the CompNet Dataset

The 8th Vintage of the CompNet Dataset includes some innovations in respect to the previous editions which improve the availability and the quality of the data and their user-friendliness. The most important innovations include:

Production Functions:

- The entire Production Function Module has been rationalized. It now includes only 5 different TFP estimates as indicated in Section 5.3.1, in the appendix, and a non-parametric estimation based on cost shares.
- All estimates are made at the 2-digit NACE industry level, better capturing similarities of production processes between firms.
- The outcome variable is always the log of gross output (turnover). Value added is no longer used as outcome variable.
- Specifications including intangible capital as separate input factor were discontinued. We now include intangibles in the general capital stock.
- Unconditional descriptive statistics report standard deviations for all variables, “Kehrig’s dispersion measures” have been discontinued.

Joint Distributions (JDs):

- Created content-specific subgroups of JDs computed within an population of firms.
- Reorganized a large number of dummy conditions into new categorical conditions.
- 2-digits sector level JDs are computed by quintiles of the distribution instead of deciles.

New Variables:

- The share of the 10 biggest firms of the selected variable is available for more variables as concentration measure
- Expanded the HHI concentration indicators to cover more variables.
- More comprehensive dummies provide more details on trade timing.

New Countries and coverage:

- Germany has now a full coverage of all the macro-sectors for the period 2005-2018.

5. Appendix

5.1 List of Data Folders

The following overview presents the available data folders and number of files in the 8th Vintage of the CompNet Dataset.

Table 9 Data Files Overview

Sample & weighting	Folder	(Subfolder)	Number of files
20e_firms_unweighted	**Descriptives**		9
	JointDistributions		220
	Regressions	Export_deciles Export_premium Production_Functions	165 165 96
	Transmatrices		-
20e_firms_weighted	**Descriptives**		13
	JointDistributions		220
	Regressions	Export_deciles Export_premium Production_Functions	164 165 96
	Transmatrices		2
all_firms_unweighted	**Descriptives**		9
	JointDistributions		220
	Regressions	Export_deciles Export_premium Production_Functions	110 110 81
	Transmatrices		-
all_firms_weighted	**Descriptives**		13
	JointDistributions		220
	Regressions	Export_deciles Export_premium Production_Functions	121 121 81
	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 10 Competition Variables

Variable Code	Variable Name	Definition
		Competition
		Estimates
CE32	dm_0	Firm labor market power - Spec. 0
CE33	dm_1	Firm labor market power - Spec. 1
CE34	dm_2	Firm labor market power - Spec. 2
CE35	dm_3	Firm labor market power - Spec. 3
CE36	dm_4	Firm labor market power - Spec. 4
CE37	dm_5	Firm labor market power - Spec. 5
CE38	markup_l_0	Firm markup - labor input decision - Spec. 0
CE39	markup_l_1	Firm markup - labor input decision - Spec. 1
CE40	markup_l_2	Firm markup - labor input decision - Spec. 2
CE41	markup_l_3	Firm markup - labor input decision - Spec. 3
CE42	markup_l_4	Firm markup - labor input decision - Spec. 4
CE43	markup_l_5	Firm markup - labor input decision - Spec. 5
CE44	markup_m_0	Firm markup - intermediate input decision - Spec. 0

Competition		
Variable Code	**Variable Name**	**Definition**
CE45	markup_m_1	Firm markup - intermediate input decision - Spec. 1
CE46	markup_m_2	Firm markup - intermediate input decision - Spec. 2
CE47	markup_m_3	Firm markup - intermediate input decision - Spec. 3
CE48	markup_m_4	Firm markup - intermediate input decision - Spec. 4
CE49	markup_m_5	Firm markup - intermediate input decision - Spec. 5
CE50	markup_ml_0	markup - labor + intermed. input decision - Spec. 0
CE51	markup_ml_1	markup - labor + intermed. input decision - Spec. 1
CE52	markup_ml_2	markup - labor + intermed. input decision - Spec. 2
CE53	markup_ml_3	markup - labor + intermed. input decision - Spec. 3
CE54	markup_ml_4	markup - labor + intermed. input decision - Spec. 4
CE55	markup_ml_5	markup - labor + intermed. input decision - Spec. 5
Ratios		
CR00	top_rev_sam_C	Top10 firms' share in revenues, country level
CR01	top_rev_sam_M	Top10 firms' share in revenues, macro-sector level

Competition		
Variable Code	Variable Name	Definition
CR02	top_rev_sam_2D	Top10 firms' share in revenues, 2-digit industry level
CR03	top_rev_sam_N	Top10 firms' share in revenues, NUTS2 level
CR04	top_ifa_sam_2D	Top10 firms' share in ifa, 2-digit indust. Level
CR05	top_ifa_sam_C	Top10 firms' share in intangible fixed assets, country level
CR06	top_ifa_sam_M	Top10 firms' share in intangible fixed assets, macro-sector level
CR07	top_ifa_sam_N	Top10 firms' share in intangible fixed assets, NUTS2 level
CR08	top_l_sam_2D	Top10 firms' labor share, 2-digit industry level
CR09	top_l_sam_C	Top10 firms' labor share, country level
CR10	top_l_sam_M	Top10 firms' labor share, macro-sector level
CR11	top_l_sam_N	Top10 firms' labor share, NUTS2 level
CR12	top_lc_sam_2D	Top10 firms' share in labor costs, 2-digit industry level
CR13	top_lc_sam_C	Top10 firms' share in labor costs, country level

Variable Code	Variable Name	Definition
		Competition
CR14	top_lc_sam_M	Top10 firms' share in labor costs, macro-sector level
CR15	top_lc_sam_N	Top10 firms' share in labor costs, NUTS2 level
CR16	top_rk_sam_2D	Top10 firms' share in real capital, 2-digit industry level
CR17	top_rk_sam_C	Top10 firms' share in real capital, country level
CR18	top_rk_sam_M	Top10 firms' share in real capital, macro-sector level
CR19	top_rk_sam_N	Top10 firms' share in real capital, NUTS2 level
CR20	top_rva_sam_2D	Top10 firms' share in real value added, 2-digit industry level
CR21	top_rva_sam_C	Top10 firms' share in real value added, country level
CR22	top_rva_sam_M	Top10 firms' share in real value added, macro-sector level
CR23	top_rva_sam_N	Top10 firms' share in real value added, NUTS2 level
		Values
CV12	hhi_ifa_sam_C	Hirschman-Herfindahl Index, intangible shares, country, sample
CV13	hhi_ifa_sam_M	Hirschman-Herfindahl Index, intangible shares, mac_sector, sample

	Competition	
Variable Code	**Variable Name**	**Definition**
CV14	hhi_ifa_sam_N	Hirschman-Herfindahl Index, intangible shares, NUTS2, sample
CV15	hhi_ifa_sam_2D	Hirschman-Herfindahl Index, intangible shares, industry, sample
CV12	hhi_ifa_pop_C	Hirschman-Herfindahl Index, intangible shares, country, population
CV13	hhi_ifa_pop_M	Hirschman-Herfindahl Index, int. fix.assets, mac_sector, population
CV14	hhi_ifa_pop_N	Hirschman-Herfindahl Index, intangible shares, NUTS2, population
CV15	hhi_ifa_pop_2D	Hirschman-Herfindahl Index, intangible shares, industry, population
CV20	hhi_l_sam_C	Hirschman-Herfindahl Index, employment shares, country, sample
CV21	hhi_l_sam_M	Hirschman-Herfindahl Index, employment shares, mac_sector, sample
CV22	hhi_l_sam_N	Hirschman-Herfindahl Index, employment shares, NUTS2, sample

Competition		
Variable Code	Variable Name	Definition
CV23	hhi_l_sam_2D	Hirschman-Herfindahl Index, employment shares, industry, sample
CV20	hhi_l_pop_C	Hirschman-Herfindahl Index, employment shares, country, population
CV21	hhi_l_pop_M	Hirschman-Herfindahl Index, emp. shares, mac_sector, population
CV22	hhi_l_pop_N	Hirschman-Herfindahl Index, employment shares, NUTS2, population
CV23	hhi_l_pop_2D	Hirschman-Herfindahl Index, employment shares, industry, population
CV28	hhi_lc_sam_C	Hirschman-Herfindahl Index, nom labor cost shares, country, sample
CV29	hhi_lc_sam_M	Hirschman-Herfindahl Index, nlc, mac_sector, sample
CV30	hhi_lc_sam_N	Hirschman-Herfindahl Index, nom labor cost shares, NUTS2, sample
CV31	hhi_lc_sam_2D	Hirschman-Herfindahl Index, nom labor cost shares, industry, sample
CV28	hhi_lc_pop_C	Hirschman-Herfindahl Index, nom labor cost shares, country, pop.

Competition		
Variable Code	**Variable Name**	**Definition**
CV29	hhi_lc_pop_M	Hirschman-Herfindahl Index, nom labor cost shares, mac_sector, pop.
CV30	hhi_lc_pop_N	Hirschman-Herfindahl Index, nom labor cost shares, NUTS2, pop.
CV31	hhi_lc_pop_2D	Hirschman-Herfindahl Index, nom labor cost shares, industry, pop.
CV04	hhi_rev_sam_C	Hirschman-Herfindahl Index, nom. revenue shares, country, sample
CV05	hhi_rev_sam_M	Hirschman-Herfindahl Index, nom. revenue shares, mac_sector, sample
CV06	hhi_rev_sam_N	Hirschman-Herfindahl Index, nom. revenue shares, NUTS2, sample
CV07	hhi_rev_sam_2D	Hirschman-Herfindahl Index, nom. revenue shares, industry, sample
CV04	hhi_rev_pop_C	Hirschman-Herfindahl Index, nom. revenue shares, country, pop.
CV05	hhi_rev_pop_M	Hirschman-Herfindahl Index, nom. revenue shares, mac_sector, pop.

Competition		
Variable Code	Variable Name	Definition
CV06	hhi_rev_pop_N	Hirschman-Herfindahl Index, nom. revenue shares, NUTS2, pop.
CV07	hhi_rev_pop_2D	Hirschman-Herfindahl Index, nom. revenue shares, industry, pop.
CV36	hhi_rk_sam_C	Hirschman-Herfindahl Index, real capital shares, country, sample
CV37	hhi_rk_sam_M	Hirschman-Herfindahl Index, real capital shares, mac_sector, sample
CV38	hhi_rk_sam_N	Hirschman-Herfindahl Index, real capital shares, NUTS2, sample
CV39	hhi_rk_sam_2D	Hirschman-Herfindahl Index, real capital shares, industry, sample
CV36	hhi_rk_pop_C	Hirschman-Herfindahl Index, real capital shares, country, pop.
CV37	hhi_rk_pop_M	Hirschman-Herfindahl Index, real capital shares, mac_sector, pop.
CV38	hhi_rk_pop_N	Hirschman-Herfindahl Index, real capital shares, NUTS2, pop.

