CompNet The Competitiveness Research Network

User Guide for the 9th Vintage of the CompNet Dataset

Last Update: June 2023.

When using the data, please cite it as:

CompNet 2022, User Guide for the 9th Vintage of the CompNet Dataset.

This User Guide has been updated by Matthias Mertens, Peter Haug, Tommaso Bighelli, Jonathan Deist, and Ahmed Elhefnawy (all Halle Institute for Economic Research).
The working group on the CompNet 9th Vintage Data Collection has been coordinated by Matthias Mertens and includes: Javier Miranda, Peter Haug, Jonathan Deist, Tommaso Bighelli, Mirja Hälbig, Verena Plümpe (all Halle Institute for Economic Research), Alessandro Zona Mattioli (VU Amsterdam), and Bernardo Mottironi (London School of Economics).

Table of Content

How to Become a User 6

1. Information Included in the User Guide 7

2. The 9th Vintage CompNet Dataset 8

2.1 Sample, Time Range and Levels of Aggregation 8

2.2 Structure of the 9th Vintage of the CompNet Dataset 12

2.2.1 Descriptives 13

2.2.2 Joint Distributions 15

2.2.3 Transition Matrices 22

2.3 Naming Convention of Variables 23

3. Important Notes on Using the Dataset 28

3.1 CompNet Dataset and National Accounts 28

3.2 General Notes 29

3.2.1 Sample and Population Figures 29

3.2.2 Dummy Variables 30

3.2.3 Categorical Variables 30

3.3 Comparability 30

3.3.1 Countries and Source Data 31

3.3.2 Some Country-specific Issues 34

3.3.3 Variables and Indicators 36

4. Differences between 8th Vintage and 9th Vintage of the CompNet Dataset 37

5. Appendix 39

5.1 List of Data Folders 39

5.2 Detailed Variable Overview 40

5.2.1 Competition Variables 40

5.2.2 Finance Variables 65

5.2.3 Labor Variables 72
5.2.4 Productivity Variables 76
5.2.5 Trade Variables 84
5.2.6 Other Variables 91
5.3 Derivation of Indicators (More Complex Variables) 109
5.3.1 Production Function Estimation, TFP, and Marginal Products 109
5.3.2 Allocative Efficiency: Static and Dynamic 117
5.3.3 Further Decompositions 119
5.3.4 Distressed Firms 119
5.3.5 How to Compare Productivity across Industries, Sectors, Regions, and Countries 120
5.3.6 Indicators of Credit Constraint 121
5.3.7 Indicators of Market Imperfection 124
5.3.8 Job Creation Rates (JCR) and Job Destruction Rates (JDR) 125
5.3.9 Revenue Dynamism 126
5.4 Data Collection and Harmonization 126
5.4.1 Confidentiality 126
5.4.2 Outlier Treatment 128
5.4.3 Weighting Procedure 129
5.4.4 Data Sources 131
5.4.5 Harmonization of Input Data 141
5.4.6 List of Macro Sectors and Industries 155
6. References 162

List of Tables

TABLE 1: COUNTRIES, SAMPLES AND TIME SPAN 9
TABLE 2: SIZE-CLASS DEFINITIONS 11
TABLE 3: FIRMS' AGE DIMENSION 11
TABLE 4: VARIABLES IN THE JOINT DISTRIBUTIONS, BY GROUP 16
TABLE 5: OVERVIEW OF THE JOINT DISTRIBUTION GROUP COMBINATIONS, USING THE SHORT NAMES (TABLE 4) 20
TABLE 6: COMPNET IDENTIFICATION CODE 24
TABLE 7: WEIGHTING METHODS USED FOR THE DECOMPOSITIONS 26
TABLE 8: SUFFIXES 26
TABLE 9: MACRO SECTORS COVERED BY COUNTRY FOR THE HEADCOUNT VARIABLE 33
TABLE 10: DATA FILES OVERVIEW 39
TABLE 11: COMPETITION VARIABLES 40
TABLE 12: FINANCE VARIABLES 65
TABLE 13: LABOR VARIABLES 72
TABLE 14: PRODUCTIVITY VARIABLES 76
TABLE 15: TRADE VARIABLES 84
TABLE 16: OTHER VARIABLES 91
TABLE 17: LIST OF VARIABLES IN THE UNCONDITIONAL DESCRIPTIVE FILES 92
TABLE 18: LIST OF DECOMPOSITION VARIABLES – FOSTER 105
TABLE 19: LIST OF DECOMPOSITION VARIABLES – OP 107
TABLE 20: OVERVIEW OF DECOMPOSITIONS 117
TABLE 21 OLLEY-PAKES INPUT WEIGHT TYPES 118
TABLE 22: COUNTRY-SPECIFIC DATA SOURCES 131
TABLE 23: RAW INPUT VARIABLES - DEFINITIONS 141
TABLE 23A: LIST OF DEFLATED VARIABLES 149
TABLE 24: COUNTRY-SPECIFIC DEFINITIONS OF INPUT VARIABLES 151
TABLE 25: LIST OF MACRO-SECTORS AND INDUSTRIES INCLUDED IN THE 9TH VINTAGE 155
TABLE 26: CATEGORIZATION OF INDUSTRIES BY TECHNOLOGICAL INTENSITY 159

List of Figures

FIGURE 1: STRUCTURE OF THE DATASET 12
FIGURE 2: PERCENTILES OF LOG LABOR PRODUCTIVITY DISTRIBUTION BY FIRM SIZE CENTILES IN DENMARK, 2020. 22

How to Become a User

To receive access, it is necessary to fill out 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 and access the 9th (sample period up to 2021, but most countries up to 2020) 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), 7th (up to 2017), and 8th (up to 2019) 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 9th Vintage of the CompNet dataset with all the 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, target population, and the naming convention of the data files and variables. Chapter 3 provides an overview of the caveats and possible limitations of this version of the CompNet dataset. Chapter 4 illustrates key differences between the 8th and the 9th Vintage. The above chapters are augmented by an extensive appendix that provides detailed information, lists, and tables on:

- List of data folders (5.1),
- Detailed variable overview (5.2),
- Derivation of parametric indicators (5.3),[2]
- CompNet Data Collection and Harmonization (5.4):
  - Confidentiality,
  - Outlier routine,
  - Weighting procedure,
  - Data sources,
  - Harmonization of input data/data preparation,
  - List of macro sectors and industries (NACE Rev.2).

[2] The expressions “indicators” and “variables” are in many occasions used interchangeably. Especially with respect to naming conventions 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 9th 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 9th Vintage of CompNet dataset is an unbalanced panel dataset covering 22 European countries. The dataset includes a rich set of indicators from six roughly defined categories: Competition, Productivity, Labor, Trade, Finance, 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.[3] CompNet reports "20e" to accommodate these differences and improve comparability between countries. Table 1 shows the samples and time spans available in the 9th Vintage of the CompNet dataset.

[3] 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.

Table 1: Countries, Samples and Time Span

Country	All firms	20e	Time Span
Belgium	**X**	**X**	2000–2020
Croatia	**X**	**X**	2002–2021
Czech Republic	**X**	**X**	2005–2020
Denmark	**X**	**X**	2001–2020
Finland	**X**	**X**	1999–2020
France	**X**[1]	**X**	2003–2020
Germany		**X**[2]	2001–2018[3]
Hungary	**X**	**X**	2003–2020
Italy	**X**	**X**	2006–2020
Latvia	**X**	**X**	2007-2019
Lithuania	**X**	**X**	2000–2020
Malta	**X**	**X**	2010-2020[4]
Netherlands	**X**	**X**	2007–2019
Poland		**X**	2002–2020
Portugal	**X**	**X**	2004–2020[5]
Romania		**X**	2005–2020
Slovakia		**X**	2000–2020
Slovenia	**X**	**X**	2002–2021
Spain	**X**	**X**	2008–2020
Sweden	**X**	**X**	2003–2020
Switzerland	**X**	**X**	2009–2020
United Kingdom		**X**[6]	1997-2019

[1] **France:** all firms sample covers the period 2008-2020.
[2] **Germany:** Only weighted version is available.
[3] **Germany:** Macro-sector coverage: Manufacturing (2001-2018), Wholesale and Retail Trade and Accommodation and Food Service Activities (2005-2018), other macro-sectors (2003-2018).
[4] **Malta:** The macro-sector: Real Estate Activities in the 20e sample covers the period 2017-2020.
[5] **Portugal:** A significant number of indicators could not be calculated for the period 2004-2009.
[6] **United Kingdom:** Only weighted version is available.

Target Population:

The CompNet dataset covers non-financial corporations with at least one employee covering the macroeconomic sectors as in table 25. This definition is consistent with category S.11 in

the European System of Accounts, except for sector 19, which, due to its small number of firms, is not covered by the CompNet dataset.[4] It consists of institutional units which are independent legal entities[5] and market producers, whose main activity is the production of goods and non-financial services (excluding sole proprietors). We refer to these independent legal entities as *firms* henceforth, and we elaborate further about the units of observation in section 3.3.1. 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.

Most countries in CompNet provide a sample from the underlying target population. To iron out sampling differences within and across countries, CompNet applies a granular reweighting procedure to generate its micro-data-based aggregate statistics for the target population. More details on the weighting procedure are given in Section 5.4.3. Generally, population-weighted datasets reflect the underlying population to the best extent possible, whereas unweighted statistics show characteristics of the underlying sample directly.

Levels of Aggregation (or Dimensions):

Indicators available in the CompNet dataset are aggregated to different levels, e.g. according to different sector definitions or firm sizes. The available levels of aggregation are Country, Macro-Sector[6], Macro-Sector-Size-Class, Industry 2-digits[7], NUTS 2[8], firm age[9], and technology and knowledge intensity.

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

[4] 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.
[5] The Netherlands is an exception because financial data could be provided only at the firm group level instead of the unconsolidated firm-level.
[6] Corresponding to NACE Rev. 2 sections.
[7] Corresponding to 2-digit NACE Rev. 2 sectors.
[8] Corresponding to basic regions for the application of regional policies based on the Nomenclature of Territorial Units for Statistics (NUTS).
[9] This dimension indicates how old the firm is.

Most countries consistently used the NUTS 2 2016 classification for all periods. However, in the case of the Netherlands, no NUTS-classification could be provided[10]. In some countries, it is possible that the regional classification is missing for a subgroup of firms. CompNet tables typically report statistics on this "missing" category, too, which allows users to assess the regional representativeness of the underlying firm sample. It should also be noted that, diverging from the official Eurostat regional statistics, the NUTS 2 classification for multi-plant enterprises is not assigned according to the location of the local production units, but rather according to the location of the firm`s headquarter.

Please note that all total factor productivity variables and (non-monetary) marginal productivity variables are only available at the 2-digit industry aggregation level.[11]

The size-class definitions, shown in Table 2, follow the Eurostat classification system, except that the Eurostat SBS size class 1 also contains firms with 0 employees:

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.

The age of firms dimension classifies firms into two categories, as shown in Table 3:

Table 3: Firms' Age Dimension

Young firms = 1	Old firms = 2
5 years or younger.	Older than 5 years.

The technology and knowledge intensity dimension classifies firms based on the level of technological intensity within the manufacturing sector and the level of knowledge intensity within the service sectors, following Eurostat[12] in its aggregation of the manufacturing sector and services, according to the technological intensity and knowledge intensity, respectively, at the 2-digit industry level. Table 26 illustrates the classification of 2-digit industries by

[10] In addition, Belgium, Italy, and Spain use the NUTS 3 classification instead. In the case of Latvia and Malta, the NUTS 1 (the whole country) and the NUTS 2 classification are identical for all years. Furthermore, for Lithuania, the NUTS 2 classification has been identical with the national level until the fourth revision (NUTS 2016) and therefore a specific NUTS 2 classification is available only for the years 2017-2020.

[11] This concerns the TFP variables with codes PEb0 – PEb2, PEb4 – PEb5, and PEj0 – PEj3 (cf. Table 14).

[12] https://ec.europa.eu/eurostat/cache/metadata/EN/htec_esms.htm

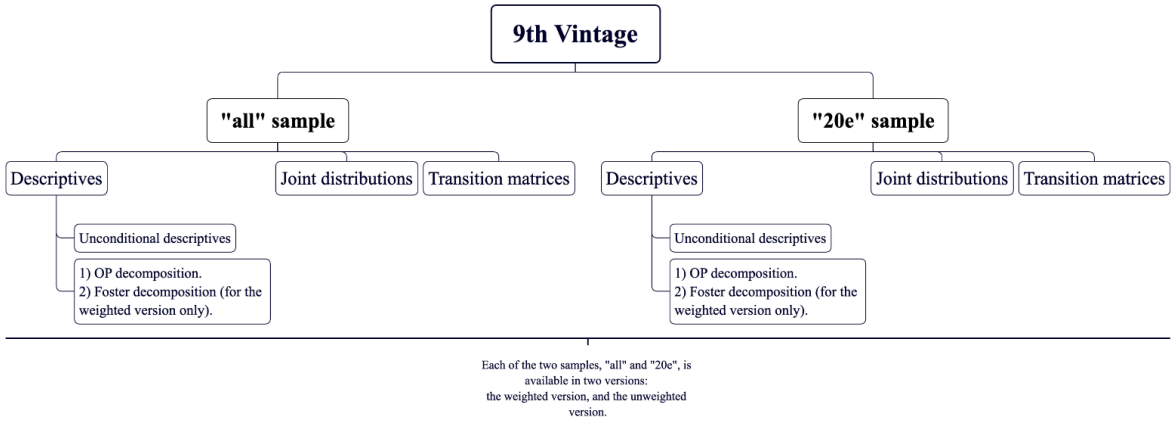
technological intensity aggregated within the manufacturing sector and within the service sectors.

2.2 Structure of the 9th Vintage of the CompNet Dataset

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

The files in the dataset are provided as Stata (.dta) files.[13] 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.

Figure 1: Structure of the Dataset



<u>The Naming Convention of the Data Files:</u>

In general, file names follow the pattern *content_dimension_sample_weighting.dta* to specify

1. ***Content*** The dataset's theme, and, if applicable, the main variable of interest

   (Examples: *unconditional* in the Descriptives folder, *jd_[group]*, for a specific file in JointDistributions; or *tm* for a Transition Matrix).

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

[13]The .dta files are compatible with Stata version 13 or higher. In addition, many statistical software packages like RStudio are capable of importing Stata files to other statistical software.

(i.e. country, macro-sector, macro-sector-size class, NUTS2, 2-digit industry, firm age, and technological intensity).

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 -unweighted sample version, respectively. Section 3.2.1 and Section 5.4.3 provide details about the population and the sample, and the weighting procedure.