Competition		
Variable Code	**Variable Name**	**Definition**
CV39	hhi_rk_pop_2D	Hirschman-Herfindahl Index, real capital shares, industry, pop.
CV44	hhi_rva_pos_sam_C	Hirschman-Herfindahl Index, rva shares, country, sample
CV45	hhi_rva_pos_sam_M	Hirschman-Herfindahl Index, rva shares, mac_sector, sample
CV46	hhi_rva_pos_sam_N	Hirschman-Herfindahl Index, real value-added shares, NUTS2, sample
CV47	hhi_rva_pos_sam_2D	Hirschman-Herfindahl Index, rva shares, industry, sample
CV44	hhi_rva_pos_pop_C	Hirschman-Herfindahl Index, real value-added shares, country, pop.
CV45	hhi_rva_pos_pop_M	Hirschman-Herfindahl Index, rva shares, mac_sector, pop.
CV46	hhi_rva_pos_pop_N	Hirschman-Herfindahl Index, real value-added shares, NUTS2, pop.
CV47	hhi_rva_pos_pop_2D	Hirschman-Herfindahl Index, real value-added shares, industry, pop.

5.2.2 Productivity Variables

Table 11 Productivity Variables

	Productivity	
Variable Code	**Variable Name**	**Definition**
	Estimates	
PEb0	tfp_0	TFP - Specification 0 (CD, industry-level cost shares)
PEb1	tfp_1	TFP - Specification 1 (CD, industry-level OLS)
PEb2	tfp_2	TFP - Specification 2 (TL, industry-level OLS)
PEb3	tfp_3	TFP - Specification 3 (CD, industry-year-level OLS)
PEb4	tfp_4	TFP - Specification 4 (CD, industry-level ACF)
PEb5	tfp_5	TFP - Specification 5 (TL, industry-level ACF)
PEb7	mpk_0	marg. product of capital - Specification 0
PEb8	mpk_1	marg. Product of capital - Specification 1
PEb9	mpk_2	marg. product of capital - Specification 2
PEc0	mpk_3	marg. Product of capital - Specification 3
PEc1	mpk_4	marg. Product of capital - Specification 4
PEc2	mpk_5	marg. Product of capital - Specification 5

Productivity		
Variable Code	**Variable Name**	**Definition**
PEc3	mpl_0	marg. product of labor - Specification 0
PEc4	mpl_1	marg. Product of labor - Specification 1
PEc5	mpl_2	marg. product of labor - Specification 2
PEc6	mpl_3	marg. Product of labor - Specification 3
PEc7	mpl_4	marg. Product of labor - Specification 4
PEc8	mpl_5	marg. Product of labor - Specification 5
PEc9	mpm_0	marg. product of intermediates - Specification 0
PEd0	mpm_1	marg. product of intermediates - Specification 1
PEd1	mpm_2	marg. product of intermediates - Specification 2
PEd2	mpm_3	marg. product of intermediates - Specification 3
PEd3	mpm_4	marg. product of intermediates - Specification 4

Productivity		
Variable Code	**Variable Name**	**Definition**
PEd4	mpm_5	marg. product of intermediates - Specification 5
PEd5	mrpk_0	marg. revenue product of capital - Specification 0
PEd6	mrpk_1	marg. revenue product of capital - Specification 1
PEd7	mrpk_2	marg. revenue product of capital - Specification 2
PEd8	mrpk_3	marg. revenue product of capital - Specification 3
PEd9	mrpk_4	marg. revenue product of capital - Specification 4
PEe0	mrpk_5	marg. revenue product of capital - Specification 5
PEe1	mrpl_0	marg. revenue product of labor - Specification 0
PEe2	mrpl_1	marg. revenue product of labor - Specification 1
PEe3	mrpl_2	marg. revenue product of labor - Specification 2
PEe4	mrpl_3	marg. revenue product of labor - Specification 3
PEe5	mrpl_4	marg. revenue product of labor - Specification 4
PEe6	mrpl_5	marg. revenue product of labor - Specification 5
PEe7	oe_k_0	output elasticity w.r.t. capital - Specification 0

	Productivity	
Variable Code	**Variable Name**	**Definition**
PEe8	oe_k_0_va	value-added elasticity w.r.t. capital - Specification 0
PEe9	oe_k_1	output elasticity w.r.t. capital - Specification 1
PEf0	oe_k_1_va	value-added elasticity w.r.t. capital - Specification 1
PEf1	oe_k_2	output elasticity w.r.t. capital - Specification 2
PEf2	oe_k_2_va	value-added elasticity w.r.t. capital - Specification 2
PEf3	oe_k_3	output elasticity w.r.t. capital - Specification 3
PEf4	oe_k_3_va	value-added elasticity w.r.t. capital - Specification 3
PEf5	oe_k_4	output elasticity w.r.t. capital - Specification 4
PEf6	oe_k_4_va	value-added elasticity w.r.t. capital - Specification 4
PEf7	oe_k_5	output elasticity w.r.t. capital - Specification 5
PEf8	oe_k_5_va	value-added elasticity w.r.t. capital - Specification 5
PEf9	oe_l_0	output elasticity w.r.t. labor - Specification 0
PEg0	oe_l_0_va	value-added elasticity w.r.t. labor - Specification 0
PEg1	oe_l_1	output elasticity w.r.t. labor - Specification 1

	Productivity	
Variable Code	Variable Name	Definition
PEg2	oe_l_1_va	value-added elasticity w.r.t. labor - Specification 1
PEg3	oe_l_2	output elasticity w.r.t. labor - Specification 2
PEg4	oe_l_2_va	value-added elasticity w.r.t. labor - Specification 2
PEg5	oe_l_3	output elasticity w.r.t. labor - Specification 3
PEg6	oe_l_3_va	value-added elasticity w.r.t. labor - Specification 3
PEg7	oe_l_4	output elasticity w.r.t. labor - Specification 4
PEg8	oe_l_4_va	value-added elasticity w.r.t. labor - Specification 4
PEg9	oe_l_5	output elasticity w.r.t. labor - Specification 5
PEh0	oe_l_5_va	value-added elasticity w.r.t. labor - Specification 5
PEh1	oe_m_0	output elasticity w.r.t. intermediates- Specification 0
PEh2	oe_m_1	output elasticity w.r.t. intermediates- Specification 1
PEh3	oe_m_2	output elasticity w.r.t. intermediates- Specification 2

Productivity		
Variable Code	**Variable Name**	**Definition**
PEh4	oe_m_3	output elasticity w.r.t. intermediates- Specification 3
PEh5	oe_m_4	output elasticity w.r.t. intermediates- Specification 4
PEh6	oe_m_5	output elasticity w.r.t. intermediates- Specification 5
PEh7	ps_0	MRPL-wage Gap (deflated) - Specification 0
PEh8	ps_1	MRPL-wage Gap (deflated) - Specification 1
PEh9	ps_2	MRPL-wage Gap (deflated) - Specification 2
PEi0	ps_3	MRPL-wage Gap (deflated) - Specification 3
PEi1	ps_4	MRPL-wage Gap (deflated) - Specification 4
PEi2	ps_5	MRPL-wage Gap (deflated) - Specification 5
PEi3	rts_0	RTS - Specification 0 (CD, industry-level cost shares)
PEi4	rts_1	RTS - Specification 1 (CD, industry-level OLS)
PEi5	rts_2	RTS - Specification 2 (TL, industry-level OLS)
PEi6	rts_3	RTS - Specification 3 (CD, industry-year-level OLS)

	Productivity	
Variable Code	**Variable Name**	**Definition**
PEi7	rts_4	RTS - Specification 4 (CD, industry-level ACF)
PEi8	rts_5	RTS - Specification 5 (TL, industry-level ACF)
PEi9	ln_tfp_0	Log(TFP) - Specification 0 (CD, industry-level cost shares)
PEj0G1	ln_tfp_1	Growth rate (from t-1): ln_tfp_1
PEj0	ln_tfp_1	Log(TFP) - Specification 1 (CD, industry-level OLS)
PEj1	ln_tfp_2	Log(TFP) - Specification 2 (TL, industry-level OLS)
PEj2	ln_tfp_3	Log(TFP)- Specification 3 (CD, industry-year-level OLS)
PEj3	ln_tfp_4	Log(TFP) - Specification 4 (CD, industry-level ACF)
PEj4	ln_tfp_5	Log(TFP) - Specification 5 (TL, industry-level ACF)
	Values	
PV00	kprod_va	Capital productivity, computed as rva/nk
PV01	lnkprod_va	Log capital productivity real value added based: ln(rva/nk)
PV02G1	lnlprod_rev	Growth rate (from t-1): log. labor prod., real revenue based

Productivity		
Variable Code	**Variable Name**	**Definition**
PV02	lnlprod_rev	Log labor productivity, real revenue based: ln(rrev/l)
PV03G1	lnlprod_va	Growth rate (from t-1): log. labor prod., real value-added based
PV03	lnlprod_va	Log labor productivity, real value added based: ln(rva/l)
PV04	lnsr	Log. Solow residual, weights in CD: labor 2/3, rk 1/3
PV05	lnsr_cs	Log. Solow residual, weights in CD from cost shares
PV06	lprod_rev	Labor productivity, real revenue based, computed as rrev/l
PV07	lprod_va	Labor productivity, real value added based, computed as rva/l
PV08	solowres	Solow residual, weights in CD: labor 2/3, rk 1/3
PV09	solowres_cs	Solow residual, weights in CD from cost shares

5.2.3 Labour Variables

Table 12 Labour Variables

		Labour
Variable Code	**Variable Name**	**Definition**
		Dummies
LD00	high_growth	D = 1, if firm had high employment growth in last 3 years
LD01	t10_l_C	D = 1, if Top10 firm by employee-number, country level
LD02	t10_l_M	D = 1, if Top10 firm by employee-number, macsec-level
LD03	t10_l_2D	D = 1, if Top10 firm by employee-number, sector-level
		Ratios
LR00	lc_rev	Ratio: wageshare: nom. labor cost / nom. Revenue
LR01	lc_va	Ratio: wageshare: nom. labor cost / nom. value-added
LR02	tertshare	Share of employees with tertiary education
LR03	ulc	Ratio: Unit labor costs: nom. labor cost / real value-added
		Values
LV21GH	firm	Haltiwanger-Davis-Schuh [27] growth rate (from t-1): labor.
LV21GH	firm_neg	Haltiwanger-Davis-Schuh growth rate (from t-1): labor, only negative

[27] See 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

Labour		
Variable Code	**Variable Name**	**Definition**
LV21GH	firm_pos	Haltiwanger-Davis-Schuh growth rate (from t-1): labor, only positive
LV24	Rwage	Real wage
LV00	avg_wage	Ratio: wage as average labor cost per employee and year (nlc/l)
LV01	jcr_pop_C	Job creation rates, country level, population
LV02	jcr_pop_M	Job creation rates, mac-sector level, population
LV03	jcr_pop_MS	Job creation rates, macsec-szclass level, population
LV04	jcr_pop_N	Job creation rates, NUTS2 level, population
LV05	jcr_pop_2D	Job creation rates, sector level, population
LV01	jcr_sam_C	Job creation rates, country level, sample
LV02	jcr_sam_M	Job creation rates, mac-sector level, sample
LV03	jcr_sam_MS	Job creation rates, macsec-szclass level, sample
LV04	jcr_sam_N	Job creation rates, NUTS2 level, sample
LV05	jcr_sam_2D	Job creation rates, sector level, sample
LV11	jdr_pop_C	Job destruction rates, country level, population

Labour		
Variable Code	**Variable Name**	**Definition**
LV12	jdr_pop_M	Job destruction rates, mac-sector level, population
LV13	jdr_pop_MS	Job destruction rates, macsec-szclass level, population
LV14	jdr_pop_N	Job destruction rates, NUTS2 level, population
LV15	jdr_pop_2D	Job destruction rates, sector level, population
LV11	jdr_sam_C	Job destruction rates, country level, sample
LV12	jdr_sam_M	Job destruction rates, mac-sector level, sample
LV13	jdr_sam_MS	Job destruction rates, macsec-szclass level, sample
LV14	jdr_sam_N	Job destruction rates, NUTS2 level, sample
LV15	jdr_sam_2D	Job destruction rates, sector level, sample
LV21G1	L	Growth rate (from t-1): labor = number of employees
LV21G3	L	Growth rate (from t-3): labor = number of employees
LV21	L	Labor: number of employees in headcounts
LV22	wage_premium_pop_2D	wage premium as % deviation from sector median, population

Labour		
Variable Code	**Variable Name**	**Definition**
LV23	wage_premium_sam_2D	wage premium as % deviation from industry median, sample

5.2.4 Trade Variables

Table 13 Trade Variables

Trade		
Variable Code	**Variable Name**	**Definition**
	Categorical	
TC00	exp_dest	Categorical variable indicating export destinations
TC01	exp_imp_rel	Categorical variable indicating the direction of trade
TC02	exp_time_3y	Categorical variable indicating the timing of exports
TC03	exp_top10	Categorical variable indicating large exporters
TC04	imp_dest	Categorical variable indicating the origins of imports
TC05	imp_time_3y	Categorical variable indicating the timing of imports
TC06	imp_top10	Categorical variable indicating large importers

Trade		
Variable Code	**Variable Name**	**Definition**
		Dummies
TD01	2w_exterior_adj	D = 1, if exEU exports & imports > inEU exp & imp, adj.
TD03	2w_extersale_adj	D = 1: exEU exp. > inEU exp. & exEU imp. < inEU imp., adj
TD07	2w_interior_adj	D = 1: inEU trade vol. (exp. & imp.) > exEU trade vol., adj.
TD09	2w_intersale_adj	D = 1: inEU exp. > exEU exp. & inEU imp. < exEU imp., adj
TD13	2w_total_adj	D = 1, if firm is twoway trader (exporting & importing), adj.
TD14	Exp	D = 1, if exporting
TD15	exp_adj	D = 1, if exporting, adj.
TD15	exp_adj_con2	D = 1, if exporting now and the year before, adj.
TD17	exp_adj_con3	D = 1, if 3 years consecutive exporter intra-EU (t-2, t-1, t), adj.
TD18	exp_adj_net	D = 1, if net exporter (exports>imports), adj.
TD19	exp_adj_new2	D = 1, if new exporter in t (and no exports in t-1), adj.
TD21	exp_adj_non2	D = 1, if not exporting (t-1, t), adj.
TD22	exp_adj_non3a	D = 1, if not exporting (t-1, t, t+1), adj.
TD23	exp_adj_stop1	D = 1, if exporter in t-1, but not in t, adj.

Trade		
Variable Code	**Variable Name**	**Definition**
TD24	exp_adj_stop3a	D = 1, if exporter in t-1 & t but not in t+1, adj.
TD26	t10_exp_adj_C	D = 1, if Top10 exporter, country level, adj.
TD27	t10_exp_adj_2D	D = 1, if Top10 exporter, sector level, adj.
TD30	exp_ex	D = 1, if exporting extra-EU
TD31	exp_ex_adj	D = 1, if exporting extra-EU, adj.
TD54	exp_in	D = 1, if exporting intra-EU
TD55	exp_in_adj	D = 1, if exporting intra-EU, adj.
TD88	Imp	D = 1, if importing
TD89	imp_adj	D = 1, if importing , adj.
TD90	imp_adj_con2	D = 1, if 2 years consecutive importer (t-1, t), adj.
TD93	t10_imp_adj_C	D = 1, if Top10 importer, country level, adj.
TD94	t10_imp_adj_2D	D = 1, if Top10 importer, sector level, adj.
TD97	imp_ex	D = 1, if importing extra-EU
TD98	imp_ex_adj	D = 1, if importing extra-EU, adj.
TDa7	imp_in	D = 1, if importing intra-EU
TDa8	imp_in_adj	D = 1, if importing intra-EU, adj.
TDc0	exp_adj_new3	D = 1, if new exporter in t (and no exports in both t-2 and t-1)

Trade		
Variable Code	**Variable Name**	**Definition**
TDc1	exp_adj_non3b	D = 1, if not exporting (t-2, t-1, t), adj.
TDc2	exp_adj_stop3b	D = 1, if exports in t-2, t-1 but not in t, adj.
TDc3	exp_val_swi	D = 1, if exports in t-1, but not in t-2 and not in t, adj.
TDc4	imp_adj_con3	D = 1, if 3 years consecutive importer (in t-2, t-1 and t), adj.
TDc5	imp_adj_new2	D = 1, if imports in t, but no imports in t-1
TDc6	imp_adj_new3	D = 1, if imports in t, but no imports in t-2 and t-1
TDc7	imp_adj_non2	D = 1, if no imports in t-1 and t
TDc8	imp_adj_non3a	D = 1, if no imports in t-1, t, and t+1
TDc9	imp_adj_non3b	D = 1, if non-importer (no imports in t-2, t-1, and t)
TDd0	imp_adj_stop3b	D = 1, if firm imported in both t-2 and t-1, but not in t, adj.
TDd1	imp_adj_swi	D = 1, if firm imported in both t-2 and t, but not in t-1, adj.
TDd2	imp_adj_stop1	D = 1, if imports in t-1, but not in t, adj.
Ratios		

	Trade	
Variable Code	**Variable Name**	**Definition**
TR00	exp_adj_pop_C	Ratio: exports, adj., share of total, country level, pop.
TR01	exp_adj_pop_2D	Ratio: exports, adj., share of total, sector level, pop.
TR02	exp_adj_rev	Ratio: Export Ratio: exports adj. / nom. Revenue
TR03	exp_adj_sam_C	Ratio: exports, adj., share of total, country level, sample
TR04	exp_adj_sam_2D	Ratio: exports, adj., share of total, sector level, sample
TR05	exp_adj_va_rev	Ratio: value added in export (adj.) revenue: exp*nva/nrev
TR36	imp_adj_pop_C	Ratio: imports, adj., share of total, country level, pop.
TR37	imp_adj_pop_2D	Ratio: imports, adj., share of total, sector level, pop.
TR38	imp_adj_rev	Ratio: import Ratio: imports adj. / nom. Revenue
TR39	imp_adj_sam_C	Ratio: imports, adj., share of total, country level, sample
TR40	imp_adj_sam_2D	Ratio: imports, adj., share of total, sector level, sample
TR67	imp_exp_adj	Ratio: import intensity = imp/exp, adj.
	Values	
TV02	exp	Exports
TV02G1	exp_val_adj	Growth rate (from t-1): adjusted exports

	Trade	
Variable Code	Variable Name	Definition
TV02G1	exp_val	Growth rate (from t-1): exports
TV03	exp_adj	Exports, adj.
TV04	exp_ex	Exports extra-EU
TV05	exp_ex_adj	Exports extra-EU, adj.
TV06	exp_in	Exports intra-EU
TV07	exp_in_adj	Exports intra-EU, adj.
TV08	imp	Imports
TV09	imp_adj	Imports, adj.
TV10	imp_ex	Imports extra-EU
TV11	imp_ex_adj	Imports extra-EU, adj.
TV12	imp_in	Imports intra-EU
TV13	imp_in_adj	Imports intra-EU, adj.

5.2.5 Finance Variables

Table 14 Finance Variables

	Finance	
Variable Code	Variable Name	Definition
	Estimates	
FC07	y_zombie_intcov_pos	Categorical: Duration of current spell as zombie
	Dummies	
FD00	absconstr	D = 1, if firm is absolutely credit constrained

Finance		
Variable Code	**Variable Name**	**Definition**
FD01	safe	D = 1, if firm is financially constrained
FD02	t10_rev_C	D = 1, if firm is among Top10 revenue firms, country level
FD03	t10_rev_M	D = 1, if firm is among Top10 revenue firms, mac-sector level
FD04	t10_rev_2D	D = 1, if firm is among Top10 revenue firms, sector level
FD05	zombie_intcov_pos	D = 1, if int. payed > op. profit > 0 & no high growth for 3 years
FD06	zombie_intcov	D = 1, if int. payed > op. profits & no high growth for 3 years
FD07	zombie_negprof	D = 1, if op. profits < 0 & no high labor growth for 3 years (BoE)
FD08	t10_lc_nom_C	D = 1, if Top10 nom. labor cost firm, country level
FD09	t10_rk_C	D = 1, if firm is among Top10 real capital firms, country level
FD10	t10_rva_C	D = 1, if firm is among Top10 real value added firms, country level
FD11	t10_ifa_M	D = 1, if Top10 intangible fixed assets, mac-sector level

Variable Code	Variable Name	Finance Definition
FD12	t10_lc_nom_M	D = 1, if Top10 nom. labor cost firm, mac-sector level
FD13	t10_rk_M	D = 1, if firm is among Top10 real capital firms, mac-sector level
FD14	t10_rva_M	D = 1, if firm is among Top10 real value added firms, macsec level
FD15	t10_ifa_2D	D = 1, if Top10 intangible fixed assets, sector level
FD16	t10_lc_nom_2D	D = 1, if Top10 nom. labor cost firm, sector level
FD17	t10_rk_2D	D = 1, if firm is among Top10 real capital firms, sector level
FD18	t10_rva_2D	D = 1, if firm is among Top10 real value added firms, sector level
FD19	t10_ifa_C	D = 1, if Top10 intangible fixed assets, country level
		Ratios
FR00	capcost_m	Ratio: capital cost / intermediate inputs
FR01	cash_ta	Ratio: cash / total assets
FR02	cashflow_ta	Ratio: cash flow / total assets
FR03	collateral_ta	Ratio: nominal capital / total assets
FR04	costcov_lc_m	Cost coverage rate 1 = nrev / nlc + nm