Sample and weighting information are consistent with the overall folders of the complete dataset.

Within each file, the observations are identified by *country – dimension – year,* where *dimension* is indicating the level of aggregation, which is only featured on levels below the country level itself. Depending on the respective file types, additional identifiers are featured which are described in the following, alongside more details and explanations about the specific contents of the three different thematic folders.

2.2.1 Descriptives

The descriptive folder 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 for all indicators available in CompNet. 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, skewness, and kurtosis), and the number of firms in the respective panel subset, as well as the

sum of weights, i.e. the underlying population of firms per cell in the population-weighted versions.

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, 2-digit NACE Rev. 2 industry, NUTS2, firms' age, and technological intensity dimensions, each for both the "all" sample and the "20e" sample and both as weighted and unweighted versions and identified by the according suffix as follows: "*unconditional_dimension_sample_[un]weighted.*" Table 17, in the Appendix, provides a list of all variables available in the unconditional descriptive files.

Decompositions

The decompositions are datasets that 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 content-prefix *"fhk_decomp" and "op_decomp".* These stand for decompositions in line with the methodology in Foster et al. (2006) and Olley and Pakes (1996), respectively. Both decompositions are available for seven different levels of aggregation: country, macro-sector, macro-sector-size-class, industry 2-digits, NUTS2, firms' age, and technological intensity. The Foster decompositions are available only as population-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 5.3.2 of the appendix and the original articles. Table 18 and Table 19, in the Appendix, provide the lists of variables in the decomposition dataset. Additionally to *country-[dimension]-year* these files feature the *weight_type* as an additional identifier reflecting novel alternative input-weighting approaches of the 9th vintage.

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.[14] 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. Hence, in the Joint Distributions observations are additionally identified by the by_var, i.e. the conditioning variable, and the by_var_value, i.e. its characteristic value.

A joint distribution requires that each analyzed firm reports all variables within a joint-distribution file in a given year. In order to analyze the largest possible number of firms, the 9th Vintage provides joint distributions for variables from ten separate thematic groups, shown in Table 4. Each group consists of conditional variables and summarized variables. The conditional distributions are provided for both, but the latter are not used as conditioning variables. Please note that the files are identified by the abbreviated short group names, given in column two of Table 4, such that the content-prefix of the file name is *jd_[group_name].*

The unweighted conditional distributions describe the firm sample. The weighted versions use inverse probability weighting within the strata defined in the calculation, i.e., a country-year-2-digit-industry-size-class bin (this is not necessarily the strata of the survey data collection) to describe the population of firms.[15] 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).[16] 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* necessarily comparable across group files, because the samples and weights (if applicable) differ. So, this caveat applies to both the unweighted version and the weighted one.[17]

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

[15] 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). This is the case if the weighting strata and collection scheme of the underlying firm surveys are in line.

[16] This ensures within-group consistency while maximizing the number of available observations in each group.

[17] 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.

Table 4: Variables in the Joint Distributions, by Group

Group Topic	Short name	Summary	Variables	Description
Input	inp	Labor Labor cost Wages Capital Top10-dummies	LV21_l	Labor: number of employees in headcounts
			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 deflated nom. value added; nom. va = nom. revenue – nom. intermediate inputs
			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)
			FV14_rk	Real capital
			PV05_lnsr_cs	Log. Solow residual, weights in CD from cost shares
			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	D = 1, if Top10 firm by employee-number, mac-sector level.

Group Topic	Short name	Summary	Variables	Description
			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, mac-sector level
Productivity	prod	TFP Markup & markdown	PEi9_ln_tfp_0	Log TFP - Specification 0 (CD, cost shares, quant.)
			PEj0_ ln_tfp_1	Log TFP - Specification 1 (CD, OLS, quant.)
			PEj1_ln_tfp_2	Log TFP - Specification 2 (TL, OLS, quant.)
			CE32_markdown_l_0	Labor markdown – Specification 0 (CD, cost shares, quant.)
			CE33_markdown_l_1	Labor markdown – Specification 1 (CD, OLS, quant.)
			CE34_markdown_l_2	Labor markdown – Specification 2 (TL, OLS, quant.)
			CE44_markup_0	Markup – Specification 0 (CD, cost shares, quant.)
			CE45_markup_1	Markup - Specification 1 (CD, OLS, quant.)
			CE46_markup_2	Markup – Specification 2 (TL, OLS, quant.)
			CE58_markdown_l_5	Labor markdown – Specification 5 (TL, OLS, rev.)
			CE62_markdown_m_5	Intermediate input markdown – Specification 5 (TL, OLS, rev.)
			CE56_markdown_k_5	Capital markdown – Specification 5 (TL, OLS, rev.)
Finance	fin	Access to credit	FD01_safe	D = 1, if firm is financially constrained
Trade	trad	Import/Export ratios	**Conditioning variables**	
			TC03_exp_top10	Categorical variable indicating large exporters (see 5.2.5)

Group Topic	Short name	Summary	Variables	Description
		Categorical variables for large importers/exporters Imports/Exports relative to revenue	TC06_imp_top10	Categorical variable indicating large importers (see 5.2.5)
			TC01_exp_imp_rel	Categorical variable indicating the direction of trade (see 5.2.5)
			TC00_exp_dest	Categorical variable indicating export destinations (see 5.2.5)
			TC04_imp_dest	Categorical variable indicating the origins of imports (see 5.2.5)
			TD15_exp_adj	D = 1, if exporting, adj.
			TD89_imp_adj	D = 1, if importing, adj.
			Additional summarized variables	
			TV03_exp_adj	Exports, adj.
			TV09_imp_adj	Imports, adj.
			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	
			TV03_exp_adj	Exports, adj.
			TV09_imp_adj	Imports, adj.
			TR02_exp_adj_rev	Ratio: exports adj. / nom. revenue
			TR38_imp_adj_rev	Ratio: imports adj. / nom. revenue
Growth Rates	grow	Indicators for growth of productivity, firm size, and exports	PV03G1_lprod_va	Growth rate (from t-1): labor prod., real value-added based
			TV02G1_exp_val_adj	Growth rate (from t-1): adjusted exports
			PEb0G1_tfp_0	Growth rate (from t-1): TFP – Specification 0 (CD, cost shares, quant.)
			FV08GH_dhs_rev_growth	Davis-Haltiwanger-Schuh growth rate (from t-1): revenue

Group Topic	Short name	Summary	Variables	Description
			LV21GH_dhs_labor_growth	Davis-Haltiwanger-Schuh growth rate (from t-1): labor
Firm Demography	demo	Firm age Growth rate of young and old firms	OC00_firm_age	1 “0-2 years", 2 "3-5 years", 3 "6-25 years", 4 "more than 25 years"
			CC01_young_high	1: high-growth young firms, 2: low-growth young firms, 3: high-growth old firms, 4: low-growth old firms.
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	Ratio: intangible capital to nominal revenue
Energy	ene	Energy revenue and costs	FR40_ener_costs	Ratio: energy costs / labor and intermediate costs
			FR41_ener_rev	Ratio: energy costs / nom. revenue

To enable cross-group comparisons based on consistent samples without compromising the sample sizes, additional conditional distributions are available for each unique pairwise combination of groups (e.g. Finance-Trade [*fin_trad*]). These pairs, shown in Table 5, 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 and Finance-Trade) is possible for the weighted versions, under the assumption that the population weighting eliminates sample selection issues.

Therefore, these groups including all their pairwise combinations are featured in the JD files for all aggregation levels, as well as both samples and population-weighting versions. The following examples illustrate the logic of the naming convention for them. The joint distribution files are available at the country, macro-sector, macro-sector-size-class, industry

2-digits, NUTS2, firms' age, and technological intensity dimensions for both the "all" sample 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 5 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 (input group: inp) conditional on firm age (demography group: demo) 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 and their combinations result in a total number of 45 *group*s, listed in Table 5. 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.[18] This naming scheme and the use of a continuous joint distribution are illustrated in example 4 below.

Table 5: Overview of the Joint Distribution Group Combinations, Using the Short Names (Table 4)

Group	inp	prod	fin	trad	trca	grow	demo	zomb	ifa	ene
inp	inp	inp_prod	inp_fin	inp_trad	inp_trca	inp_grow	inp_demo	inp_zomb	inp_ifa	inp_ene
prod		prod	prod_fin	prod_trad	prod_trca	prod_grow	prod_demo	prod_zomb	prod_ifa	prod_ene
fin			fin	fin_trad	fin_trca	fin_grow	fin_demo	fin_zomb	fin_ifa	fin_ene
trad				trad	trad_trca	trad_grow	trad_demo	trad_zomb	trad_ifa	trad_ene
trca					trca	trca_grow	trca_demo	trca_zomb	trca_ifa	trca_ene
grow						grow	grow_demo	grow_zomb	grow_ifa	grow_ene
demo							demo	demo_zomb	demo_ifa	demo_ene
zomb								zomb	zomb_ifa	zomb_ene
ifa									ifa	ifa_ene
ene										ene

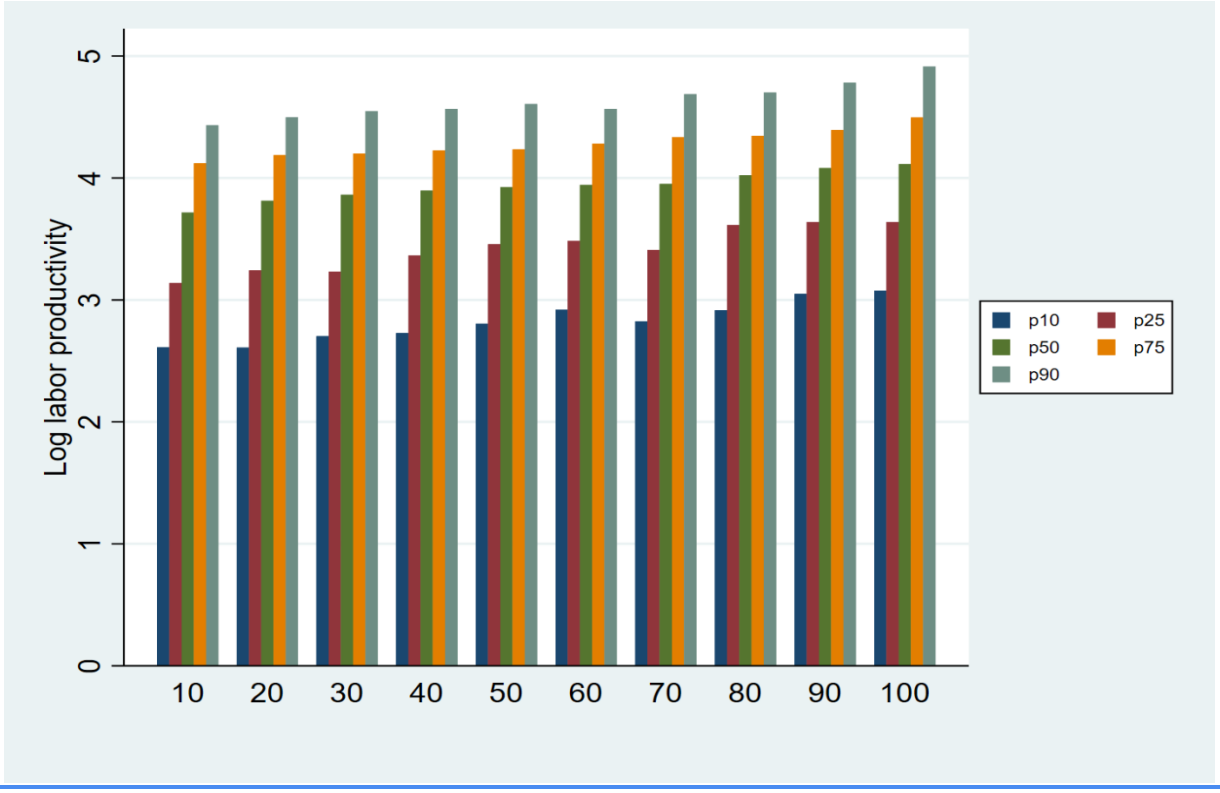
[18] Group "pairs" like input-input would be similarly redundant and thus do not exist.

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 country-dimension-year strata conditional upon the firm being a zombie or not a zombie. To condition on a continuous variable, for example, labor 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' labor productivity within the respective country-dimension-year.

Example 4: **Working with Joint Distributions**

Figure 2 shows an example of a (continuous) joint distribution.[19] In particular, it uses the variable *PV03_lnlprod_va* to show the labor productivity distribution of firms for different deciles of the firm size (*LV21_l*; taken from the conditional variable called *by_var*) in Denmark in 2020. The conditional distribution shows that, at each percentile, log labor productivity seems to be slightly increasing, on average, when the size decile increases. However, the difference between the 10% least productive firms and the 90% most productive firms does not appear to be changing when the size decile increases.

Figure 2: Percentiles of Log Labor Productivity Distribution by Firm Size Centiles in Denmark, 2020.



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 (and others, see below) 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 based on the firm number of employees for year *t* and *t-3*, respectively, both within a given

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

country-dimension strata. 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 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"* and are available at the macro-sector, and newly added in this vintage for technological intensity, and firms' age dimensions, all in weighted and unweighted versions. The size quintile in t-3 is indicated by *qt_l3* and the size quintile in *t* is indicated by *qt, which is thus both additional identifiers in these datasets*. 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, for comparison, the files with the prefix "tm_unc" provide unconditional summary statistics for the specific firm sample that was used to compute the transition matrices, but without conditioning on the employment quintile transitions.

Novel additions in this vintage besides the new dimensions are also different *tm_var* versions. Previously the quintiles were only provided for the size distribution, but now they are also offered based on the distributions for value-added-based labor productivity or real revenue, within the respective dimensions.

2.3 Naming Convention of Variables

The naming convention for variables in the 9th Vintage of the CompNet dataset continues the scheme of the 8th Vintage. The code system from the last vintage is maintained and uniquely and consistently identifies variables by a combination of 4 characters, also across both vintages. Discontinued variables' codes are not featured any longer and new variable additions from the 9th vintage have received a new code. This "identification code" can always be found at the beginning of each variable.[20]

9th Vintage Variable Naming Scheme:

Every variable included in the 9th Vintage of the CompNet dataset follows the pattern:

IdentificationCode_IndividualName_[Weightedby]_Suffix

[20] See Section 5.2 Detailed Variable Overview, where the names of the new variables (column "Variable Name") which were assigned new identification codes are in **bold**.

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

CategoryTypeNumericalCode

The first letter of the code represents the category; the second letter indicates the variable type and the last two digits are arbitrary numbers (or lower-case letters) for unique identification.