Finance		
Variable Code	**Variable Name**	**Definition**
FR05	costcov_all	Cost coverage rate 2 = nrev / nlc + nm + capcost
FR06	depr_ta	Ratio: depreciation / total assets
FR07	div_ta	Ratio: dividends / total assets
FR08	equity_debt	Ratio: equity / debt
FR09	equity_ta	Equity ratio: equity / total assets
FR10	fingap	Ratio: Financial gap: (nom. Investment (ninvest) - cashflow)/nrev
FR11	ifa_k	Ratio: nom. intangible fixed assets / nom. capital
FR12	inte_debt	Ratio: interest paid / 0.5*(debt(t-1)+debt(t)) (implicit rate)
FR13	inv_rev	Ratio: inventories / nom. revenue
FR15	lc_capcost	Ratio: nom. labor cost / nom. capital cost
FR17	lc_m	Ratio: nom. labor cost / nom. intermediate inputs
FR18	leverage	Ratio: Leverage: debt (long-term & short-term) / total assets
FR19	op_inte	Ratio: operating profits / interest payments

	Finance	
Variable Code	**Variable Name**	**Definition**
FR20	pcm_kfix	Price cost margin without capital costs, assuming fixed k
FR21	pcm_kvar	Price cost margin incl. capital cost
FR22	profitmargin	Ratio: Operating profits / nom. Revenue
FR23	rd_costs	Ratio: nom. R&D expenditure / total costs
FR24	rd_m	Ratio: nom. R&D expenditure / nom. intermediate inputs
FR25	rev_capcost	Ratio: nom. revenue / capital costs
FR26	rev_ener	Ratio: nom. revenue / nom. energy inputs
FR27	rev_lc	Ratio: nom. revenue / nom. labor cost
FR28	rev_lc_m	Ratio: nom. revenue / nom. labor cost + nom. intermediate inputs
FR29	rev_m	Ratio: nom. revenue / nom. intermediate inputs
FR30	rk_l	Ratio: capital intensity: real capital / labor
FR31	roa	Ratio: return on total assets = op. profit / 0.5*(ta(t-1)+ta(t))

Variable Code	Variable Name	Definition
		Finance
FR32	trade_credit	Ratio: accounts payable / total assets
FR33	trade_debt	Ratio: accounts receivable / total assets
FR34	va_ener	Ratio: nom. value-added / nom. energy inputs
FR35	va_rev	Ratio: nom. value-added / nom. revenue
FR36	ifa_rev	Share of intangible k to rev_nom
FR37	invest_k	Ratio of nominal investment to nominal capital
FR38	invest_rev	Ratio of nominal investment to nominal revenue
FR39	rd_share_rev	Ratio: nom. R&D expenditure / nominal revenue
		Values
FV14G1	rk	Growth rate (from t-1): real capital
FV14G3	rk	Growth rate (from t-3): real capital
FV31	rinvest_intan	real intangible investment
FV00	capcost	Capital cost = depr. + interest paid + imputed int. on equity
FV01	debt	Debt: Long-term and short-term debts
FV02	debt_fin	Debt: current + non-current liabilities - accounts payable

	Finance	
Variable Code	Variable Name	Definition
FV03	n_ener	nominal energy inputs (also abbreviated as nei)
FV04	nk	Nominal capital stock
FV04	nk	Nominal capital stock
FV05	nlc	Nominal labor costs
FV05	nlc	Nominal labor costs
FV06	nm	Nominal intermediate inputs
FV06	nm	Nominal intermediate inputs
FV07	nrd	nominal Research & Development expenditure
FV08G1	nrev	Growth rate (from t-1): nom. revenue
FV08	nrev	Nominal revenue
FV08	nrev	Nominal revenue
FV10	nva	nominal value-added, computed as nrev - nm
FV10	nva	nominal value-added, computed as nrev - nm
FV11	nva_pos	nominal value-added, computed as nrev - nm, only positive values
FV12	nvi	nominal variable inputs (i.e. labor & intermediate inputs)
FV13	rifa	real intangible fixed assets
FV14	rk	real capital
FV15	rlc	real labor costs
FV16	rm	real intermediate inputs
FV17	rrev	real revenue

	Finance	
Variable Code	**Variable Name**	**Definition**
FV18	rva	real value-added, computed as rev - m
FV19	rva_pos	real value-added, only positive values
FV20	ta	Total assets
FV21	y_zombie_intcov_pos	years designated as int. > prof. > 0 zombie (D_zombie_intcov_pos=1)
FV22	y_zombie_intcov	years designated as int > profits zombie (D_zombie_intcov=1)
FV23	y_zombie_negprof	years designated as negative profit zombie (D_zombie_negprof = 1)
FV24	etr	Effective tax rate
FV25	invest_intan	Nominal intangible investments
FV26	ninvest	Nominal investment
FV27	lc_nom_l	Ratio of nominal labor costs to labor
FV28	rcapcost	Real capital cost
FV29	rinvest	Real investment
FV30	rrd	Real R&D expenditure

5.2.6 Other Variables

Table 15 Other Variables

Variable Code	Variable Name	Definition
		Other
		Categorical
OC00	firm_age	`1 "0-2 years" 2 "3-5 years" 3 "6-25 years" 4 "more than 25 years"'
OC	legal	Categorical variable providing the legal form of the firm
		Dummies
OD00	exit	D = 1, if firm exits in t or t+1
OD01	firm_age_medium	D = 1, if medium aged firm (age > 5 & <= 25)
OD02	firm_age_new	D = 1, if new firm (age < 3)
OD03	firm_age_old	D = 1, if old firm (age > 25 years)
OD04	firm_age_young	D = 1, if young firm (age >=3 & <=5)
OD05	foreign_own	D = 1, if >50% of firm is owned by foreigner(s)
OD06	d_llc	D = 1, if firm with limited liability
OD07	publ_own	D = 1, if >50% of firm is owned by government
		Values
OV00	firm_age	Age of firm in years
OV01	firm_age_atexit	Age of exiting firm, in years
OV02	years_till_exit	Amount of years until firm exit

Table 16 List of Variables in the Unconditional Descriptive Files

Unconditional Variables
CE32_dm_0
CE33_dm_1
CE34_dm_2
CE35_dm_3
CE36_dm_4
CE37_dm_5
CE38_markup_l_0
CE39_markup_l_1
CE40_markup_l_2
CE41_markup_l_3
CE42_markup_l_4
CE43_markup_l_5
CE44_markup_m_0
CE45_markup_m_1
CE46_markup_m_2
CE47_markup_m_3
CE48_markup_m_4
CE49_markup_m_5
CE50_markup_ml_0
CE51_markup_ml_1
CE52_markup_ml_2
CE53_markup_ml_3
CE54_markup_ml_4
CE55_markup_ml_5
FD00_absconstr
FD01_safe
FD02_t10_rev_C
FD03_t10_rev_M
FD04_t10_rev_2D
FD05_zombie_intcov_pos
FD06_zombie_intcov
FD07_zombie_negprof
FD08_t10_lc_nom_C
FD09_t10_rk_C
FD10_t10_rva_C
FD11_t10_ifa_M
FD12_t10_lc_nom_M
FD13_t10_rk_M
FD14_t10_rva_M
FD15_t10_ifa_2D
FD16_t10_lc_nom_2D
FD17_t10_rk_2D
FD18_t10_rva_2D

FD19_t10_ifa_C
FR00_capcost_m
FR01_cash_ta
FR02_cashflow_ta
FR03_collateral_ta
FR04_costcov_lc_m
FR05_costcov_all
FR06_depr_ta
FR07_div_ta
FR08_equity_debt
FR09_equity_ta
FR10_fingap
FR11_ifa_k
FR12_inte_debt
FR13_inv_rev
FR15_lc_capcost
FR17_lc_m
FR18_leverage
FR19_op_inte
FR20_pcm_kfix
FR21_pcm_kvar
FR22_profitmargin
FR23_rd_costs
FR24_rd_m
FR25_rev_capcost
FR26_rev_ener
FR27_rev_lc
FR28_rev_lc_m
FR29_rev_m
FR30_rk_l
FR31_roa
FR32_trade_credit
FR33_trade_debt
FR34_va_ener
FR35_va_rev
FR36_ifa_rev
FR37_invest_k
FR38_invest_rev
FR39_rd_share_rev
FV00_capcost
FV01_debt
FV02_debt_fin
FV03_n_ener
FV04_nk
FV05_nlc

FV06_nm
FV07_nrd
FV08G1_nrev
FV08_nrev
FV10_nva
FV11_nva_pos
FV12_nvi
FV13_rifa
FV14G1_rk
FV14G3_rk
FV14_rk
FV15_rlc
FV16_rm
FV17_rrev
FV18_rva
FV19_rva_pos
FV20_ta
FV21_y_zombie_intcov_pos
FV22_y_zombie_intcov
FV23_y_zombie_negprof
FV24_etr
FV25_invest_intan
FV26_ninvest
FV27_lc_nom_l
FV28_rcapcost
FV29_rinvest
FV30_rrd
FV31_rinvest_intan
LD00_high_growth
LD01_t10_l_C
LD02_t10_l_M
LD03_t10_l_2D
LR00_lc_rev
LR01_lc_va
LR02_tertshare
LR03_ulc
LV00_avg_wage
LV21G1_l
LV21G3_l
LV21GH_firm
LV21GH_firm_neg
LV21GH_firm_pos
LV21_l
LV22_wage_premium_pop_2D
LV23_wage_premium_sam_2D

LV24_rwage
OD00_exit
OD01_firm_age_medium
OD02_firm_age_new
OD03_firm_age_old
OD04_firm_age_young
OD05_foreign_own
OD06_d_llc
OD07_publ_own
OV00_firm_age
OV01_firm_age_atexit
OV02_years_till_exit
PEb0_tfp_0
PEb1_tfp_1
PEb2_tfp_2
PEb3_tfp_3
PEb4_tfp_4
PEb5_tfp_5
PEb7_mpk_0
PEb8_mpk_1
PEb9_mpk_2
PEc0_mpk_3
PEc1_mpk_4
PEc2_mpk_5
PEc3_mpl_0
PEc4_mpl_1
PEc5_mpl_2
PEc6_mpl_3
PEc7_mpl_4
PEc8_mpl_5
PEc9_mpm_0
PEd0_mpm_1
PEd1_mpm_2
PEd2_mpm_3
PEd3_mpm_4
PEd4_mpm_5
PEd5_mrpk_0
PEd6_mrpk_1
PEd7_mrpk_2
PEd8_mrpk_3
PEd9_mrpk_4
PEe0_mrpk_5
PEe1_mrpl_0
PEe2_mrpl_1
PEe3_mrpl_2