The **category** corresponds to the topics covered by the CompNet dataset. Different **types** of variables contain information about the variable format and measurement. It is possible to directly understand from the **identification code** whether a variable is e.g. based on an estimate or represents a ratio. The **numerical code system** is used to distinguish different indicators within each category and variable group. For example, LC00_lc_rev, one approximation of the wage share in the CompNet dataset, is the ratio of labor costs over real revenue. It belongs to the "labor" category denote by the first letter "L" and represents a ratio, which is denoted by second letter "R" for ratio. 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 8th and 9th vintages.[21] In most cases, 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 8th 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 – Labor	E – Estimates[7]	02	…	a2	a8	b4	c0	G3
P – Productivity	R – Ratios	03	09	a3	a9	b5	c1	
T – Trade	V – Values[8]	04	10	a4	b0	b6	c2	
O – Other		05	… 99	a5	b1	b7	…	

[21] They are not necessarily unique compared to vintages older than the 8th vintage.

[7]Estimates are defined as any variable which is based on a production function
[8]Defined as a continuous 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 "ener_rev" in the variable name "FR41_ener_rev_*" stands for the ratio of energy costs to sales revenue. 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 rates of variables (calculated for selected variables) are reported as any other summarized variable in the dataset. To facilitate comparisons with the underlying variable, they do not receive their own code. Instead, 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 FV14**G1**_rk and FV14**G3**_rk, respectively, where 1 and 3 denote the time dimension. Growth rates calculated according to Davis, Haltiwanger and Schuh (1996) are denoted by the third alternative suffix **GH.** These are one-year growth rates bounded between -2 and +2.

Weightedby is used only for the aggregate variable decompositions within the decomposition files[22] and indicates the applied input-weight. 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 "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

[22] The decomposition data files are part of the "Descriptives" data files and denoted by "fhk/op_decomp_ *dimension_sample*_[un]weighted.dta." See section 2.2 for details about the structure of the CompNet dataset and the naming convention of the data files.

reports the different input-weighting methods used in the 9th Vintage of the CompNet dataset for the decomposition variables[23] including the applied abbreviation.

Table 7: Weighting Methods Used for the Decompositions

Abbreviation	Meaning
rrv	Real revenue
ntc	Nominal total costs
nlc	Nominal labor cost
nrv	Nominal revenue
l	Labor
rva	Real value-added
nm	Nominal intermediate inputs
nvi	Nominal variable inputs
nen	Nominal energy inputs
nva	Nominal value-added
fte	Full time equivalents of firm employment
rk	Real capital
kc	Capital cost
nk	Nominal capital

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

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

cov	The covariance term in the OP decomposition
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
flag	Flag indicating whether moments are missing due to confidentiality reasons.
iwmn	Indicates the mean value of the input weight; only used in the decompositions.

3. Important Notes on Using the Dataset

This chapter highlights a few important features of the 9th 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 9th 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 9th Vintage of the CompNet dataset addresses a multitude of caveats existing in the 8th 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 older CompNet data vintages can be found in the CompNet cross-country comparability report (2018). Despite being based on older vintages, this report provides a comprehensive general discussion on data comparability.

Furthermore, CompNet provides an autonomous e-learning training course to give an overview of some of the most useful research tools offered by the CompNet dataset, available at three levels: beginners, intermediate, and advanced. See the website of CompNet for more details about the training. The users can access the training course when applying to CompNet data.

3.1 CompNet Dataset and 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 that also cover firms outside of the CompNet dataset's target population.[24] Hence, the two datasets might show similar patterns but are vastly different because they measure different slices of economic activity.

[24] For more details see Section 2.1

3.2 General Notes

3.2.1 Sample and Population Figures

The data user must be aware that the applied weighting procedure gears the descriptive statistics of the CompNet indicators toward describing the total population, not the underlying sample.[25] Consequently, one frequent question is how to compute aggregate statistics with the variables provided in the weighted datasets in the 9th Vintage of the CompNet dataset.

To compute aggregate statistics for the underlying population, one first needs to differentiate between variables reported in levels and variables reported in ratios. Ratios cannot be easily aggregated and must either be aggregated by separately calculating totals of the numerator and denominator or by using appropriate weights (see our training courses which users can access when applying for the data). For variables in levels, users need to make use of variables ending with "_sw", standing for "summed_weights". To calculate, e.g., 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 number of firms in the cell that was available to construct a given variable in the underlying raw firm data.

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 a 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.

Notably, as the raw data underlying the CompNet dataset differs in macro-sector coverage for some countries, country-level results will not always be perfectly comparable across countries and against other sources. One can therefore apply the above-mentioned approach to calculate totals from the data, to construct a harmonized set of country-level results across countries. For a further discussion on this, we refer the interested reader to Bighelli et al.

[25] 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.

(2022). Furthermore, also our training courses provide further details on working with the data.

3.2.2 Dummy Variables

The CompNet dataset contains many dummy variables. 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 9th 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 type ***C*** (for example O**C**00_*firm_age*). The conditional distributions contain moments for each level of the categorical variables.

3.3 Comparability

We explain the data collection and harmonization in more detail in Section 5.4, in the appendix.

By using existing national data sources without accessing the underlying firm data ourselves, we ensure full confidentiality of the firm data and provide a unique database that combines representative data across multiple European countries. Yet, the data collection process also has a 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 decide on their relative importance, as well as suggesting strategies to mitigate the potential biases of their estimations based on CompNet data. For that reason, CompNet has invested to produce detailed meta-data and to analyze 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. See Section 5.4.5 to see the difference in raw input variables' definitions adopted by each country, and 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 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 significant breaks or changes are 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) 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 or statistical units. The data providing countries for Eurostat's Structural Business Statistics (SBS) database used the smallest legal unit as a statistical unit for firms for the years before 2016 or 2017. With the reference year 2017 or 2018 – depending on the country – the national statistical offices changed to the enterprise level as the new statistical unit in accordance with the statistical unit Regulation N°

696/93.[26] This change of units causes breaks in the Eurostat SBS time series after 2016. The enterprise level is used by a selected number of data providers of CompNet as well, but the majority use the legal unit.[27] 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.[28] In addition, the CompNet team is preparing quality reports, by country, to discuss issues related to the data, which we will publish in the near future.

Differences in sectoral coverage across countries

Not all countries provide information on all macro-sectors for all years. This affects comparability at the country level and can vary, depending on the variable of interest. Table 9 provides an overview of the macro-sectors covered by country for the headcount variable[29]. We recommend data users to check the sectoral coverage for the variables they study. This can be easily done by studying the unconditional data file for the macro-sector level.

[26] "The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit." 'Council Regulation (EEC) 696/93' (1993).

[27] See appendix Section 5.4.4

[28] For an assessment, see the CompNet cross-country comparability report (2018).

[29] We use the weighted 20-e sample. The file used is *unconditional_mac_sector_20e_weighted.dta*.

Table 9: Macro Sectors Covered by Country for the Headcount Variable

Country	Macro Sectors Covered in CompNet	Years
Belgium	All	2000-2020
Croatia	All	2002-2021
Czech Republic	All	2005-2020
Denmark	All	2001-2020
Finland	All	1999-2020
France	All	2003-2020
Germany	All, **Except**:	2003-2018
	1-Manufacturing	2001-2018
	3-Wholesale and retail trade, and 5-Accommodation and food service activities	2006-2018
Hungary	All	2003-2020
Italy	All	2006-2020
Latvia	All	2007-2017
Lithuania	All	2000-2020
Malta	All, **Except**:	2010-2020
	7-Real estate activities	2017-2020
Netherlands	All	2007-2019
Poland	All	2002-2020
Portugal	All	2010-2020
Romania	All 7-Real estate activities is not covered.	2005-2020
Slovakia	All	2000-2020

Slovenia	All[30]	2002-2021
Spain	All	2008-2020
Sweden	All	2003-2020
Switzerland	All	2009-2020
United Kingdom	All	1997-2019

For the macro-sectors included in CompNet dataset, see section 5.4.6.

3.3.2 Some Country-specific Aspects

We understand the differences in the methodology and the pre-existing infrastructure between countries. Nevertheless, we work on identifying the issues and anomalies present in the data from each data provider as a part of our recent effort to assess and evaluate the quality of the data from each country, and we **constantly update** the issues when we encounter them. Please keep in mind that the following list of countries and their specific issues is not exhaustive.[31]

Belgium

In the 9th Vintage, some **raw** input variables[32], or one of their components, were excluded due to either being fully missing or having high shares of missing values, namely intangible fixed assets; (current) other financial assets (including financial assets held for trading and derivatives); variation in stocks of finished goods and work in progress (includes the change in inventories of production recognized in the income statement); provisions (of non-current liabilities); (non-current) bonds and similar obligations (bonds and similar securities not due to be settled within 12 months after the reporting period); and (non-current) other financial creditors (funding from other financial creditors not due to be settled within 12 months after the reporting period).

Slovakia

[30] The macro-sector 7-Real estate activities is covered. However, the observations are missing for the years 2016-2021.

[31] For further country-specific information please see https://www.comp-net.org/eu-technical-support-instrument-tsi/visualization/

[32] See Table 23.

The variables in 2020 are originally missing, and the values were imputed using growth rates of the missing variables calculated based on the Bisnode database[33]. The variables are total fixed assets, capital, intangible fixed assets, and depreciation. The Bisnode data covers missing variables, including the year 2020, for almost the entire population of firms. However, it is available for only a few recent years, so it is not sufficient for the "all firms" sample.

In addition, the Bisnode data differ in identifying some variables in the preceding years due to slightly different balance sheet items' definitions.

Slovenia

Some **raw** variables are available only from 2006 as a point in time, namely Other fixed assets; Current assets; Current liabilities; Other current liabilities; Non-current liabilities; Other non-current liabilities; Total debt; Long-term debt; and Accounts payable.

In calculating gross output, subsidies on products and production were included in the calculation of gross output. However, including subsidies on production in output is a discrepancy with national accounts statistics, which include only subsidies on products. Normally, this might not be a problem, as the amount of subsidies is not significant. However, in 2020, the level of subsidies on production increased significantly. The output of sectors, which were important beneficiaries of subsidies on production will probably differ substantially from national accounts statistics. Another issue might arise with the use of value-added deflators for transforming nominal into real variables: value-added deflators are compiled within national accounts statistics and do not include subsidies on production, hence higher subsidies on production in 2020 will be translated into real output.

Sweden

Some observations exist where the first percentile of full-time equivalents of firm employment (LV25_fte_p1) is zero, despite the exclusion of all firms that have less than one employee. This is because most of the active enterprises within the Structured Business Statistics (SBS) survey[34] have zero full-time equivalent even if some of these firms have one employee. The reason is the difference in the definition of employment variable in Sweden; the employee

[33] See Table 22 for country-specific data sources.
[34] SBS survey is one of the data sources in Sweden. See Table 22.

variable, which is not part of the SBS survey, captures all enterprises with at least one employee during November month.

Few other variables show a similar pattern (e.g., values of zeros) which seem counterintuitive at the first sight. For example, the median of the share of employees with tertiary education (LR02_tertshare_p50) is zero in some years. One explanation is that the share of employees with tertiary education is strictly zero at the aggregation level at which the median is zero.

The Swedish data provider confirmed that the missing values had not been replaced by zeros.

3.3.3 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 23 includes information on all raw variables and their possible definitions
- Table 24 highlights the used definitions for each country included in the dataset.

4. Differences between 8th Vintage and 9th Vintage of the CompNet Dataset

The 9th 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:

New Countries and Coverage:

- Three new countries: United Kingdom, Latvia, and Malta.
- Coverage up to 2021, but for most countries up to 2020.

New Dimensions:

- Include two new dimensions: firms' age, and technological intensity. Table 3 and Table 26 show the details of firms' age and technological intensity dimensions, respectively.

New Variables:

- Firms' market shares for various variables.
- Revenue dynamism. See section 5.3.9.
- Measure for intermediate input market imperfections.
- Measure for capital market imperfections.
- Energy expenditures over material and labor costs.
- Energy expenditures over sales.
- Energy expenditures over value added.
- FTE-based labor productivity.

See section 5.2 for a complete list of variables, where the new variables' names, which were assigned a new identification code, are in **bold**.

Discontinued Variables:

- Marginal product of capital, marginal product of labor, and marginal product of intermediates.
- Value-added elasticity.
- MRPL-wage gap.

- Dummies for the top 10% of firms in terms of revenue, nominal labor cost, real capital, real value-added, intangible fixed assets, employee-number, being exporter, and being importer.
- Price cost margin without capital costs, assuming fixed capital.
- Nominal revenue over nominal energy inputs, and nominal value-added over nominal energy inputs.
- Solow residual, weights in CD: labor 2/3, real capital 1/3.
- The legal form of the firm (categorical variable).

New Decompositions:

- Harmonic averages for market power parameters.
- Weighted average in OP decomposition with weights fixed in certain years, similar to the within-term from the Foster et al. (2006) decomposition.
- Adjust the decompositions files to include the macro-sector-size-class dimension.

Joint Distributions (JDs):

- Adjust the joint distributions files to include the macro-sector-size-class dimension.
- Replaced the effective tax rate group with the energy group. See Table 5.
- Removal of the following variables: unit labor costs variable from the input group; the variables: ratio of nominal investment to nominal revenue, ratio of operating profits to interest payments, and real intangible investment from the finance group; growth rate of TFP specification 1 from the growth rates group.
- Adding the following variables: TFP specifications 1 and 2 to the productivity group; the growth rate of TFP specification 0 to the growth rates group.

Production Functions:

- Removal of time-varying OLS estimation of CD production function specification.
- Adding two new production function specifications estimating revenue elasticities.

Regression Output:

- Discontinue trade regressions and production function regression output.

5. Appendix

5.1 List of Data Folders

The following overview presents the available data folders and number of files in the 9th Vintage of the CompNet Dataset.

Table 10: Data Files Overview

Sample & weighting	Folder	(Subfolder)	Number of Files
20e_firms_unweighted	**Descriptives**		14
	JointDistributions		385
	Transmatrices		6
20e_firms_weighted	**Descriptives**		21
	JointDistributions		385
	Transmatrices		6
all_firms_unweighted	**Descriptives**		14
	JointDistributions		385
	Transmatrices		6
all_firms_weighted	**Descriptives**		21
	JointDistributions		385
	Transmatrices		6

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 11: Competition Variables

Competition		
Variable Code	**Variable Name**	**Definition**
	Categorical	
CC01	**young_high**	Age (young/old) and High-growth (Y/N)
	Dummies	
CD01	**old_high_0**	D = 1, if old firm (>5) & high growth (>0.25), Def. 0
CD02	**old_high_1_pop_2D**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, industry2d
CD02	**old_high_1_sam_2D**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, industry2d
CD03	**old_high_1_pop_C**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, country
CD03	**old_high_1_sam_C**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, country
CD04	**old_high_1_pop_M**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, macro sector
CD04	**old_high_1_sam_M**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, macro sector

Competition		
Variable Code	**Variable Name**	**Definition**
CD05	**old_high_1_pop_MS**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, macsec_szcl
CD05	**old_high_1_sam_MS**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, macsec_szcl
CD06	**old_high_1_pop_N**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, NUTS2 level
CD06	**old_high_1_sam_N**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, NUTS2 level
CD07	**old_high_1_pop_T**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, techno. knowl.
CD07	**old_high_1_sam_T**	D = 1, if old firm (>5) & high labor (>p95), Def. 1, techno. knowl.
CD08	**old_low_0**	D = 1, if old firm (>5) & low growth (≤0.25), Def. 0
CD09	**old_low_1_pop_2D**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, industry2d
CD09	**old_low_1_sam_2D**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, industry2d
CD10	**old_low_1_pop_C**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, country

Competition		
Variable Code	**Variable Name**	**Definition**
CD10	**old_low_1_sam_C**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, country
CD11	**old_low_1_pop_M**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, macro sector
CD11	**old_low_1_sam_M**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, macro sector
CD12	**old_low_1_pop_MS**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, macsec_szcl
CD12	**old_low_1_sam_MS**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, macsec_szcl
CD13	**old_low_1_pop_N**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, NUTS2 level
CD13	**old_low_1_sam_N**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, NUTS2 level
CD14	**old_low_1_pop_T**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, techno. knowl.
CD14	**old_low_1_sam_T**	D = 1, if old firm (>5) & low labor (<p5), Def. 1, techno. knowl.
CD15	**young_high_0**	D = 1, if young firm (≤5) & high growth (>0.25), Def. 0

Competition		
Variable Code	**Variable Name**	**Definition**
CD16	**young_high_1_pop_2D**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, industry2d
CD16	**young_high_1_sam_2D**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, industry2d
CD17	**young_high_1_pop_C**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, country
CD17	**young_high_1_sam_C**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, country
CD18	**young_high_1_pop_M**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, macro sector
CD18	**young_high_1_sam_M**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, macro sector
CD19	**young_high_1_pop_MS**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, macsec_szcl
CD19	**young_high_1_sam_MS**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, macsec_szcl
CD20	**young_high_1_pop_N**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, NUTS2 level

Competition		
Variable Code	**Variable Name**	**Definition**
CD20	**young_high_1_sam_N**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, NUTS2 level
CD21	**young_high_1_pop_T**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, techno. knowl.
CD21	**young_high_1_sam_T**	D = 1, if young firm (≤5) & high labor (>p95), Def. 1, techno. knowl.
CD22	**young_low_0**	D = 1, if young firm (≤5) & low growth (≤0.25), Def. 0
CD23	**young_low_1_pop_2D**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, industry2d
CD23	**young_low_1_sam_2D**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, industry2d
CD24	**young_low_1_pop_C**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, country
CD24	**young_low_1_sam_C**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, country
CD25	**young_low_1_pop_M**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, macro sector
CD25	**young_low_1_sam_M**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, macro sector

Competition		
Variable Code	**Variable Name**	**Definition**
CD26	**young_low_1_pop_MS**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, macsec_szcl
CD26	**young_low_1_sam_MS**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, macsec_szcl
CD27	**young_low_1_pop_N**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, NUTS2 level
CD27	**young_low_1_sam_N**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, NUTS2 level
CD28	**young_low_1_pop_T**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, techno. knowl.
CD28	**young_low_1_sam_T**	D = 1, if young firm (≤5) & low labor (<p5), Def. 1, techno. knowl.
	Estimates	
CE32	markdown_l_0	Labor markdown - Spec. 0 (CD, cost shares, quant.)
CE33	markdown_l_1	Labor markdown - Spec. 1 (CD, OLS, quant.)
CE34	markdown_l_2	Labor markdown - Spec. 2 (TL, OLS, quant.)
CE36	markdown_l_3	Labor markdown - Spec. 3 (CD, ACF, quant.)
CE37	markdown_l_4	Labor markdown - Spec. 4 (TL, ACF, quant.)

Competition		
Variable Code	**Variable Name**	**Definition**
CE44	markup_0	Markup - Spec. 0 (CD, cost shares, quant.)
CE45	markup_1	Markup - Spec. 1 (CD, OLS, quant.)
CE46	markup_2	Markup - Spec. 2 (TL, OLS, quant.)
CE48	markup_3	Markup - Spec. 3 (CD, ACF, quant.)
CE49	markup_4	Markup - Spec. 4 (TL, ACF, quant.)
CE56	**markdown_k_5**	Capital markdown - Spec. 5 (TL, OLS, rev.)
CE57	**markdown_k_6**	Capital markdown - Spec. 6 (TL, ACF, rev.)
CE58	**markdown_l_5**	Labor markdown - Spec. 5 (TL, OLS, rev.)
CE59	**markdown_l_6**	Labor markdown - Spec. 6 (TL, ACF, rev.)
CE60	**markup_5**	Markup - Spec. 5 (TL, OLS, rev.)
CE61	**markup_6**	Markup - Spec. 6 (TL, ACF, rev.)
CE62	**markdown_m_5**	Intermediate input markdown - Spec. 5 (TL, OLS, rev.)
CE63	**markdown_m_6**	Intermediate input markdown - Spec. 6 (TL, ACF, rev.)
Ratios		

Competition		
Variable Code	**Variable Name**	**Definition**
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
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

Competition		
Variable Code	**Variable Name**	**Definition**
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
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
CR24	**top_ifa_sam_A**	Top10 firms' share in intangible fixed assets, firm age

Competition		
Variable Code	**Variable Name**	**Definition**
CR25	**top_lc_sam_A**	Top10 firms' share in labor costs, firm age
CR26	**top_rev_sam_A**	Top10 firms' share in revenues, firm age
CR27	**top_rk_sam_A**	Top10 firms' share in real capital, firm age
CR28	**top_rva_sam_A**	Top10 firms' share in real value added, firm age
CR29	**top_l_sam_A**	Top10 firms' labor share, firm age
CR30	**top_ifa_sam_T**	Top10 firms' share in intangible fixed assets, techno. knowledge
CR31	**top_lc_sam_T**	Top10 firms' share in labor costs, techno. knowledge
CR32	**top_rev_sam_T**	Top10 firms' share in revenues, techno. knowledge
CR33	**top_rk_sam_T**	Top10 firms' share in real capital, techno. knowledge
CR34	**top_rva_sam_T**	Top10 firms' share in real value added, techno. knowledge
CR35	**top_l_sam_T**	Top10 firms' labor share, techno. knowledge
CR40	**mrktsha_l_pop_A**	Market share, number of employees, pop., firm age

Competition		
Variable Code	**Variable Name**	**Definition**
CR40	**mrktsha_l_sam_A**	Market share, number of employees, sample, firm age
CR41	**mrktsha_l_pop_C**	Market share, number of employees, pop., country
CR41	**mrktsha_l_sam_C**	Market share, number of employees, sample, country
CR42	**mrktsha_l_pop_M**	Market share, number of employees, pop., macro sector
CR42	**mrktsha_l_sam_M**	Market share, number of employees, sample, macro sector
CR43	**mrktsha_l_pop_MS**	Market share, number of employees, pop., macro sector size class
CR43	**mrktsha_l_sam_MS**	Market share, number of employees, sample, macro sector size class
CR44	**mrktsha_l_pop_N**	Market share, number of employees, pop., NUTS2
CR44	**mrktsha_l_sam_N**	Market share, number of employees, sample, NUTS2
CR45	**mrktsha_l_pop_2D**	Market share, number of employees, pop., industry2d
CR45	**mrktsha_l_sam_2D**	Market share, number of employees, sample, industry2d

Competition		
Variable Code	**Variable Name**	**Definition**
CR46	**mrktsha_l_pop_T**	Market share, number of employees, pop., techno. knowledge
CR46	**mrktsha_l_sam_T**	Market share, number of employees, sample, techno. knowledge
CR54	**mrktsha_rev_pop_A**	Market share, nominal revenue, pop., firm age
CR54	**mrktsha_rev_sam_A**	Market share, nominal revenue, sample, firm age
CR55	**mrktsha_rev_pop_C**	Market share, nominal revenue, pop., country
CR55	**mrktsha_rev_sam_C**	Market share, nominal revenue, sample, country
CR56	**mrktsha_rev_pop_M**	Market share, nominal revenue, pop., macro sector
CR56	**mrktsha_rev_sam_M**	Market share, nominal revenue, sample, macro sector
CR57	**mrktsha_rev_pop_MS**	Market share, nominal revenue, pop., macro sector size class
CR57	**mrktsha_rev_sam_MS**	Market share, nominal revenue, sample, macro sector size class
CR58	**mrktsha_rev_pop_N**	Market share, nominal revenue, pop., NUTS2
CR58	**mrktsha_rev_sam_N**	Market share, nominal revenue, sample, NUTS2

Competition		
Variable Code	**Variable Name**	**Definition**
CR59	**mrktsha_rev_pop_2D**	Market share, nominal revenue, pop., industry2d
CR59	**mrktsha_rev_sam_2D**	Market share, nominal revenue, sample, industry2d
CR60	**mrktsha_rev_pop_T**	Market share, nominal revenue, pop., techno. knowledge
CR60	**mrktsha_rev_sam_T**	Market share, nominal revenue, sample, techno. knowledge
CR68	**mrktsha_va_pop_A**	Market share, nominal value added, pop., firm age, only positive
CR68	**mrktsha_va_sam_A**	Market share, nominal value added, sample, firm age, only positive
CR69	**mrktsha_va_pop_C**	Market share, nominal value added, pop., country, only positive
CR69	**mrktsha_va_sam_C**	Market share, nominal value added, sample, country, only positive
CR70	**mrktsha_va_pop_M**	Market share, nominal value added, pop., macro sector, only positive
CR70	**mrktsha_va_sam_M**	Market share, nominal value added, sample, macro sector, only positive

Competition		
Variable Code	**Variable Name**	**Definition**
CR71	**mrktsha_va_pop_MS**	Market share, nom. va, pop., macro sector size class, only positive
CR71	**mrktsha_va_sam_MS**	Market share, nom. va, sample, macro sector size class, only positive
CR72	**mrktsha_va_pop_N**	Market share, nominal value added, pop., NUTS2, only positive
CR72	**mrktsha_va_sam_N**	Market share, nominal value added, sample, NUTS2, only positive
CR73	**mrktsha_va_pop_2D**	Market share, nominal value added, pop., industry2d, only positive
CR73	**mrktsha_va_sam_2D**	Market share, nominal value added, sample, industry2d, only positive
CR74	**mrktsha_va_pop_T**	Market share, nom. va, pop., techno. knowledge, only positive
CR74	**mrktsha_va_sam_T**	Market share, nom. va, sample, techno. knowledge, only positive
Values		
CV02	**hhi_rev_pop_T**	Hirschman-Herfindahl Index, nom. revenue shares, techno. knowl., pop.

Competition		
Variable Code	**Variable Name**	**Definition**
CV02	**hhi_rev_sam_T**	Hirschman-Herfindahl Index, nom. revenue shares, techno. knowl., pop.
CV03	**hhi_rev_pop_A**	Hirschman-Herfindahl Index, nom. revenue shares, firm age, pop.
CV03	**hhi_rev_sam_A**	Hirschman-Herfindahl Index, nom. revenue shares, firm age, sample
CV04	hhi_rev_pop_C	Hirschman-Herfindahl Index, nom. revenue shares, country, pop.
CV04	hhi_rev_sam_C	Hirschman-Herfindahl Index, nom. revenue shares, country, sample
CV05	hhi_rev_pop_M	Hirschman-Herfindahl Index, nom. revenue shares, mac_sector, pop.
CV05	hhi_rev_sam_M	Hirschman-Herfindahl Index, nom. revenue shares, mac_sector, sample
CV06	hhi_rev_pop_N	Hirschman-Herfindahl Index, nom. revenue shares, NUTS2, pop.
CV06	hhi_rev_sam_N	Hirschman-Herfindahl Index, nom. revenue shares, NUTS2, sample

Competition		
Variable Code	**Variable Name**	**Definition**
CV07	hhi_rev_pop_2D	Hirschman-Herfindahl Index, nom. revenue shares, industry, pop.
CV07	hhi_rev_sam_2D	Hirschman-Herfindahl Index, nom. revenue shares, industry, sample
CV10	**hhi_ifa_pop_T**	Hirschman-Herfindahl Index, intangible shares, techno. knowledge, pop.
CV10	**hhi_ifa_sam_T**	Hirschman-Herfindahl Index, intangible shares, techno. knowl., sample
CV11	**hhi_ifa_pop_A**	Hirschman-Herfindahl Index, intangible shares, firm age, pop.
CV11	**hhi_ifa_sam_A**	Hirschman-Herfindahl Index, intangible shares, firm age, sample
CV12	hhi_ifa_pop_C	Hirschman-Herfindahl Index, intangible shares, country, population
CV12	hhi_ifa_sam_C	Hirschman-Herfindahl Index, intangible shares, country, sample
CV13	hhi_ifa_pop_M	Hirschman-Herfindahl Index, int. fix.assets, mac_sector, population

Competition		
Variable Code	**Variable Name**	**Definition**
CV13	hhi_ifa_sam_M	Hirschman-Herfindahl Index, intangible shares, mac_sector, sample
CV14	hhi_ifa_pop_N	Hirschman-Herfindahl Index, intangible shares, NUTS2, population
CV14	hhi_ifa_sam_N	Hirschman-Herfindahl Index, intangible shares, NUTS2, sample
CV15	hhi_ifa_pop_2D	Hirschman-Herfindahl Index, intangible shares, industry, population
CV15	hhi_ifa_sam_2D	Hirschman-Herfindahl Index, intangible shares, industry, sample
CV18	**hhi_l_pop_T**	Hirschman-Herfindahl Index, emp. shares, techno. knowledge, pop.
CV18	**hhi_l_sam_T**	Hirschman-Herfindahl Index, emp. shares, techno. knowledge, sample
CV19	**hhi_l_pop_A**	Hirschman-Herfindahl Index, employment shares, firm age, pop.
CV19	**hhi_l_sam_A**	Hirschman-Herfindahl Index, employment shares, firm age, sample