PEe4_mrpl_3
PEe5_mrpl_4
PEe6_mrpl_5
PEe7_oe_k_0
PEe9_oe_k_1
PEf1_oe_k_2
PEf3_oe_k_3
PEf5_oe_k_4
PEf7_oe_k_5
PEf9_oe_l_0
PEg1_oe_l_1
PEg3_oe_l_2
PEg5_oe_l_3
PEg7_oe_l_4
PEg9_oe_l_5
PEh1_oe_m_0
PEh2_oe_m_1
PEh3_oe_m_2
PEh4_oe_m_3
PEh5_oe_m_4
PEh6_oe_m_5
PEh7_ps_0
PEh8_ps_1
PEh9_ps_2
PEi0_ps_3
PEi1_ps_4
PEi2_ps_5
PEi3_rts_0
PEi4_rts_1
PEi5_rts_2
PEi6_rts_3
PEi7_rts_4
PEi8_rts_5
PEi9_ ln_tfp_0
PEj0G1_ ln_tfp_1
PEj0_ ln_tfp_1
PEj1_ ln_tfp_2
PEj2_ ln_tfp_3
PEj3_ ln_tfp_4
PEj4_ ln_tfp_5
PV00_kprod_va
PV01_lnkprod_va
PV02G1_lnlprod_rev
PV02_lnlprod_rev
PV03G1_lnlprod_va

PV03_lnlprod_va
PV04_lnsr
PV05_lnsr_cs
PV06_lprod_rev
PV07_lprod_va
PV08_solowres
PV09_solowres_cs
TD01_2w_exterior_adj
TD03_2w_extersale_adj
TD07_2w_interior_adj
TD09_2w_intersale_adj
TD13_2w_total_adj
TD14_exp
TD15_exp_adj_con2
TD17_exp_adj_con3
TD18_exp_adj_net
TD19_exp_adj_new2
TD21_exp_adj_non2
TD22_exp_adj_non3a
TD23_exp_adj_stop1
TD24_exp_adj_stop3a
TD26_t10_exp_adj_C
TD27_t10_exp_adj_2D
TD30_exp_ex
TD31_exp_ex_adj
TD54_exp_in
TD55_exp_in_adj
TD88_imp
TD90_imp_adj_con2
TD93_t10_imp_adj_C
TD94_t10_imp_adj_2D
TD97_imp_ex
TD98_imp_ex_adj
TDa7_imp_in
TDa8_imp_in_adj
TDc0_exp_adj_new3
TDc1_exp_adj_non3b
TDc2_exp_adj_stop3b
TDc3_exp_val_swi
TDc4_imp_adj_con3
TDc5_imp_adj_new2
TDc6_imp_adj_new3
TDc7_imp_adj_non2
TDc8_imp_adj_non3a
TDc9_imp_adj_non3b

TDd0_imp_adj_stop3b
TDd1_imp_adj_swi
TDd2_imp_adj_stop1
TR00_exp_adj_pop_C
TR01_exp_adj_pop_2D
TR02_exp_adj_rev
TR03_exp_adj_sam_C
TR04_exp_adj_sam_2D
TR05_exp_adj_va_rev
TR36_imp_adj_pop_C
TR37_imp_adj_pop_2D
TR38_imp_adj_rev
TR39_imp_adj_sam_C
TR40_imp_adj_sam_2D
TR67_imp_exp_adj
TV02G1_exp_val_adj
TV02G1_exp_val
TV02_exp
TV03_exp_adj
TV04_exp_ex
TV05_exp_ex_adj
TV06_exp_in
TV07_exp_in_adj
TV08_imp
TV09_imp_adj
TV10_imp_ex
TV11_imp_ex_adj
TV12_imp_in
TV13_imp_in_adj

Table 17 List of Decomposition Variables

Foster decomposition Variables
CE32_dm_0_Wnlc
CE33_dm_1_Wnlc
CE34_dm_2_Wnlc
CE35_dm_3_Wnlc
CE36_dm_4_Wnlc
CE37_dm_5_Wnlc
CE38_markup_l_0_Wnlc
CE39_markup_l_1_Wnlc
CE40_markup_l_2_Wnlc
CE41_markup_l_3_Wnlc
CE42_markup_l_4_Wnlc
CE43_markup_l_5_Wnlc
CE44_markup_m_0_Wnm
CE45_markup_m_1_Wnm
CE46_markup_m_2_Wnm
CE47_markup_m_3_Wnm
CE48_markup_m_4_Wnm
CE49_markup_m_5_Wnm
CE50_markup_ml_0_Wnvi
CE51_markup_ml_1_Wnvi
CE52_markup_ml_2_Wnvi
CE53_markup_ml_3_Wnvi
CE54_markup_ml_4_Wnvi
CE55_markup_ml_5_Wnvi
FR05_costcov_all_Wntc
FR11_ifa_k_Wk_nom
FR23_rd_costs_Wntc
FR24_rd_m_Wnm
FR26_rev_ener_Wnen
FR27_rev_lc_Wnlc
FR28_rev_lc_m_Wnvi
FR30_rk_l_Wl
FR35_va_rev_Wnrv
FR36_ifa_rev_Wnrv
FR39_rd_share_rev_Wnrv
LR00_lc_rev_Wnrv
LR01_lc_va_Wnva
PEb0_tfp_0_Wrrv
PEb1_tfp_1_Wrrv
PEb2_tfp_2_Wrrv
PEb3_tfp_3_Wrrv
PEb4_tfp_4_Wrrv
PEb5_tfp_5_Wrrv

PEe7_oe_k_0_Wrrv
PEe8_oe_k_0_va_Wnva
PEe9_oe_k_1_Wrrv
PEf0_oe_k_1_va_Wnva
PEf1_oe_k_2_Wrrv
PEf2_oe_k_2_va_Wnva
PEf3_oe_k_3_Wrrv
PEf4_oe_k_3_va_Wnva
PEf5_oe_k_4_Wrrv
PEf6_oe_k_4_va_Wnva
PEf7_oe_k_5_Wrrv
PEf8_oe_k_5_va_Wnva
PEf9_oe_l_0_Wrrv
PEg0_oe_l_0_va_Wnva
PEg1_oe_l_1_Wrrv
PEg2_oe_l_1_va_Wnva
PEg3_oe_l_2_Wrrv
PEg4_oe_l_2_va_Wnva
PEg5_oe_l_3_Wrrv
PEg6_oe_l_3_va_Wnva
PEg7_oe_l_4_Wrrv
PEg8_oe_l_4_va_Wnva
PEg9_oe_l_5_Wrrv
PEh0_oe_l_5_va_Wnva
PEh1_oe_m_0_Wrrv
PEh2_oe_m_1_Wrrv
PEh3_oe_m_2_Wrrv
PEh4_oe_m_3_Wrrv
PEh5_oe_m_4_Wrrv
PEh6_oe_m_5_Wrrv
PEi9_ ln_tfp_0_Wrrv
PEj0_ ln_tfp_1_Wrrv
PEj1_ ln_tfp_2_Wrrv
PEj2_ ln_tfp_3_Wrrv
PEj3_ ln_tfp_4_Wrrv
PEj4_ ln_tfp_5_Wrrv
PV00_kprod_va_Wrk
PV06_lprod_rev_Wl
PV07_lprod_va_Wl
PV08_solowres_Wrva
PV09_solowres_cs_Wrva

Table 18 List of Decomposition Variables

OP Decomposition Variables
CE32_dm_0_Wnlc
CE32_dm_0_Wnrv
CE33_dm_1_Wnlc
CE33_dm_1_Wnrv
CE34_dm_2_Wnlc
CE34_dm_2_Wnrv
CE35_dm_3_Wnlc
CE35_dm_3_Wnrv
CE36_dm_4_Wnlc
CE36_dm_4_Wnrv
CE37_dm_5_Wnlc
CE37_dm_5_Wnrv
CE38_markup_l_0_Wnlc
CE38_markup_l_0_Wnrv
CE39_markup_l_1_Wnlc
CE39_markup_l_1_Wnrv
CE40_markup_l_2_Wnlc
CE40_markup_l_2_Wnrv
CE41_markup_l_3_Wnlc
CE41_markup_l_3_Wnrv
CE42_markup_l_4_Wnlc
CE42_markup_l_4_Wnrv
CE43_markup_l_5_Wnlc
CE43_markup_l_5_Wnrv
CE44_markup_m_0_Wnm
CE44_markup_m_0_Wnrv
CE45_markup_m_1_Wnm
CE45_markup_m_1_Wnrv
CE46_markup_m_2_Wnm
CE46_markup_m_2_Wnrv
CE47_markup_m_3_Wnm
CE47_markup_m_3_Wnrv
CE48_markup_m_4_Wnm
CE48_markup_m_4_Wnrv
CE49_markup_m_5_Wnm
CE49_markup_m_5_Wnrv
CE50_markup_ml_0_Wnrv
CE50_markup_ml_0_Wnvi
CE51_markup_ml_1_Wnrv
CE51_markup_ml_1_Wnvi
CE52_markup_ml_2_Wnrv
CE52_markup_ml_2_Wnvi
CE53_markup_ml_3_Wnrv

CE53_markup_ml_3_Wnvi
CE54_markup_ml_4_Wnrv
CE54_markup_ml_4_Wnvi
CE55_markup_ml_5_Wnrv
CE55_markup_ml_5_Wnvi
FR05_costcov_all_Wntc
FR11_ifa_k_Wnk
FR23_rd_costs_Wntc
FR24_rd_m_Wnm
FR26_rev_ener_Wnen
FR27_rev_lc_Wnlc
FR28_rev_lc_m_Wnvi
FR30_rk_l_Wl
FR35_va_rev_Wnrv
FR36_ifa_rev_Wnrv
FR39_rd_share_rev_Wnrv
LR00_lc_rev_Wnrv
LR01_lc_va_Wnva
PEb0_tfp_0_Wrrv
PEb1_tfp_1_Wrrv
PEb2_tfp_2_Wrrv
PEb3_tfp_3_Wrrv
PEb4_tfp_4_Wrrv
PEb5_tfp_5_Wrrv
PEe7_oe_k_0_Wrrv
PEe8_oe_k_0_va_Wnva
PEe9_oe_k_1_Wrrv
PEf0_oe_k_1_va_Wnva
PEf1_oe_k_2_Wrrv
PEf2_oe_k_2_va_Wnva
PEf3_oe_k_3_Wrrv
PEf4_oe_k_3_va_Wnva
PEf5_oe_k_4_Wrrv
PEf6_oe_k_4_va_Wnva
PEf7_oe_k_5_Wrrv
PEf8_oe_k_5_va_Wnva
PEf9_oe_l_0_Wrrv
PEg0_oe_l_0_va_Wnva
PEg1_oe_l_1_Wrrv
PEg2_oe_l_1_va_Wnva
PEg3_oe_l_2_Wrrv
PEg4_oe_l_2_va_Wnva
PEg5_oe_l_3_Wrrv
PEg6_oe_l_3_va_Wnva
PEg7_oe_l_4_Wrrv

PEg8_oe_l_4_va_Wnva
PEg9_oe_l_5_Wrrv
PEh0_oe_l_5_va_Wnva
PEh1_oe_m_0_Wrrv
PEh2_oe_m_1_Wrrv
PEh3_oe_m_2_Wrrv
PEh4_oe_m_3_Wrrv
PEh5_oe_m_4_Wrrv
PEh6_oe_m_5_Wrrv
PEi9_ln_tfp_0_Wrrv
PEj0_ln_tfp_1_Wrrv
PEj1_ln_tfp_2_Wrrv
PEj2_ln_tfp_3_Wrrv
PEj3_ln_tfp_4_Wrrv
PEj4_ln_tfp_5_Wrrv
PV00_kprod_va_Wrk
PV06_lprod_rev_Wl
PV07_lprod_va_Wl
PV08_solowres_Wrva
PV09_solowres_cs_Wrva

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, indicators of labour market imperfections/labour market power, 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 in this section before we describe the indicators derived from the recovered production function parameters.