Competition		
Variable Code	**Variable Name**	**Definition**
CV20	hhi_l_pop_C	Hirschman-Herfindahl Index, employment shares, country, population
CV20	hhi_l_sam_C	Hirschman-Herfindahl Index, employment shares, country, sample
CV21	hhi_l_pop_M	Hirschman-Herfindahl Index, emp. shares, mac_sector, population
CV21	hhi_l_sam_M	Hirschman-Herfindahl Index, employment shares, mac_sector, sample
CV22	hhi_l_pop_N	Hirschman-Herfindahl Index, employment shares, NUTS2, population
CV22	hhi_l_sam_N	Hirschman-Herfindahl Index, employment shares, NUTS2, sample
CV23	hhi_l_pop_2D	Hirschman-Herfindahl Index, employment shares, industry, population
CV23	hhi_l_sam_2D	Hirschman-Herfindahl Index, employment shares, industry, sample
CV26	hhi_lc_sam_T	Hirschman-Herfindahl Index, nlc shares, techno. knowledge, sample

Competition		
Variable Code	**Variable Name**	**Definition**
CV27	hhi_lc_sam_A	Hirschman-Herfindahl Index, nom labor cost shares, firm age, sample
CV28	hhi_lc_pop_C	Hirschman-Herfindahl Index, nom labor cost shares, country, pop.
CV28	hhi_lc_sam_C	Hirschman-Herfindahl Index, nom labor cost shares, country, sample
CV29	hhi_lc_pop_M	Hirschman-Herfindahl Index, nom labor cost shares, mac_sector, pop.
CV29	hhi_lc_sam_M	Hirschman-Herfindahl Index, nlc, mac_sector, sample
CV30	hhi_lc_pop_N	Hirschman-Herfindahl Index, nom labor cost shares, NUTS2, pop.
CV30	hhi_lc_sam_N	Hirschman-Herfindahl Index, nom labor cost shares, NUTS2, sample
CV31	hhi_lc_pop_2D	Hirschman-Herfindahl Index, nom labor cost shares, industry, pop.
CV31	hhi_lc_sam_2D	Hirschman-Herfindahl Index, nom labor cost shares, industry, sample

Competition		
Variable Code	**Variable Name**	**Definition**
CV34	**hhi_rk_pop_T**	Hirschman-Herfindahl Index, real capital shares, techno. knowl., sample
CV34	**hhi_rk_sam_T**	Hirschman-Herfindahl Index, real capital shares, techno. knowl., pop.
CV35	**hhi_rk_pop_A**	Hirschman-Herfindahl Index, real capital shares, firm age, pop.
CV35	**hhi_rk_sam_A**	Hirschman-Herfindahl Index, real capital shares, firm age, sample
CV36	**hhi_rk_pop_C**	Hirschman-Herfindahl Index, real capital shares, country, pop.
CV36	hhi_rk_sam_C	Hirschman-Herfindahl Index, real capital shares, country, sample
CV37	hhi_rk_pop_M	Hirschman-Herfindahl Index, real capital shares, mac_sector, pop.
CV37	hhi_rk_sam_M	Hirschman-Herfindahl Index, real capital shares, mac_sector, sample
CV38	hhi_rk_pop_N	Hirschman-Herfindahl Index, real capital shares, NUTS2, pop.

Competition		
Variable Code	**Variable Name**	**Definition**
CV38	hhi_rk_sam_N	Hirschman-Herfindahl Index, real capital shares, NUTS2, sample
CV39	hhi_rk_pop_2D	Hirschman-Herfindahl Index, real capital shares, industry, pop.
CV39	hhi_rk_sam_2D	Hirschman-Herfindahl Index, real capital shares, industry, sample
CV42	**hhi_rva_pos_pop_T**	Hirschman-Herfindahl Index, real value-added shares, tech-group, pop.
CV42	**hhi_rva_pos_sam_T**	Hirschman-Herfindahl Index, rva shares, techno. knowledge, sample
CV43	**hhi_rva_pos_pop_A**	Hirschman-Herfindahl Index, real value-added shares, firm age, pop.
CV43	**hhi_rva_pos_sam_A**	Hirschman-Herfindahl Index, rva shares, firm age, sample
CV44	hhi_rva_pos_pop_C	Hirschman-Herfindahl Index, real value-added shares, country, pop.
CV44	hhi_rva_pos_sam_C	Hirschman-Herfindahl Index, rva shares, country, sample

Competition		
Variable Code	**Variable Name**	**Definition**
CV45	hhi_rva_pos_pop_M	Hirschman-Herfindahl Index, rva shares, mac_sector, pop.
CV45	hhi_rva_pos_sam_M	Hirschman-Herfindahl Index, rva shares, mac_sector, sample
CV46	hhi_rva_pos_pop_N	Hirschman-Herfindahl Index, real value-added shares, NUTS2, pop.
CV46	hhi_rva_pos_sam_N	Hirschman-Herfindahl Index, real value-added shares, NUTS2, sample
CV47	hhi_rva_pos_pop_2D	Hirschman-Herfindahl Index, real value-added shares, industry, pop.
CV47	hhi_rva_pos_sam_2D	Hirschman-Herfindahl Index, rva shares, industry, sample
CV48	**firmrev_neg_pop_A**	Sum of firm revenue growth, firm age, pop., only negative
CV49	**firmrev_pos_pop_A**	Sum of firm revenue growth, firm age, pop., only positive
CV50	**firmrev_pos_sam_A**	Sum of firm revenue growth, firm age, sample, only positive

Competition		
Variable Code	**Variable Name**	**Definition**
CV51	**firmrev_neg_sam_A**	Sum of firm revenue growth, firm age, sample, only negative
CV52	**firmrev_neg_pop_C**	Sum of firm revenue growth, country level, pop., only negative
CV53	**firmrev_pos_pop_C**	Sum of firm revenue growth, country level, pop., only positive
CV54	**firmrev_neg_pop_M**	Sum of firm revenue growth, macro sector, pop., only negative
CV55	**firmrev_pos_pop_M**	Sum of firm revenue growth, macro sector, pop., only positive
CV56	**firmrev_pos_sam_M**	Sum of firm revenue growth, macro sector, sample, only positive
CV57	**firmrev_neg_sam_M**	Sum of firm revenue growth, macro sector, sample, only negative
CV58	**firmrev_neg_pop_MS**	Sum of firm revenue growth, macsec_szclass, pop., only negative
CV59	**firmrev_pos_pop_MS**	Sum of firm revenue growth, macsec_szclass, pop., only positive

Competition		
Variable Code	**Variable Name**	**Definition**
CV60	**firmrev_pos_pop_N**	Sum of firm revenue growth, NUTS2 level, pop., only positive
CV61	**firmrev_neg_pop_N**	Sum of firm revenue growth, NUTS2 level, pop., only negative
CV62	**firmrev_neg_pop_2D**	Sum of firm revenue growth, industry2d, pop., only negative
CV63	**firmrev_pos_pop_2D**	Sum of firm revenue growth, industry2d, pop., only positive
CV64	**firmrev_pos_sam_2D**	Sum of firm revenue growth, industry2d, sample, only positive
CV65	**firmrev_neg_sam_2D**	Sum of firm revenue growth, industry2d, sample, only negative
CV66	**firmrev_neg_pop_T**	Sum of firm revenue growth, techno. knowledge, pop., only negative
CV67	**firmrev_pos_pop_T**	Sum of firm revenue growth, techno. knowledge, pop., only positive
CV68	**firmrev_pos_sam_T**	Sum of firm revenue growth, techno. knowledge, sample, only positive

Competition		
Variable Code	**Variable Name**	**Definition**
CV69	**firmrev_neg_sam_T**	Sum of firm revenue growth, techno. knowledge, sample, only negative
CV70	**firmrev_pos_sam_C**	Sum of firm revenue growth, country, sample, only positive
CV71	**firmrev_neg_sam_C**	Sum of firm revenue growth, country, sample, only negative
CV72	**firmrev_pos_sam_MS**	Sum of firm revenue growth, macsec_szclass, sample, only positive
CV73	**firmrev_neg_sam_MS**	Sum of firm revenue growth, macsec_szclass, sample, only negative
CV74	**firmrev_pos_sam_N**	Sum of firm revenue growth, NUTS2 level, sample, only positive
CV75	**firmrev_neg_sam_N**	Sum of firm revenue growth, NUTS2 level, sample, only negative

5.2.2 Finance Variables

Table 12: Finance Variables

Finance		
Variable Code	**Variable Name**	**Definition**
	Categorical	
FC07	y_zombie_intcov_pos	Categorical: Duration of current spell as zombie_intcov_pos (i.e. interest payments larger than operating profits for 3 years but positive operating profit and no high labor growth for the last 3 years) in years. 0: Zero years spent as zombie (i.e. firm is not currently a zombie); 1: One year spent as zombie; 2: Two consecutive years spent as zombie; C=3: Three consecutive years; C=4: Four or more consecutive years.
	Dummies	
FD00	absconstr	D = 1, if firm is absolutely credit constrained
FD01	safe	D = 1, if firm is financially constrained
FD05	zombie_intcov_pos	D = 1, if int. payed > op. profit > 0 & no high growth for 3 years

Finance		
Variable Code	**Variable Name**	**Definition**
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)
	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
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

Finance		
Variable Code	**Variable Name**	**Definition**
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
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
FR27	rev_lc	Ratio: nom. revenue / nom. labor cost

Finance		
Variable Code	**Variable Name**	**Definition**
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))
FR32	trade_credit	Ratio: accounts payable / total assets
FR33	trade_debt	Ratio: accounts receivable / total assets
FR35	va_rev	Ratio: nom. value-added / nom. revenue
FR36	ifa_rev	Ratio: intangible capital to nominal revenue
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
FR40	**ener_costs**	Ratio: energy costs / labor and intermediate costs
FR41	**ener_rev**	Ratio: energy costs / sales revenue

Finance		
Variable Code	**Variable Name**	**Definition**
FR42	**ener_va**	Ratio: energy cost / nominal, positive value added
Values		
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
FV03	n_ener	nominal energy inputs (also abbreviated as nei)
FV04	nk	Nominal capital stock
FV05	nlc	Nominal labor costs
FV06	nm	Nominal intermediate inputs
FV07	nrd	nominal Research & Development expenditure
FV08	nrev	Nominal revenue
FV08G1	nrev	Growth rate (from t-1): nom. revenue
FV08GH	**dhs_rev_growth**	Davis-Haltiwanger-Schuh growth rate (from t-1): revenue
FV09GH	**dhs_rev_growth_neg**	Davis-Haltiwanger-Schuh growth rate (from t-1): revenue, only negative
FV10	nva	nominal value-added, computed as nrev - nm

Finance		
Variable Code	**Variable Name**	**Definition**
FV10GH	**dhs_rev_growth_pos**	Davis-Haltiwanger-Schuh growth rate (from t-1): revenue, only positive
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
FV14G1	rk	Growth rate (from t-1): real capital
FV14G3	rk	Growth rate (from t-3): real capital
FV15	rlc	real labor costs
FV16	rm	real intermediate inputs
FV17	rrev	real revenue
FV18	rva	Real value-added, computed as deflated version of nva
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)

Finance		
Variable Code	**Variable Name**	**Definition**
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
FV31	rinvest_intan	real intangible investment

5.2.3 Labor Variables

Table 13: Labor Variables

Labor		
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
	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	
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
LV01	jcr_sam_C	Job creation rates, country level, sample
LV02	jcr_pop_M	Job creation rates, mac-sector level, population
LV02	jcr_sam_M	Job creation rates, mac-sector level, sample

Labor		
Variable Code	**Variable Name**	**Definition**
LV03	jcr_pop_MS	Job creation rates, macsec-szclass level, population
LV03	jcr_sam_MS	Job creation rates, macsec-szclass level, sample
LV04	jcr_pop_N	Job creation rates, NUTS2 level, population
LV04	jcr_sam_N	Job creation rates, NUTS2 level, sample
LV05	jcr_pop_2D	Job creation rates, sector level, population
LV05	jcr_sam_2D	Job creation rates, sector level, sample
LV11	jdr_pop_C	Job destruction rates, country level, population
LV11	jdr_sam_C	Job destruction rates, country level, sample
LV12	jdr_pop_M	Job destruction rates, mac-sector level, population
LV12	jdr_sam_M	Job destruction rates, mac-sector level, sample
LV13	jdr_pop_MS	Job destruction rates, macsec-szclass level, population
LV13	jdr_sam_MS	Job destruction rates, macsec-szclass level, sample
LV14	jdr_pop_N	Job destruction rates, NUTS2 level, population
LV14	jdr_sam_N	Job destruction rates, NUTS2 level, sample

Labor		
Variable Code	**Variable Name**	**Definition**
LV15	jdr_pop_2D	Job destruction rates, sector level, population
LV15	jdr_sam_2D	Job destruction rates, sector level, sample
LV19GH	**dhs_labor_growth_neg**	Davis-Haltiwanger-Schuh growth rate (from t-1): labor, only negative
LV20GH	**dhs_labor_growth_pos**	Davis-Haltiwanger-Schuh growth rate (from t-1): labor, only positive
LV21	L	Labor: number of employees in headcounts
LV21G1	L	Growth rate (from t-1): labor = number of employees
LV21G3	L	Growth rate (from t-3): labor = number of employees
LV21GH	dhs_labor_growth	Davis-Haltiwanger-Schuh growth rate (from t-1): labor
LV21GH	dhs_labor_growth2	Haltiwanger-Davis growth rate (from t-1): labor
LV24	rwage	Real wage
LV25	**fte**	Full time equivalents of firm employment
LV26	**jcr_pop_A**	Job creation rates, age-group level, population
LV27	**jcr_sam_A**	Job creation rates, age-group level, sample

Labor		
Variable Code	**Variable Name**	**Definition**
LV28	**jcr_pop_T**	Job creation rates, technology-group level, population
LV29	**jcr_sam_T**	Job creation rates, technology-group level, sample
LV30	**jdr_pop_A**	Job death rates, age-group level, population
LV31	**jdr_sam_A**	Job death rates, age-group level, sample
LV32	**jdr_pop_T**	Job death rates, technology-group level, population
LV33	**jdr_sam_T**	Job death rates, technology-group level, sample

5.2.4 Productivity Variables

Table 14: Productivity Variables

Productivity		
Variable Code	**Variable Name**	**Definition**
Estimates		
PEb0	tfp_0	TFP - Spec. 0 (CD, cost shares, quant.)
PEb0G1	**tfp_0**	Growth rate (from t-1): TFP - Spec. 0 (CD, cost shares, quant.)
PEb1	tfp_1	TFP - Spec. 1 (CD, OLS, quant.)
PEb2	tfp_2	TFP - Spec. 2 (TL, OLS, quant.)
PEb4	tfp_3	TFP - Spec. 3 (CD, ACF, quant.)
PEb5	tfp_4	TFP - Spec. 4 (TL, ACF, quant.)
PEd5	mrpk_0	Marg. rev. prod. of capital - Spec. 0 (CD, cost shares, quant.)
PEd6	mrpk_1	Marg. rev. prod. of capital - Spec. 1 (CD, OLS, quant.)
PEd7	mrpk_2	Marg. rev. prod. of capital - Spec. 2 (TL, OLS, quant.)
PEd9	mrpk_3	Marg. rev. prod. of capital - Spec. 3 (CD, ACF, quant.)
PEe0	mrpk_4	Marg. rev. prod. of capital - Spec. 4 (TL, ACF, quant.)
PEe1	mrpl_0	Marg. rev. prod. of labor - Spec. 0 (CD, cost shares, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEe2	mrpl_1	Marg. rev. prod. of labor - Spec. 1 (CD, OLS, quant.)
PEe3	mrpl_2	Marg. rev. prod. of labor - Spec. 2 (TL, OLS, quant.)
PEe5	mrpl_3	Marg. rev. prod. of labor - Spec. 3 (CD, ACF, quant.)
PEe6	mrpl_4	Marg. rev. prod. of labor - Spec. 4 (TL, ACF, quant.)
PEe7	oe_k_0	Output elasticity w.r.t. capital - Spec. 0 (CD, cost shares, quant.)
PEe9	oe_k_1	Output elasticity w.r.t. capital - Spec. 1 (CD, OLS, quant.)
PEf1	oe_k_2	Output elasticity w.r.t. capital - Spec. 2 (TL, OLS, quant.)
PEf5	oe_k_3	Output elasticity w.r.t. capital - Spec. 3 (CD, ACF, quant.)
PEf7	oe_k_4	Output elasticity w.r.t. capital - Spec. 4 (TL, ACF, quant.)
PEf9	oe_l_0	Output elasticity w.r.t. labor - Spec. 0 (CD, cost shares, quant.)
PEg1	oe_l_1	Output elasticity w.r.t. labor - Spec. 1 (CD, OLS, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEg3	oe_l_2	Output elasticity w.r.t. labor - Spec. 2 (TL, OLS, quant.)
PEg7	oe_l_3	Output elasticity w.r.t. labor - Spec. 3 (CD, ACF, quant.)
PEg9	oe_l_4	Output elasticity w.r.t. labor - Spec. 4 (TL, ACF, quant.)
PEh1	oe_m_0	Output elasticity w.r.t. interm. - Spec. 0 (CD, cost shares, quant.)
PEh2	oe_m_1	Output elasticity w.r.t. interm. - Spec. 1 (CD, OLS, quant.)
PEh3	oe_m_2	Output elasticity w.r.t. interm. - Spec. 2 (TL, OLS, quant.)
PEh5	oe_m_3	Output elasticity w.r.t. interm. - Spec. 3 (CD, ACF, quant.)
PEh6	oe_m_4	Output elasticity w.r.t. interm. - Spec. 4 (TL, ACF, quant.)
PEi3	rts_0	Returns to scale - Spec. 0 (CD, cost shares, quant.)
PEi4	rts_1	Returns to scale - Spec. 1 (CD, OLS, quant.)
PEi5	rts_2	Returns to scale - Spec. 2 (TL, OLS, quant.)
PEi7	rts_3	Returns to scale - Spec. 3 (CD, ACF, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEi8	rts_4	Returns to scale - Spec. 4 (TL, ACF, quant.)
PEi9	ln_tfp_0	log. TFP - Spec. 0 (CD, cost shares, quant.)
PEj0	ln_tfp_1	log. TFP - Spec. 1 (CD, OLS, quant.)
PEj1	ln_tfp_2	log. TFP - Spec. 2 (TL, OLS, quant.)
PEj2	ln_tfp_3	log. TFP - Spec. 3 (CD, ACF, quant.)
PEj3	ln_tfp_4	log. TFP - Spec. 4 (TL, ACF, quant.)
PEk2	**mrpk_5**	Marg. rev. prod. of capital - Spec. 5 (TL, OLS, rev.)
PEk3	**mrpk_6**	Marg. rev. prod. of capital - Spec. 6 (TL, ACF, rev.)
PEk4	**mrpl_5**	Marg. rev. prod. of labor - Spec. 5 (TL, OLS, rev.)
PEk5	**mrpl_6**	Marg. rev. prod. of labor - Spec. 6 (TL, ACF, rev.)
PEk6	**mrpm_0**	Marg. rev. prod. of interm. - Spec. 0 (CD, cost shares, quant.)
PEk7	**mrpm_1**	Marg. rev. prod. of interm. - Spec. 1 (CD, OLS, quant.)
PEk8	**mrpm_2**	Marg. rev. prod. of interm. - Spec. 2 (TL, OLS, quant.)
PEk9	**mrpm_3**	Marg. rev. prod. of interm. - Spec. 3 (CD, ACF, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEl0	**mrpm_4**	Marg. rev. prod. of interm. - Spec. 4 (TL, ACF, quant.)
PEl1	**mrpm_5**	Marg. rev. prod. of interm. - Spec. 5 (TL, OLS, rev.)
PEl2	**mrpm_6**	Marg. rev. prod. of interm. - Spec. 6 (TL, ACF, rev.)
PEl3	**oe_k_5**	Output elasticity w.r.t. capital - Spec. 5 (TL, OLS, rev.)
PEl4	**oe_k_6**	Output elasticity w.r.t. capital - Spec. 6 (TL, ACF, rev.)
PEl5	**oe_l_5**	Output elasticity w.r.t. labor - Spec. 5 (TL, OLS, rev.)
PEl6	**oe_l_6**	Output elasticity w.r.t. labor - Spec. 6 (TL, ACF, rev.)
PEl7	**oe_m_5**	Output elasticity w.r.t. interm. - Spec. 5 (TL, OLS, rev.)
PEl8	**oe_m_6**	Output elasticity w.r.t. interm. - Spec. 6 (TL, ACF, rev.)
PEm0	**re_k_0**	Revenue elasticity: capital - Spec. 0 (CD, cost shares, quant.)
PEm1	**re_k_1**	Revenue elasticity: capital - Spec. 1 (CD, OLS, quant.)
PEm2	**re_k_2**	Revenue elasticity: capital - Spec. 2 (TL, OLS, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEm3	**re_k_3**	Revenue elasticity: capital - Spec. 3 (CD, ACF, quant.)
PEm4	**re_k_4**	Revenue elasticity: capital - Spec. 4 (TL, ACF, quant.)
PEm5	**re_k_5**	Revenue elasticity: capital - Spec. 5 (TL, OLS, rev.)
PEm6	**re_k_6**	Revenue elasticity: capital - Spec. 6 (TL, ACF, rev.)
PEm7	**re_l_0**	Revenue elasticity: labor - Spec. 0 (CD, cost shares, quant.)
PEm8	**re_l_1**	Revenue elasticity: labor - Spec. 1 (CD, OLS, quant.)
PEm9	**re_l_2**	Revenue elasticity: labor - Spec. 2 (TL, OLS, quant.)
PEn0	**re_l_3**	Revenue elasticity: labor - Spec. 3 (CD, ACF, quant.)
PEn1	**re_l_4**	Revenue elasticity: labor - Spec. 4 (TL, ACF, quant.)
PEn2	**re_l_5**	Revenue elasticity: labor - Spec. 5 (TL, OLS, rev.)
PEn3	**re_l_6**	Revenue elasticity: labor - Spec. 6 (TL, ACF, rev.)
PEn4	**re_m_0**	Revenue elasticity: intermediates - Spec. 0 (CD, cost shares, quant.)
PEn5	**re_m_1**	Revenue elasticity: intermediates - Spec. 1 (CD, OLS, quant.)

Productivity		
Variable Code	**Variable Name**	**Definition**
PEn6	**re_m_2**	Revenue elasticity: intermediates - Spec. 2 (TL, OLS, quant.)
PEn7	**re_m_3**	Revenue elasticity: intermediates - Spec. 3 (CD, ACF, quant.)
PEn8	**re_m_4**	Revenue elasticity: intermediates - Spec. 4 (TL, ACF, quant.)
PEn9	**re_m_5**	Revenue elasticity: intermediates - Spec. 5 (TL, OLS, rev.)
PEo0	**re_m_6**	Revenue elasticity: intermediates - Spec. 6 (TL, ACF, rev.)
Ratios		
PR00	**rev_tot_costs**	Ratio: nominal revenue to total costs
PR01	**va_lc**	Ratio: nominal (positive) value-added to nominal labor cost
PR02	**va_tot_costs**	Ratio: nominal (positive) value-added to total costs
Values		
PV00	kprod_va	Capital productivity, computed as rva/nk
PV01	lnkprod_va	Log capital productivity real value added based: ln(rva/nk)

Productivity		
Variable Code	**Variable Name**	**Definition**
PV02	lnlprod_rev	Log labor productivity, real revenue based: ln(rrev/l)
PV02G1	lprod_rev	Growth rate (from t-1): labor prod., real revenue based
PV03	lnlprod_va	Log labor productivity, real value added based: ln(rva/l)
PV03G1	lprod_va	Growth rate (from t-1): labor prod., real value-added based
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
PV09	solowres_cs	Solow residual, weights in CD from cost shares
PV10	**lnlprod_va_fte**	Log labor productivity, real value added over FTE: ln(rva/fte)
PV11	**lprod_va_fte**	Labor productivity, real value added over FTE: rva/fte

5.2.5 Trade Variables

Table 15: Trade Variables

Trade		
Variable Code	**Variable Name**	**Definition**
	Categorical	
TC00	exp_dest	Categorical variable indicating export destinations: 1 only exports to EU-countries; 2 only exports to non-EU countries; 3 exports to both EU- and non-EU countries; 4 no exports at all
TC01	exp_imp_rel	Categorical variable indicating the direction of trade in t: 1 only exports, no imports; 2 only imports, no exports; 3 two-way trader: both imports and exports; 4 no imports and no exports
TC02	exp_time_3y	Categorical variable indicating the timing of exports: 1 exports only in t; 2 exports in t-2, t-1 and t, 3 exports in t-2 and t-1, but not in t; 4 no exports in t-2, t-1 and t; 5 no exports in t-2 and t, but exports in t-1
TC03	exp_top10	Categorical variable indicating large exporters: 1 top 10 exporter at country

Trade		
Variable Code	**Variable Name**	**Definition**
		level; 2 top 10 exporter at 2-digit sector level, but not at the country level; 3 firm is exporter, but not top 10 exporter; 4 non-exporter
TC04	imp_dest	Categorical variable indicating the origins of imports: 1 only imports from EU countries; 2 only imports from non-EU countries; 3 imports from both EU- and non-EU countries; 4 no imports at all
TC05	imp_time_3y	Categorical variable indicating the timing of imports: 1 imports only in t; 2 imports in t-2, t-1 and t, 3 imports in t-2 and t-1, but not in t; 4 no imports in t-2, t-1 and t; 5 no imports in t-2 and t, but exports in t-1
TC06	imp_top10	Categorical variable indicating large importers: 1 top 10 importer at country level; 2 top 10 importer at 2-digit sector level, but not at the country level; 3 firm is importer, but not top 10 importer; 4 non- importer

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
TD16	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.
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
TD90	imp_adj_con2	D = 1, if 2 years consecutive importer (t-1, t), 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)
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.

Trade		
Variable Code	**Variable Name**	**Definition**
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		
TR00	exp_adj_pop_C	Ratio: exports, adj., share of total, country level, pop.
TR00	**exp_adj_sam_C**	Ratio: exports, adj., share of total, country level, sample
TR01	exp_adj_pop_2D	Ratio: exports, adj., share of total, sector level, pop.
TR01	**exp_adj_sam_2D**	Ratio: exports, adj., share of total, sector level, sample
TR02	exp_adj_rev	Ratio: Export Ratio: exports adj. / nom. Revenue

Trade		
Variable Code	**Variable Name**	**Definition**
TR05	exp_adj_va_rev	Ratio: value added in export (adj.) revenue: exp*nva/nrev
TR06	**exp_adj_pop_A**	Ratio: exports, adj., share of total, firm age, pop.
TR07	**exp_adj_pop_T**	Ratio: exports, adj., share of total, techno. knowledge, pop.
TR08	**exp_pop_A**	Ratio: exports, share of total, firm age, pop.
TR09	**exp_pop_T**	Ratio: exports, share of total, techno. knowledge, pop.
TR11	**exp_ex_adj_pop_A**	Ratio: exports extra-EU, adj., share of total, firm age, pop.
TR12	**exp_ex_adj_pop_T**	Ratio: exports extra-EU, adj., share of total, techno. knowledge, pop.
TR15	**exp_in_adj_pop_A**	Ratio: exports intra-EU, adj., share of total, firm age, pop.
TR16	**exp_in_adj_pop_T**	Ratio: exports intra-EU, adj., share of total, techno. knowledge, pop.
TR17	**exp_in_pop_A**	Ratio: exports intra-EU, share of total, firm age, pop.
TR18	**exp_in_pop_T**	Ratio: exports intra-EU, share of total, techno. knowledge, pop.
TR36	imp_adj_pop_C	Ratio: imports, adj., share of total, country level, pop.

Trade		
Variable Code	**Variable Name**	**Definition**
TR36	**imp_adj_sam_C**	Ratio: imports, adj., share of total, country level, sample
TR37	imp_adj_pop_2D	Ratio: imports, adj., share of total, sector level, pop.
TR37	**imp_adj_sam_2D**	Ratio: imports, adj., share of total, sector level, sample
TR38	imp_adj_rev	Ratio: import Ratio: imports adj. / nom. Revenue
TR67	imp_exp_adj	Ratio: import intensity = imp/exp, adj.
Values		
TV02	exp	Exports
TV02G1	exp_val	Growth rate (from t-1): exports
TV02G1	exp_val_adj	Growth rate (from t-1): adjusted 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.