We estimate several different types of production functions in gross-output (i.e. real revenues) at the two-digit sector-level. As input variables, we always use capital (i.e. fixed assets), labour (i.e. number of employees), and materials (i.e. intermediate inputs). Variables are deflated using deflators available in Eurostat that are specific for the country, sector (2-digit), and year of each observation. Revenues and materials are deflated using sectoral value-added deflators, while capital is deflated using sectoral capital-specific deflators.

The production function is estimated according to 6 different methodologies:

0. In specification 0, we assume a CD production function with constant return to scale (CRS) and no fixed costs, and derive the output elasticity of each input as the country-sector-year median cost-share (input expenditure over total costs).
1. In specification 1, we assume a CD production function (from this specification on, no estimation imposes CRS), and estimate the output elasticities using OLS with year fixed effects (FE).
2. In specification 2, we assume a TL production function with second-degree interactions, and estimate the output elasticities using OLS with year FE.
3. In specification 3, we assume a CD production, and estimate the output elasticities using OLS with year fixed effects (FE); unlike specifications 1 and 2, here, on top of the constant, output elasticities are also allowed to change over time by interacting the inputs with year dummies.
4. In specification 4, we assume a CD production function, and estimate the output elasticities following the two-step control function approach of Ackerberg, Caves and Frazer 2015 (ACF); year FE are implemented by demeaning every variable before estimation.
5. In specification 5, we apply the approach of ACF with year-demeaned variables while assuming a TL production function.

While the cost-based and the OLS approaches are straightforward, the ACF methodology of specifications 4 and 5 may deserve some further explanation. For simplicity, we only refer to the CD case (specification 4). Notice that the original specification of ACF is in value added while we adapt the methodology to a production function in gross output.

Output y_t is produced using capital k_t, labour l_t, and materials m_t according to the following equation (in logs).

(1) $y_t = b^k k_t + b^l l_t + b^m m_t + \omega_t + \varepsilon_t$

ω_t is the component of productivity observed by the firm but not by the econometrician and ε_t is an unobserved productivity shock.

ω_t follows an AR(1) process with a productivity shock ξ_t that is observed by the firm at the beginning of the period.

(2) $\omega_t = g\ \omega_{t-1} + \xi_t$

Substituting equation (2) in equation (1), we get:

(3) $y_t = b^k k_t + b^l l_t + b^m m_t + g\omega_{t-1} + \xi_t + \varepsilon_t$

Materials demand is determined after observing the shock ξ_t , through an increasing (invertible) function.

(4) $m_t = m(k_t, l_t, \omega_t)$

Once we substitute $\omega_t = m^{-1}(k_t, l_t, m_t)$ in equation (1), the output will depend on a combination of contemporaneous levels of inputs, that we call $\phi_t = \phi_t(k_t, l_t, m_t)$. The latter can be easily approximated regressing output on a polynomial in k_t, l_t, m_t (second-degree interactions are used in the CompNet code).

(5) $y_t = \phi_t + \varepsilon_t$

Combining equations (1) and (5), we can derive ω_t as:

(6) $\omega_t = \phi_t - b^k k_t - b^l l_t - b^m m_t$

Taking the lag of equation (6) and substituting it in equation (3), we get the following equation, with $e_t = \xi_t + \varepsilon_t$ as the residual.

(7) $y_t = b^k k_t + b^l l_t + b^m m_t + g\phi_{t-1} - \ gb^k k_{t-1} - gb^l l_{t-1} - gb^m m_{t-1} + e_t$

Notice that l_t and m_t are endogenous because they depend on ξ_t. Thus, estimation via OLS would produce biased estimates. Moreover, since g interacts with the output elasticities, we cannot implement a linear 2SLS strategy.

However, we can rely on a system of four moment-conditions based on the exogeneity of k_t, ϕ_{t-1}, l_{t-1}, m_{t-1}, to identify the four parameters of interest (b^k, b^l, b^m, g) using GMM.

Notice that this solution can be only implemented using materials and a proxy variable for ω_t and introducing the "control function" ϕ_t. Otherwise, ω_t would have been part of the residual and the lagged inputs would have been endogenous as well (because they depend on ω_{t-1}, which is also a component of ω_t).

In practice, estimation follows two steps. In the first step, we estimate $\phi_t$via OLS, regressing $y_t$on a second degree polynomial of the inputs k_t, l_t, m_t. As a second step, we plug the predicted level of $\phi_t$in equation (7) and run a GMM estimation based on the system of moment conditions just described.

For a translog production function, we follow the same procedure, but the functional form of the production function is:

(8) $y_t = b^k k_t + b^l l_t + b^m m_t + b^{k2} k^2{}_t + b^{l2} l^2{}_t + b^{m2} m^2{}_t + b^{kl} k_t l_t + b^{km} k_t m_t + b^{lm} l_t m_t + \omega_t + \varepsilon_t$

So, as there are more coefficients to estimate, we also need a higher number of instruments for the moment conditions: k_t, k_t^2, ϕ_{t-1}, l_{t-1}, m_{t-1}, l_{t-1}^2, m_{t-1}^2, kl_{t-1}, lm_{t-1}, ml_{t-1}.

The estimation of the production function coefficients (the 'betas') allows deriving the following set of indicators.

Output Elasticities (θ^m, θ^k, θ^l)

For the CD case, these are simply the coefficients of the production function:

(9) $\theta^m = b^m$

(10) $\theta^k = b^k$

(11) $\theta^l = b^l$

For the TL, these are given by:

(12) $\theta^m = b^m + 2b^{m2} m^2 + b^{km} k + b^{lm} l$

(13) $\theta^k = b^k + 2b^{k2} k^2 + b^{km} m + b^{kl} l$

(14) $\theta^l = b^l + 2b^{l2} l^2 + b^{lm} m + b^{kl} k$

Returns to Scale (RTS)

This is given by the sum of the output elasticities of all inputs:

(15) RTS = $\theta^m + \theta^k + \theta^l$

Total Factor Productivity (TFP)

TFP can be retrieved as the difference between the actual and predicted level of output (in logs):

(16) $tfp_t = y_t - (b^k k_t + b^l l_t + b^m m_t)$

(17) $tfp_t = y_t - (b^k k_t + b^l l_t + b^m m_t + b^{k2} k^2{}_t + b^{l2} l^2{}_t + b^{m2} m^2{}_t + b^{kl} k_t l_t + b^{km} k_t m_t + b^{lm} l_t m_t)$

Where the first and second lines stand for the CD and TL case, respectively, and we use estimated coefficients in the term in brackets.

Markups

Markups are generally defined as the ratio between the final good price and the marginal cost of production. In CompNet we estimate mark-ups using the methodology by De Loecker & Warzinsky (2012). According to this methodology, it is possible to retrieve a measure of markup for each flexible input of the production process, therefore we compute markup on intermediate input and labour input (μ_m, μ_l); moreover, we include a measure of labour market power from Dobbelaere & Mairesse (2013) (μ_{DM}):

(18) $\mu_{mt} = \theta^m \frac{p_t q_t}{v_t}$

(19) $\mu_{lt} = \theta^l \frac{p_t q_t}{w_t l_t}$

(20) $\mu_{DMt} = \frac{\mu_{lt}}{\mu_{mt}}$

Where $p_t q_t$ is nominal revenues (price times quantity), v_t nominal material cost and w_t nominal labour cost per worker.

Marginal Product

From the production function estimation we can also retrieve the Marginal Product (MP) of each production input:

(21) $$MP_k = \theta^k \frac{y_t}{k_t}$$

(22) $$MP_m = \theta^m \frac{y_t}{m_t}$$

(23) $$MP_l = \theta^l \frac{y_t}{l_t}$$

Marginal Revenue Product

Similarly to the previous indicator, we also compute the Marginal Revenue Product of capital and labour (MRPK and MRPL) as in Mertens (2020).

(24) $$MRPK_t = \frac{\theta^k \frac{p_t q_t}{c_t}}{\mu_{mt}} rc_t$$

(25) $$MRPL_t = \mu_{DMt} \frac{w_t}{HICP_t}$$

Where in (24), c_t and rc_t are respectively nominal and real capital cost, and the other variables are the same as those introduced in earlier. Similarly, in (25) $HICP_t$ stands for the Harmonised Index of Consumer Prices.

Petrin-Sivadasan Gap

We compute the Petrin-Sivadasan gap (PS_t) as constructed in Petrin & Sivadasan (2013):

(26) $$PS_t = \left| \theta^l \; \frac{p_t q_t - w_t}{l_t} \right| rc_t$$

Value Added approximated output elasticities

Finally, we produce a measure of production function elasticity adjusted by the inverse of the ratio of nominal value added over nominal turnover:

(27) $$\theta^i{}_{VA} = \theta^i \frac{p_t q_t}{va_t}$$

We compute this measure for each of the production function inputs, therefore $i = k, l, m$.

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 firm size and productivity of firm i at time t, respectively, $\theta_{st} = \sum_{i \in S}(\theta_{it} - \bar{\theta})(\omega_{it} - \bar{\omega}_{st})$ bars indicate unweighted means of variables.

The decomposition splits the weighted average of firm productivity into two components: the unweighted mean and the covariance between productivity and firm 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.

Note that for defining firm size, we always apply denominator weights, i.e. in case of labour productivity size is defined by the labour input, whereas in case of TFP, size is defined in terms of gross output.