5.2.6 Other Variables

Table 16: Other Variables

Other		
Variable Code	**Variable Name**	**Definition**
	Categorical	
OC00	firm_age	1 "0-2 years" 2 "3-5 years" 3 "6-25 years" 4 "more than 25 years"
	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	llc	D = 1, if firm with limited liability, i.e. company or partnership 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 17: List of Variables in the Unconditional Descriptive Files

Unconditional variables
CD01_old_high_0
CD02_old_high_1_pop_2D
CD02_old_high_1_sam_2D
CD03_old_high_1_pop_C
CD03_old_high_1_sam_C
CD04_old_high_1_pop_M
CD04_old_high_1_sam_M
CD05_old_high_1_pop_MS
CD05_old_high_1_sam_MS
CD06_old_high_1_pop_N
CD06_old_high_1_sam_N
CD07_old_high_1_pop_T
CD07_old_high_1_sam_T
CD08_old_low_0
CD09_old_low_1_pop_2D
CD09_old_low_1_sam_2D
CD10_old_low_1_pop_C
CD10_old_low_1_sam_C
CD11_old_low_1_pop_M
CD11_old_low_1_sam_M
CD12_old_low_1_pop_MS
CD12_old_low_1_sam_MS
CD13_old_low_1_pop_N
CD13_old_low_1_sam_N
CD14_old_low_1_pop_T
CD14_old_low_1_sam_T
CD15_young_high_0
CD16_young_high_1_pop_2D
CD16_young_high_1_sam_2D
CD17_young_high_1_pop_C
CD17_young_high_1_sam_C
CD18_young_high_1_pop_M
CD18_young_high_1_sam_M
CD19_young_high_1_pop_MS
CD19_young_high_1_sam_MS
CD20_young_high_1_pop_N
CD20_young_high_1_sam_N
CD21_young_high_1_pop_T
CD21_young_high_1_sam_T
CD22_young_low_0
CD23_young_low_1_pop_2D
CD23_young_low_1_sam_2D
CD24_young_low_1_pop_C
CD24_young_low_1_sam_C

CD25_young_low_1_pop_M
CD25_young_low_1_sam_M
CD26_young_low_1_pop_MS
CD26_young_low_1_sam_MS
CD27_young_low_1_pop_N
CD27_young_low_1_sam_N
CD28_young_low_1_pop_T
CD28_young_low_1_sam_T
CE32_markdown_l_0
CE33_markdown_l_1
CE34_markdown_l_2
CE36_markdown_l_3
CE37_markdown_l_4
CE44_markup_0
CE45_markup_1
CE46_markup_2
CE48_markup_3
CE49_markup_4
CE56_markdown_k_5
CE57_markdown_k_6
CE58_markdown_l_5
CE59_markdown_l_6
CE60_markup_5
CE61_markup_6
CE62_markdown_m_5
CE63_markdown_m_6
CR00_top_rev_sam_C
CR01_top_rev_sam_M
CR02_top_rev_sam_2D
CR03_top_rev_sam_N
CR04_top_ifa_sam_2D
CR05_top_ifa_sam_C
CR06_top_ifa_sam_M
CR07_top_ifa_sam_N
CR08_top_l_sam_2D
CR09_top_l_sam_C
CR10_top_l_sam_M
CR11_top_l_sam_N
CR12_top_lc_sam_2D
CR13_top_lc_sam_C
CR14_top_lc_sam_M
CR15_top_lc_sam_N
CR16_top_rk_sam_2D
CR17_top_rk_sam_C
CR18_top_rk_sam_M
CR19_top_rk_sam_N

CR20_top_rva_sam_2D
CR21_top_rva_sam_C
CR22_top_rva_sam_M
CR23_top_rva_sam_N
CR24_top_ifa_sam_A
CR25_top_lc_sam_A
CR26_top_rev_sam_A
CR27_top_rk_sam_A
CR28_top_rva_sam_A
CR29_top_l_sam_A
CR30_top_ifa_sam_T
CR31_top_lc_sam_T
CR32_top_rev_sam_T
CR33_top_rk_sam_T
CR34_top_rva_sam_T
CR35_top_l_sam_T
CR40_mrktsha_l_pop_A
CR40_mrktsha_l_sam_A
CR41_mrktsha_l_pop_C
CR41_mrktsha_l_sam_C
CR42_mrktsha_l_pop_M
CR42_mrktsha_l_sam_M
CR43_mrktsha_l_pop_MS
CR43_mrktsha_l_sam_MS
CR44_mrktsha_l_pop_N
CR44_mrktsha_l_sam_N
CR45_mrktsha_l_pop_2D
CR45_mrktsha_l_sam_2D
CR46_mrktsha_l_pop_T
CR46_mrktsha_l_sam_T
CR54_mrktsha_rev_pop_A
CR54_mrktsha_rev_sam_A
CR55_mrktsha_rev_pop_C
CR55_mrktsha_rev_sam_C
CR56_mrktsha_rev_pop_M
CR56_mrktsha_rev_sam_M
CR57_mrktsha_rev_pop_MS
CR57_mrktsha_rev_sam_MS
CR58_mrktsha_rev_pop_N
CR58_mrktsha_rev_sam_N
CR59_mrktsha_rev_pop_2D
CR59_mrktsha_rev_sam_2D
CR60_mrktsha_rev_pop_T
CR60_mrktsha_rev_sam_T
CR68_mrktsha_va_pop_A
CR68_mrktsha_va_sam_A

CR69_mrktsha_va_pop_C
CR69_mrktsha_va_sam_C
CR70_mrktsha_va_pop_M
CR70_mrktsha_va_sam_M
CR71_mrktsha_va_pop_MS
CR71_mrktsha_va_sam_MS
CR72_mrktsha_va_pop_N
CR72_mrktsha_va_sam_N
CR73_mrktsha_va_pop_2D
CR73_mrktsha_va_sam_2D
CR74_mrktsha_va_pop_T
CR74_mrktsha_va_sam_T
CV02_hhi_rev_pop_T
CV02_hhi_rev_sam_T
CV03_hhi_rev_pop_A
CV03_hhi_rev_sam_A
CV04_hhi_rev_pop_C
CV04_hhi_rev_sam_C
CV05_hhi_rev_pop_M
CV05_hhi_rev_sam_M
CV06_hhi_rev_pop_N
CV06_hhi_rev_sam_N
CV07_hhi_rev_pop_2D
CV07_hhi_rev_sam_2D
CV10_hhi_ifa_pop_T
CV10_hhi_ifa_sam_T
CV11_hhi_ifa_pop_A
CV11_hhi_ifa_sam_A
CV12_hhi_ifa_pop_C
CV12_hhi_ifa_sam_C
CV13_hhi_ifa_pop_M
CV13_hhi_ifa_sam_M
CV14_hhi_ifa_pop_N
CV14_hhi_ifa_sam_N
CV15_hhi_ifa_pop_2D
CV15_hhi_ifa_sam_2D
CV18_hhi_l_pop_T
CV18_hhi_l_sam_T
CV19_hhi_l_pop_A
CV19_hhi_l_sam_A
CV20_hhi_l_pop_C
CV20_hhi_l_sam_C
CV21_hhi_l_pop_M
CV21_hhi_l_sam_M
CV22_hhi_l_pop_N
CV22_hhi_l_sam_N

CV23_hhi_l_pop_2D
CV23_hhi_l_sam_2D
CV26_hhi_lc_sam_T
CV27_hhi_lc_sam_A
CV28_hhi_lc_pop_C
CV28_hhi_lc_sam_C
CV29_hhi_lc_pop_M
CV29_hhi_lc_sam_M
CV30_hhi_lc_pop_N
CV30_hhi_lc_sam_N
CV31_hhi_lc_pop_2D
CV31_hhi_lc_sam_2D
CV34_hhi_rk_pop_T
CV34_hhi_rk_sam_T
CV35_hhi_rk_pop_A
CV35_hhi_rk_sam_A
CV36_hhi_rk_pop_C
CV36_hhi_rk_sam_C
CV37_hhi_rk_pop_M
CV37_hhi_rk_sam_M
CV38_hhi_rk_pop_N
CV38_hhi_rk_sam_N
CV39_hhi_rk_pop_2D
CV39_hhi_rk_sam_2D
CV42_hhi_rva_pos_pop_T
CV42_hhi_rva_pos_sam_T
CV43_hhi_rva_pos_pop_A
CV43_hhi_rva_pos_sam_A
CV44_hhi_rva_pos_pop_C
CV44_hhi_rva_pos_sam_C
CV45_hhi_rva_pos_pop_M
CV45_hhi_rva_pos_sam_M
CV46_hhi_rva_pos_pop_N
CV46_hhi_rva_pos_sam_N
CV47_hhi_rva_pos_pop_2D
CV47_hhi_rva_pos_sam_2D
CV48_firmrev_neg_pop_A
CV49_firmrev_pos_pop_A
CV50_firmrev_pos_sam_A
CV51_firmrev_neg_sam_A
CV52_firmrev_neg_pop_C
CV53_firmrev_pos_pop_C
CV54_firmrev_neg_pop_M
CV55_firmrev_pos_pop_M
CV56_firmrev_pos_sam_M
CV57_firmrev_neg_sam_M

CV58_firmrev_neg_pop_MS
CV59_firmrev_pos_pop_MS
CV60_firmrev_pos_pop_N
CV61_firmrev_neg_pop_N
CV62_firmrev_neg_pop_2D
CV63_firmrev_pos_pop_2D
CV64_firmrev_pos_sam_2D
CV65_firmrev_neg_sam_2D
CV66_firmrev_neg_pop_T
CV67_firmrev_pos_pop_T
CV68_firmrev_pos_sam_T
CV69_firmrev_neg_sam_T
CV70_firmrev_pos_sam_C
CV71_firmrev_neg_sam_C
CV72_firmrev_pos_sam_MS
CV73_firmrev_neg_sam_MS
CV74_firmrev_pos_sam_N
CV75_firmrev_neg_sam_N
FD00_absconstr
FD01_safe
FD05_zombie_intcov_pos
FD06_zombie_intcov
FD07_zombie_negprof
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
FR21_pcm_kvar
FR22_profitmargin
FR23_rd_costs
FR24_rd_m
FR25_rev_capcost

FR27_rev_lc
FR28_rev_lc_m
FR29_rev_m
FR30_rk_l
FR31_roa
FR32_trade_credit
FR33_trade_debt
FR35_va_rev
FR36_ifa_rev
FR37_invest_k
FR38_invest_rev
FR39_rd_share_rev
FR40_ener_costs
FR41_ener_rev
FR42_ener_va
FV00_capcost
FV01_debt
FV02_debt_fin
FV03_n_ener
FV04_nk
FV04_nk
FV05_nlc
FV05_nlc
FV06_nm
FV06_nm
FV07_nrd
FV08G1_nrev
FV08GH_dhs_rev_growth
FV08_nrev
FV08_nrev
FV09GH_dhs_rev_growth_neg
FV10GH_dhs_rev_growth_pos
FV10_nva
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
LR00_lc_rev
LR01_lc_va
LR02_tertshare
LR03_ulc
LV00_avg_wage
LV01_jcr_pop_C
LV01_jcr_sam_C
LV02_jcr_pop_M
LV02_jcr_sam_M
LV03_jcr_pop_MS
LV03_jcr_sam_MS
LV04_jcr_pop_N
LV04_jcr_sam_N
LV05_jcr_pop_2D
LV05_jcr_sam_2D
LV11_jdr_pop_C
LV11_jdr_sam_C
LV12_jdr_pop_M
LV12_jdr_sam_M
LV13_jdr_pop_MS
LV13_jdr_sam_MS
LV14_jdr_pop_N
LV14_jdr_sam_N
LV15_jdr_pop_2D
LV15_jdr_sam_2D
LV19GH_dhs_labor_growth_neg
LV20GH_dhs_labor_growth_pos
LV21G1_l
LV21G3_l
LV21GH_dhs_labor_growth
LV21GH_dhs_labor_growth2
LV21_l
LV24_rwage

LV25_fte
LV26_jcr_pop_A
LV27_jcr_sam_A
LV28_jcr_pop_T
LV29_jcr_sam_T
LV30_jdr_pop_A
LV31_jdr_sam_A
LV32_jdr_pop_T
LV33_jdr_sam_T
OD00_exit
OD01_firm_age_medium
OD02_firm_age_new
OD03_firm_age_old
OD04_firm_age_young
OD05_foreign_own
OD06_llc
OD07_publ_own
OV00_firm_age
OV01_firm_age_atexit
OV02_years_till_exit
PEb0G1_tfp_0
PEb0_tfp_0
PEb1_tfp_1
PEb2_tfp_2
PEb4_tfp_3
PEb5_tfp_4
PEd5_mrpk_0
PEd6_mrpk_1
PEd7_mrpk_2
PEd9_mrpk_3
PEe0_mrpk_4
PEe1_mrpl_0
PEe2_mrpl_1
PEe3_mrpl_2
PEe5_mrpl_3
PEe6_mrpl_4
PEe7_oe_k_0
PEe9_oe_k_1
PEf1_oe_k_2
PEf5_oe_k_3
PEf7_oe_k_4
PEf9_oe_l_0
PEg1_oe_l_1
PEg3_oe_l_2
PEg7_oe_l_3
PEg9_oe_l_4

PEh1_oe_m_0
PEh2_oe_m_1
PEh3_oe_m_2
PEh5_oe_m_3
PEh6_oe_m_4
PEi3_rts_0
PEi4_rts_1
PEi5_rts_2
PEi7_rts_3
PEi8_rts_4
PEi9_ln_tfp_0
PEj0_ln_tfp_1
PEj1_ln_tfp_2
PEj2_ln_tfp_3
PEj3_ln_tfp_4
PEk2_mrpk_5
PEk3_mrpk_6
PEk4_mrpl_5
PEk5_mrpl_6
PEk6_mrpm_0
PEk7_mrpm_1
PEk8_mrpm_2
PEk9_mrpm_3
PEl0_mrpm_4
PEl1_mrpm_5
PEl2_mrpm_6
PEl3_oe_k_5
PEl4_oe_k_6
PEl5_oe_l_5
PEl6_oe_l_6
PEl7_oe_m_5
PEl8_oe_m_6
PEm0_re_k_0
PEm1_re_k_1
PEm2_re_k_2
PEm3_re_k_3
PEm4_re_k_4
PEm5_re_k_5
PEm6_re_k_6
PEm7_re_l_0
PEm8_re_l_1
PEm9_re_l_2
PEn0_re_l_3
PEn1_re_l_4
PEn2_re_l_5
PEn3_re_l_6