Table 19 Overview of Decompositions

op_decomp_	country_	weighted_ or unweighted_
	mac_sector_	weighted_ or unweighted_
	nuts2_	weighted_ or unweighted_
	industry2d_	weighted_ or unweighted_
foster_decomp_	country_	weighted_
	mac_sector_	weighted_
	nuts2_	weighted_
	Industry2d_	weighted_

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}, size shares in the respective aggregation level 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, industry 2-digits, 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.

Hsieh-Klenow Indicator (Hsieh and Klenow 2009)

Building on the work of Hsieh and Klenow (2009), we also provide the dispersion of marginal revenue products in our unconditional data files, 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 industry 2-digits, macro-sector, and country level standard deviation of marginal revenue products.

5.3.3 Further decompositions

In this vintage, we apply the static productivity decomposition of Olley & Pakes (1996), as explained above, to further variables, including labour shares and markups. This provides data users with size-weighted aggregates of these variables and allows to understand whether changes in these aggregates are due to changes in unweighted means of variables or due to changes in the covariance between firm size and firm-level values of the variable of interest.

5.3.4 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.

Variable:	*Zombie_intcov_pos*
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*
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.

5.3.5 How to compare productivity across industries, sectors, regions, and countries

The CompNet data provides variables measuring productivity at the industry, sector, region, and country level. These productivity measures can be divided into production-function-based measures and productivity variables directly calculated from the data. When comparing these estimates across aggregation levels (industries, sectors, regions, and countries) in CompNet, several aspects have to be considered.

As all production functions are estimated separately for the two-digit industry level in CompNet, the parameters of the production function vary between industries. This induces cross-industry variation in productivity variables derived from these production functions that does not results from true productivity differences between industries, but rather from

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

differences in production function parameters (the production technology) of an industry. This makes it impossible to compare levels of production-function-based productivity variables across industries.[29] We therefore recommend to use non-production-function-based variables for comparing productivity levels across industries, like our labour productivity variable.

A way to mitigate these issues of comparing levels between industries is to rely on comparing percentage changes of productivity between industries.[30] If the production function is time-constant, this will eliminate level shifts in productivity between industries due to differences in industry-specific production functions. For time-varying production functions (e.g. the time-varying Cobb-Douglas, or cost shares), productivity will, however, still exhibit jumps between industries due to changes in the production processes of industries that are unrelated to changes in true productivity.

Hence, when it comes to comparing changes, we recommend to either use productivity measures that are not based on production function estimates or productivity variables based on production function estimates with time-constant parameters.

Note that these issues of comparability do not extent to monetary and dimensionless variables that are derived from the production function, e.g. markups or marginal revenue products. Take markups as an example. Although specific production technologies might be associated with higher or lower markups, such markup differences, as opposed to associated differences in total factor productivity, reflect differences in true markups.

Take a high tech industry that manufactures aircrafts and an industry that manufactures ready-mixed concrete as an example. To manufacture one aircraft the firm will need an enormous amount of labour and capital compared to a firm manufacturing one package of ready-mixed concrete. As total factor productivity levels are related to the definition of output units, the aircraft firm will have a much lower total factor productivity than the firm producing ready-mixed concrete. This productivity difference is, however, not informative on differences

[29] Note that also the industry-specific cost-share based Solow residual is based on a production function. The simpler Solow residual based on fixed cost shares of 1/3 and 2/3 for capital and labour can, however, be compared across industries as here the parameters of the production function (1/3 and 2/3) are identical across all industries. Yet, the latter involves the strong assumption that output elasticities are constant and identical across all firms within Europe.

[30] This will only solve the issue of comparing productivity across industries for productivity estimates based on time-constant production functions.

in technical efficiency between these two firms. Yet, markup differences between these two firms will still correctly measure the percentage difference between prices and marginal costs for both firms, and data users can correctly assess differences in price setting market power between these firms. This is because the markup in CompNet is defined as the ratio between prices and marginal costs and is not related to the output unit, nor to the level of costs (as it is a ratio and not a difference between prices and marginal costs).

Due to these comparability issues of production-function-based productivity variables, the 8th vintage of the CompNet data does not report production-function-based productivity variables beyond the industry level. Hence, for higher aggregation levels, the CompNet data does only contain non-production-function-based productivity variables.

5.3.6 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 the value 1 if 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)[31] 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 specifically, the regression equation is the following:

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

[31] 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, industry, 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 are 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.

While the resulting *safe* variable itself is a binary dummy, the dataset reports its mean, which gives 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 *safe* credit constrained indicator, the dataset reports the mean of the *absconstr* binary variable, giving 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.

5.3.7 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.[32] 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

[32] 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 (see Section 5.3.1). Arguably, the most sophisticated version of our markup estimates is the one based on the translog production function. 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). When using our markup estimates we also recommend having a look at cost-share based estimates of output elasticities and markups and the non-parametric competition indicators that we provide. The latter contain price-cost margins, Hirschman-Herfindahl indices, and profit margins.

Labour Market Power

Following a recent stream of work (e.g. Dobbelaere & Mairesse (2013), Mertens (2020)), 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).

5.3.8 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) $X_{it} = 0.5 \cdot (E_{it} + E_{it-1})$ 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 industry 2-digits, 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) $Firm\ weight = \frac{X_{it}}{X_{st}}$ and the weighted growth rate is $Firm\ weight \cdot g_{it}$

where X_{st} is the average employment for a particular sector. Therefore, at the two-digit 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, the test for statistical dominance, 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 8th 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:

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

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

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 8th Vintage of the CompNet dataset uses a new weighting procedure which includes population weights derived from business registries to calculate the appropriate weights. Weights are based on the number of firms in a given industry and size class.

To illustrate the general weighting procedure[33], 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

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

with *firms_y_z* as the number of firms (of a certain year) of size class y and industry 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, industry and year. The sum of the weights (= *variable*_sum_weights[34] in the output dataset) is then

$$(2) \qquad V = \sum_{i=1}^{n} v_i \ .$$

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

$$(3) \qquad w_i = v_i \, \frac{n}{V} \, .$$

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

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

The weighted sample variance s^2 is

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

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

[33] 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.

[34] 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.

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

Consequently, weighted skewness is defined as

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

and the weighted kurtosis as

$$(8) \qquad m_4 / (m_2)^2.$$

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 $W_i > P$.

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

5.4.4 Data Sources

Table 20 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	Business Registry	FINA	Financial Agency Croatia	Croatian National Bank	All (census)	Almost all raw data are from FINA
	Court Registry	FINA	Financial Agency Croatia			Birth_year and death_year
Czech Republic	P5-01 survey	P501	Czech Statistical Office	Czech National Bank		Annual CZSO survey of businesses, used for compiling structural business statistics. NACE2 see RES, below.
	Register of Economic Subjects	RE5	Czech Statistical Office	Czech National Bank		Own NACE2 concordance system – years 2005-2007 backfilled based on simultaneous classification in 2008 in source dataset
	Foreign trade dataset	FT	Czech Statistical Office	Czech National Bank		Firm-by-product-by-destination data on imports and exports, based on customs or Intrastat declarations
Denmark	Accounts statistics	FIRE	Statistics Denmark	Central Bank of Denmark	Link	NACE2 Classification provided by Statistics Denmark
	General enterprise statistics	FIRM	Statistics Denmark		Link	

Finland	Structural business and financial statement statistics data	SBS	Statistics Finland	Statistics Finland	all	Breaks in 2006, 2013 Own NACE2 concordance used
	International trade statistics data	ITS	Statistics Finland	Finnish Customs	enterprises trading goods	Intrastat thresholds
	Employment statistics data	FOLK	Statistics Finland	Finnish Centre for Pensions, Statistics Finland	all	
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	Amtliche Firmendaten in Deutschland	AFID	Destatis	Federal Statistical Office of Germany and Federal Statistical Offices of the German Länder	NFC drawn from total economy	Manufacturing: only firms with more than 19 employees.
	Kostenstrukturerhebung im Bauhaupt- und Ausbaugewerbe					Firms with at least 17.5K
	Jahreserhebung der Gastgewerbestatistik					
	Jahreserhebung der Handelsstatistik					
	Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und					

	Gewinnung von Steinen und Erden					
Hungary	Tax registry database of National Tax and Customs Administration	NAV	National Tax and Customs Administration	Central Bank of 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			
	Export-Import data of Hungarian Enterprises	Külker	Statistics Hungary			
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 of Lithuania	NFC drawn from total economy	
	Business Register	BR	Centre of Register			
	Customs declaration	CU	Customs of the Republic of Lithuania			
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	Report on revenues, costs and financial result as well as on expenditure on fixed assets	F-01	Statistics Poland	Central Bank Poland		Compnet provided correspondence code used
	Annual enterprise survey	SP	Statistics Poland			Compnet provided correspondence code used
Portugal	Integrated Business Account System	SCIE/IBAS	Statistic Portugal	GEE – Ministry of Economic Portugal	NFC drawn from total economy	
Romania	Balance sheet information on non-financial enterprises	Bal. Sheet	Ministry of Public Finances	National Bank of 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. NACE2 concordance: Own system
	Statistical register of organizations	Register	Statistics Slovakia	National Bank of Slovakia	NFC drawn from total economy	Exclusion of firms with less than 20 employees. NACE2 concordance: Own system
	Foreign trade statistics	Customs	Statistics Slovakia	National Bank of Slovakia	NFC drawn from total economy	Exclusion of firms with less than 20 employees. NACE2 concordance: Own system
Slovenia	Agency of the Republic of Slovenia for Public Legal Records and Related Services	AJPES (Link)	Institute of Macroeconomic Analysis and Development of the Republic of Slovenia (IMAD)	Institute of Macroeconomic Analysis and Development of the Republic of Slovenia (IMAD)	Only considering Companies data (100 % of them; not the whole Business Register)**; for	

					the Period: 2002-2019	
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	NCF drawn from total economy	
	International trade in goods	ITGS	Statistics Sweden			
	Business register	BR	Statistics Sweden			
	Business Demography	BD	Statistics Sweden	Statistics Sweden		
	Labour statistics based on administrative sources	RAMS	Statistics Sweden	Statistics Sweden / Tax Authority		
Switzerland						

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 23.

Table 21 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!) + intangible fixed assets (see definition below; acquired - not developed in-house - intellectual property (patents, licences, copyrights, trademarks) and goodwill)	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	Basically all fixed assets (yearly average), that could not be subsumed under tangible fixed assets or intangible fixed assets: contains mainly long-term financial assets such as shares in affiliated enterprises, loans to affiliated enterprises, stocks, securities or bonds held not for immediate sale and unpaid capital	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. These include cash means, bank accounts, marketable securities, commercial paper, treasury bills and short-term government bonds with a maturity date of three months or less.	Cash and cash equivalents at a particular point in time	

Variable	First Best	Second Best	Third Best
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 in the ordinary course of business (work in process), and materials and supplies that are consumed in production (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.	
Other current assets	Basically all current assets (yearly average) that could not be subsumed under accounts receivables and inventories; contains for example cash and cash equivalent (see definition above), 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 (non-current assets) (yearly average) and should 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	Basically 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: A liability shall be classified as current when it satisfies any of the following criteria: (a) it is expected to be settled in the entity’s normal operating cycle; (b) it is held primarily for the purpose of being ; traded; (c) it is due to be settled within twelve months after the balance sheet date; or (d) the entity	Current liabilities at a particular point in time	short-term debt + accounts payable

Variable	First Best	Second Best	Third Best
	does not have an unconditional right to defer settlement of the liability for at least twelve months after the balance sheet date. All other liabilities shall be classified as non-current. Should be equal to short-term debt + accounts payable + other current liabilities.		
Total debt	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 + 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	Basically all current liabilities (yearly average) that could not be subsumed under short-term debt and accounts payable: current income tax liabilities, provisions, advance payments received from customers, outstanding wages, outstanding social security contributions etc.; should be equal to current liabilities - short-term debt - accounts payable	Other current liabilities at a particular point in time	
Gross Output	Gross output includes: 1) Turnover at factor cost: gross sales revenues minus customer discounts, returns and allowances; excluding indirect taxes but including subsidies on products and production. (Sales include: revenues from selling manufactured finished- or semi-finished goods, revenues from selling goods bought for resale, and revenues from services offered.) + 2) increase in the stock/inventory of manufactured finished - or semi-finished products + 3) Capitalized internal activities, i.e. increase in the value of total assets by construction of own machinery, self-constructed buildings or other self-constructed investment goods (excluding software, licenses, patents, copyrights developed in-house). This definition does not include other non-financial revenues (e.g. revenues from liquidating reserves, unexpected payments of demands that have been already written off etc. or revenues from selling tangible or intangible non-financial	Valued at market prices	

Variable	First Best	Second Best	Third Best
	fixed assets). Furthermore, financial revenues are also excluded.		
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, including all costs incurred from hiring labour, i.e. social security contributions, payroll taxes, benefits… - should be equal to position "employee benefits expense" in the statement of profit and loss (nature of expense method!). If possible, do not include share payment systems or payments to non-active staff (e.g. pension payments).	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. Definition includes all expenses for raw materials and consumables, expenses for components, expenses for energy, expenses for goods intended for resale and expenses for hired services. (If items from income statement are used: expenses for purchased materials and hired services only according to the classification of expenses by nature method.)	Intermediate inputs valued at market prices	
Energy Input	Sub-item of intermediate inputs; all expenses of the firm for energy covering all sorts of fuels, heat or electricity (e.g. solid fuels like coal or wood, liquid fuels like gasoline, gas fuels like natural gas). It should refer to operating expenses, ideally excluding expenditures for further resale or expenditures used as inputs for further production (e.g. coke from coal or ammonia from natural gas).		
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)	IAS 1.92 EBIT (Earnings Before Interest and Taxes) according to the "cost of goods sold approach" = Revenues - Costs of goods sold + Other income -Distribution costs -Administrative expenses - Other expenses; IAS 1.91 EBIT according to the "nature of expense method" = Revenue + Other income +/- Changes in inventories of finished goods and work in progress - Raw materials and consumables used - Employee benefits	revenues (turnover) - intermediate inputs - labour cost - depreciation	

Variable	First Best	Second Best	Third Best
	expense - Depreciation and amortisation expense - Other expenses (including purchased services)		
Interest paid and financial charges	All interest payable on any borrowings, i.e. bonds, loans, convertible debt or lines of credit		
Depreciation	Includes depreciation (ordinary or extraordinary) of the capital variable, i.e. depreciation of fixed tangible assets and depreciation/amortization of intangible fixed assets. Variable does not include depreciation/impairment of financial (non-current) 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 [e.g. interest received] - financial costs [e.g. interest paid] +/- equity in earnings of subsidiaries	Operating profit/loss- interest paid + 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		
Gross Investment	Total gross investment (tangible and intangible fixed assets) of a firm = Total gross increase in the value of tangible and intangible fixed assets during the calendar year. This includes the total value of acquired or self-constructed land, machinery, equipment, buildings and other durables (including assets under construction; does not include long-term financial assets!) plus the acquisitions of intangible fixed assets (acquired - not developed in-house - intellectual property like copyrights, patents, licenses, software etc.)		

Variable	First Best	Second Best	Third Best
Effective Tax Rate	Ratio of corporate taxes on pre-taxes income		
	Trade		
Export value	Exports valued at factor cost: Nominal export turnover (see definition of turnover; unadjusted exports) excluding indirect taxes, tariffs etc., but including subsidies on products and production. (The unadjusted value represents the value from the balance sheet or customs source that depending on the source may already be adjusted by the country specific annual threshold, but not the country specific maximum threshold that will be applied by the code.)	Valued at market prices: including	
Exports to extra-EU	Valued at factor costs: Nominal export turnover (unadjusted exports) outside EU (see definition of exports and turnover) excluding indirect taxes, tariffs etc., but including subsidies on products and production. (The unadjusted value represents the value from the balance sheet or customs source that depending on the source may already be adjusted by the country specific annual threshold, but not the country specific maximum threshold that will be applied by the code.)	Valued at market prices	
Exports to intra-EU (new in CompNet)	Valued at factor costs: Nominal export turnover (unadjusted exports) within EU (see definition of exports and turnover) excluding indirect taxes, tariffs etc., but including subsidies on products and production. (The unadjusted value represents the value from the balance sheet or customs source that depending on the source may already be adjusted by the country specific annual threshold, but not the country specific maximum threshold that will be applied by the code.)	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 include purchases of goods intended for resale.	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. Imports include purchases of goods intended for resale. Note that the sum of intra- and extra-EU imports should be equal to the total import value	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. Imports include purchases of goods intended for resale. Note	Valued at market prices	

Variable	First Best	Second Best	Third Best
	that the sum of intra- and extra-EU imports should be equal to the total import value		
	Other		
Industry 2digits	Two-digit division number according to NACE Rev. 2		
NUTS2	Four-digit code (combination of country and region) according to *Commission Regulation (EU) 2016/2066 of 21 November 2016 amending the annexes to Regulation (EC) No 1059/2003 of the European Parliament and of the Council on the establishment of a common classification of territorial units for statistics (NUTS)*		
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 (e.g. nationalised firms, publicly owned firms, state or local authority monopolies); unknown = missing.		
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		

Variable	First Best	Second Best	Third Best
	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 22 Country specific Definitions of Input Variables

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

Variable / Country	BE*	HR*	CZ	DK	FI	FR	DE	HU	IT*	LT	NL	PL	PT	RO	SK	SI	ES	SE	CH
Interest paid and financial charges	1	1	1	1	1	1	1	1	Other	1		1	1		1	1	1	1	
Depreciation	Other	1	1	2	1	1	1	1	Other	1		1	1		1	1	2	1	
Profits and losses before taxes	Other	2	1	2	1	1	1	1	Other	2		1	1		2	1	Other	1	
Cash flow (from profit/loss statement)	Other	3	1	3	1	3	1	1	0	1		3	1		3	1	Other	3	
Dividends	0	0	1	1	1	1	0	1	0	1		N/A	Unknowns		1	N/A	0	1	
Gross Investment		1	1	1	1	2	1	1	1	1		1	1		1	N/A		1	
Effective tax rate		1	1	0	1	1	1	1	1	1		1	1		1	1		1	
Export value	0	1	1	2	1	1	1	1	0	1		1	Other		1	1	0	1	
Exports to extra-EU	0	0	1	0	1	1	0	0	0	1		N/A	Other		1	1	0	1	
Exports to intra-EU	0	0	1	0	1	1	0	0	0	1		N/A	Other		1	1	0	1	
Import value	0	1	1	2	1	1	0	0	0	1		1	Other		1	N/A	0	1	
Imports from extra-EU	0	0	1	0	1	1	0	0	0	1		N/A	Other		1	N/A	0	1	
Imports from intra-EU	0	0	1	N/A	1	1	0	0	0	1		N/A	Other		1	N/A	0	1	
Industry	1	1	1	1	1	1	1	1	1	1		1	1		1	Other	1	1	
nuts2	1	1	1	1	1	1	1	1	1	1		0	1		1	Other	1	1	
Firm's birth year	Other	1	1	1	N/A	1	1	1	Other	1		N/A	Other		1	Other	Other	1	
Firm's exit year	1	1	1	1	N/A	0	0	1	1	0		N/A	Other		1	N/A	1	1	
Foreign ownership	0	1	1	1	1	0	0	1	0	0		1	Other		1	N/A	0	1	
Labour	1	2	1	2	1	1	2	1	Other	1		2	1		1	3	1	2	
Legal form	1	1	1	1	1	0	0	1	1	1		N/A	1		1	N/A	1	1	
Public ownership	0	1	1	N/A	1	0	0	2	0	0		1	0		1	N/A	0	1	
Share of skilled labour	0	0	N/A	N/A	0	0	0	0	0	0		N/A	0		0	N/A	0	1	

Notes: 0: “not available” – 1: “first-best definition”according to Table 22 – 2: “second-best definition” – 3: “third-best definition”[35]

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

5.4.6 List of Macro Sectors and Industries

Table 23 List of Macro-Sectors and Industries Included in the 8th Vintage

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Industry 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
			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
			25	Manufacture of fabricated metal products, except machinery and equipment
			26	Manufacture of computer, electronic and optical products

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Industry in CompNet	Description
C	1	Manufacturing	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
			52	Warehousing and support activities for transportation
			53	Postal and courier activities
I	5		55	Accommodation

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Industry in CompNet	Description
		Accommodation and food service activities	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
N	9	Administrative and support service activities	77	Rental and leasing activities
			78	Employment activities

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Industry in CompNet	Description
			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

5.5 Approaches to measure allocative efficiency

The CompNet database includes provides estimates of allocative efficiency based on three different measurement approaches. They are described in detail in the following subsections.

5.5.1 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]

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. two-digit industry-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.

5.5.2 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

[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)).

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.

5.5.3 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:

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

where s_{it} is the share of economic activity. Ω_{ij}, $\overline{\omega}_{jt}$, 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 $\overline{\omega}_{jt}$ 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.

[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).

6. References

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. Econometrica, 83(6), 2411-2451.

Boisvert, R. N. (1982) The Translog Production Function: Its properties, its several interpretations and estimation problems. *Department of Agricultural Economics, Cornell University*: A.E. Res. 82-28

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 Economic 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, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," Econometrica, 64, 1263–1295.

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.