PEn4_re_m_0
PEn5_re_m_1
PEn6_re_m_2
PEn7_re_m_3
PEn8_re_m_4
PEn9_re_m_5
PEo0_re_m_6
PR00_rev_tot_costs
PR01_va_lc
PR02_va_tot_costs
PV00_kprod_va
PV01_lnkprod_va
PV02G1_lprod_rev
PV02_lnlprod_rev
PV03G1_lprod_va
PV03_lnlprod_va
PV05_lnsr_cs
PV06_lprod_rev
PV07_lprod_va
PV09_solowres_cs
PV10_lnlprod_va_fte
PV11_lprod_va_fte
TD01_2w_exterior_adj
TD03_2w_extersale_adj
TD07_2w_interior_adj
TD09_2w_intersale_adj
TD13_2w_total_adj
TD14_exp
TD16_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
TD30_exp_ex
TD31_exp_ex_adj
TD54_exp_in
TD55_exp_in_adj
TD88_imp
TD90_imp_adj_con2
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
TR00_exp_adj_sam_C
TR01_exp_adj_pop_2D
TR01_exp_adj_sam_2D
TR02_exp_adj_rev
TR05_exp_adj_va_rev
TR06_exp_adj_pop_A
TR07_exp_adj_pop_T
TR08_exp_pop_A
TR09_exp_pop_T
TR11_exp_ex_adj_pop_A
TR12_exp_ex_adj_pop_T
TR15_exp_in_adj_pop_A
TR16_exp_in_adj_pop_T
TR17_exp_in_pop_A
TR18_exp_in_pop_T
TR36_imp_adj_pop_C
TR36_imp_adj_sam_C
TR37_imp_adj_pop_2D
TR37_imp_adj_sam_2D
TR38_imp_adj_rev
TR67_imp_exp_adj
TV02G1_exp_val
TV02G1_exp_val_adj
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 18: List of Decomposition Variables – Foster

Foster decomposition Variables
CE32_markdown_l_0_Wnlc
CE33_markdown_l_1_Wnlc
CE34_markdown_l_2_Wnlc
CE36_markdown_l_3_Wnlc
CE37_markdown_l_4_Wnlc
CE44_markup_0_Wnm
CE45_markup_1_Wnm
CE46_markup_2_Wnm
CE48_markup_3_Wnm
CE49_markup_4_Wnm
CE56_markdown_k_5_Wcap_cost
CE57_markdown_k_6_Wcap_cost
CE58_markdown_l_5_Wnlc
CE59_markdown_l_6_Wnlc
CE60_markup_5_Wnm
CE61_markup_6_Wnm
CE62_markdown_m_5_Wnm
CE63_markdown_m_6_Wnm
FR05_costcov_all_Wntc
FR11_ifa_k_Wk_nom
FR22_profitmargin_Wnrv
FR23_rd_costs_Wntc
FR24_rd_m_Wnm
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
PEb4_tfp_3_Wrrv
PEb5_tfp_4_Wrrv
PEe7_oe_k_0_Wrrv
PEe9_oe_k_1_Wrrv
PEf1_oe_k_2_Wrrv
PEf5_oe_k_3_Wrrv
PEf7_oe_k_4_Wrrv
PEf9_oe_l_0_Wrrv
PEg1_oe_l_1_Wrrv

PEg3_oe_l_2_Wrrv
PEg7_oe_l_3_Wrrv
PEg9_oe_l_4_Wrrv
PEh1_oe_m_0_Wrrv
PEh2_oe_m_1_Wrrv
PEh3_oe_m_2_Wrrv
PEh5_oe_m_3_Wrrv
PEh6_oe_m_4_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
PEl3_oe_k_5_Wrrv
PEl4_oe_k_6_Wrrv
PEl5_oe_l_5_Wrrv
PEl6_oe_l_6_Wrrv
PEl7_oe_m_5_Wrrv
PEl8_oe_m_6_Wrrv
PR00_rev_tot_costs_Wntc
PR01_va_lc_Wnlc
PR02_va_tot_costs_Wntc
PV00_kprod_va_Wrk
PV06_lprod_rev_Wl
PV07_lprod_va_Wl
PV09_solowres_cs_Wrva
PV11_lprod_va_fte_Wfte

Table 19: List of Decomposition Variables – OP

OP Decomposition Variables
CE32_markdown_l_0_Wnlc
CE33_markdown_l_1_Wnlc
CE34_markdown_l_2_Wnlc
CE36_markdown_l_3_Wnlc
CE37_markdown_l_4_Wnlc
CE44_markup_0_Wnm
CE45_markup_1_Wnm
CE46_markup_2_Wnm
CE48_markup_3_Wnm
CE49_markup_4_Wnm
CE56_markdown_k_5_Wcap_cost
CE57_markdown_k_6_Wcap_cost
CE58_markdown_l_5_Wnlc
CE59_markdown_l_6_Wnlc
CE60_markup_5_Wnm
CE61_markup_6_Wnm
CE62_markdown_m_5_Wnm
CE63_markdown_m_6_Wnm
FR05_costcov_all_Wntc
FR11_ifa_k_Wk_nom
FR22_profitmargin_Wnrv
FR23_rd_costs_Wntc
FR24_rd_m_Wnm
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
PEb4_tfp_3_Wrrv
PEb5_tfp_4_Wrrv
PEe7_oe_k_0_Wrrv
PEe9_oe_k_1_Wrrv
PEf1_oe_k_2_Wrrv
PEf5_oe_k_3_Wrrv
PEf7_oe_k_4_Wrrv
PEf9_oe_l_0_Wrrv
PEg1_oe_l_1_Wrrv

PEg3_oe_l_2_Wrrv
PEg7_oe_l_3_Wrrv
PEg9_oe_l_4_Wrrv
PEh1_oe_m_0_Wrrv
PEh2_oe_m_1_Wrrv
PEh3_oe_m_2_Wrrv
PEh5_oe_m_3_Wrrv
PEh6_oe_m_4_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
PEl3_oe_k_5_Wrrv
PEl4_oe_k_6_Wrrv
PEl5_oe_l_5_Wrrv
PEl6_oe_l_6_Wrrv
PEl7_oe_m_5_Wrrv
PEl8_oe_m_6_Wrrv
PR00_rev_tot_costs_Wntc
PR01_va_lc_Wnlc
PR02_va_tot_costs_Wntc
PV00_kprod_va_Wrk
PV06_lprod_rev_Wl
PV07_lprod_va_Wl
PV09_solowres_cs_Wrva
PV11_lprod_va_fte_Wfte

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 input market imperfections, markups, job creation and destruction rate indicators, and measures of revenue dynamism.

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, market power, and allocative efficiency. 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), labor (i.e. number of employees), and materials (i.e. intermediate inputs). In order to proxy physical quantities, nominal values are deflated using deflators available in EU-Klems that are specific for each variable (output, capital, intermediates), country, sector (2-digit), and year of observation.

The production function is estimated according to 5 different methodologies (0-4). Additionally, we estimate two revenue functions (5-6), where output is only deflated using the GDP deflator. In this case, elasticities capture the effect of input usage not only on production but also on product prices. They are then used to derive measures of market power.

0. In specification 0, we assume a Cobb-Douglas (CD) production function with constant return to scale (CRS) and derive the output elasticity of each input as the country-sector-year median cost-share (input expenditure over total costs[35]).
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).

[35] Total costs are computed as the sum of the expenditures in the input variables mentioned above, i.e. fixed assets, labor and intermediate inputs.

2. In specification 2, we assume a translog (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 function, and estimate the output elasticities following the two-step control function approach of Ackerberg, Caves and Frazer 2015 (ACF).
4. In specification 4, we apply the approach of ACF while assuming a TL production function.
5. In specification 5, we assume a TL revenue function, and estimate elasticities using OLS with year FE.
6. In specification 6, we assume a TL revenue function, and estimate elasticities with the approach of ACF.

Specification 0 is the only one involving a nonparametric estimation. Its approach, based on cost shares, requires two assumptions which are unnecessary in other specifications: constant returns to scale and no input market power.

Firms produce quantities of output, Q, while engaging in a standard cost minimization problem. For each input $j = \{K, L, M\}$, purchased in quantities q^j, at the market price p^j, and whose output elasticity is denoted by θ^j, the first-order condition implies:

(1) $p^j = \lambda \theta^j \frac{Q}{q^j}$

Since the multiplier λ here represents the marginal cost, one can introduce the markup (i.e. price over marginal cost, $\mu = \frac{P}{\lambda}$), rearrange and write:

(2) $\theta^j = \mu \frac{p^j q^j}{PQ}$

CRS implies that the sum of output elasticities equals unity. Thus, we obtain:

(3) $\mu \frac{\Sigma_j p^j q^j}{PQ} = 1$

Combining equations (2) and (3), we obtain our nonparametric estimator for each input's output elasticity, as its cost share.

(4) $\theta^j = \frac{p^j q^j}{\Sigma_j p^j q^j}$

In principle, this approach allows us to derive firm-year specific estimators. However, heterogeneity in firm-level cost shares may also reflect temporary adjustments or measurement errors. Thus, we assume output elasticities to be constant across firms belonging to the same 2-digit sector, and measure them using the median value of each sector-year cluster.

Regarding parametric estimations, we distinguish Cobb-Douglas production functions (specifications 1 and 3), translog production functions (specifications 2 and 4) and translog revenue functions (specifications 5 and 6). For each of these three subgroups, we implement two approaches: OLS and, to deal with transmission bias, ACF.

While the OLS approach (specifications 1, 2 and 5) is straightforward, the ACF methodology (specifications 3, 4 and 6) may deserve some further explanation. For simplicity, we only refer to the CD case. 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, labor l_t, and materials m_t according to the following equation (small letters represent logs, while capital letters represent levels).

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

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 a first-order Markov process with a productivity shock ξ_t that is observed by the firm at the beginning of the period. While our estimation approximates this process using a second-order polynomial in ω_{t-1} which allows nonlinear persistence, here for simplicity we represent the linear case of an AR(1) process.

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

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

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

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

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

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 (third-degree interactions are used in the CompNet code).

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

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

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

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.

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

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 expansion 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:

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

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 (specifications 1-4) and the revenue function (specifications 5-6) allows deriving the following set of indicators.

Output Elasticities ($\theta^m, \theta^k, \theta^l$)

Output elasticities for specifications 1 and 3 (Cobb-Douglas production functions), are simply the estimated coefficients:

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

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

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

Output elasticities for specifications 2 and 4 (translog production functions) are given by:

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

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

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

$$\theta^m = \eta^m \mu \theta^k = \eta^k \mu \theta^l = \eta^l \mu$$

Additionally, $\theta^k = \eta^k \mu \theta^l = \eta^l \mu$

Additionally, $\theta^l = \eta^l \mu$

Additionally,

(1) $\theta^m = \eta^m \mu$

(2) $\theta^k = \eta^k \mu$

(3) $\theta^l = \eta^l \mu$

Revenue Elasticities (η^m, η^k, η^l)

Revenue elasticities for specifications 5 and 6 (translog revenue functions) are given by:

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

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

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

Finally, we compute revenue elasticities for specifications 0 to 4 as the ratio of each input's output elasticity η to the markup μ.

(4) $\eta^k = \frac{\theta^k}{\mu}$

(5) $\eta^l = \frac{\theta^l}{\mu}$

(6) $\eta^m = \frac{\theta^m}{\mu}$

Returns to Scale (RTS)

This is given by the sum of the output elasticities of all inputs (only computed for production functions, i.e. specifications 0-4):

(1) 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):

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

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

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

TFP is only computed from production functions (specifications 0-4).

Market Power

Markups are generally defined as the ratio between the final good price and the marginal cost of production. In CompNet, we estimate mark-ups (μ) following the methodology by De Loecker & Warzinsky (2012), implemented using the optimality condition of the choice of intermediate inputs. A similar approach, based on Dobbelaere & Mairesse (2013) and Yeh et al. (2022), is used to measure labor markdowns (γ^l) assuming no market power on intermediate inputs:

(1) $\mu_t = \theta_t^m \frac{P_t R_t}{z_t M_t}$

(2) $\gamma_t^l = \frac{\theta_t^l}{\theta_t^m} \frac{z_t M_t}{w_t L_t}$

Where $P_t Q_t$ is nominal revenues (price times quantity), z_t is the price of materials and w_t the price of labor.

Additionally, we use revenue elasticities from specifications 5 and 6 to directly measure input market power, with no assumption on intermediate inputs.

(4) $\gamma_t^k = \eta_t^k \frac{P_t Q_t}{r_t K_t}$

(5) $\gamma_t^l = \eta_t^l \frac{P_t Q_t}{w_t L_t}$

(6) $\gamma_t^m = \eta_t^k \frac{P_t Q_t}{z_t M_t}$

We also compute markups from specifications 5 and 6 as the inverse of the sum of all revenue elasticities, assuming constant returns to scale.

(7) $\mu_t = \left(\eta_t^k + \eta_t^l + \eta_t^m\right)^{-1} = \left(\frac{\theta_t^k + \theta_t^l + \theta_t^m}{\mu_t}\right)^{-1}$

Marginal Product

We retrieve the Marginal Product (MP) of each production input from output elasticities and the Marginal Revenue Product (MRP) from revenue elasticities:

(1) $MP_k = \theta^k \frac{Q_t}{K_t}$

(2) $MP_m = \theta^m \frac{Q_t}{M_t}$

(3) $MP_l = \theta^l \frac{Q_t}{L_t}$

(4) $MRP_k = \eta^k \frac{P_t Q_t}{K_t}$

(5) $MRP_m = \eta^m \frac{P_t Q_t}{M_t}$

(6) $MRP_l = \eta^l \frac{P_t Q_t}{L_t}$

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 the case of labor productivity size is defined by the labor input, whereas in the case of TFP, size is defined in terms of gross output.

Table 20: 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_

Additionally to the standard way of computation for the decompositions according to Olley and Pakes (1996), the 9th vintage features alternative ways to decompose variables of interest. The variable *weight_type* encompasses these changes in the *op_decomp* files and different versions of input weighting (see Table 21).

The harmonic mean version (hm) differs from the standard case by utilizing harmonic means in the aggregation rather than the standard arithmetic mean. This is done for a subset of the variables, in particular monetary ones.

Also, there are now three versions with input shares fixed to 2004, 2009, or 2014. Instead of computing the input share of a firm in a given dimension anew every year, the input weights are fixed. This allows for the analysis of dynamics in the decomposition following an initial market-size allocation. Unlike the standard case, this dynamic version of Olley-Pakes allows for firms to drop out of the sample (i.e. input-shares no longer add up to 1 except in the base year).

Table 21 Olley-Pakes input weight types

Olley-Pakes input weight types (*weight_type*)	
standard	Standard decomposition as in Olley and Pakes (1996).
hm	Harmonic mean aggregation method instead of standard arithmetic mean.
wf04	Dynamic OP decomposition weighted with fixed input shares from 2004.
wf09	Dynamic OP decomposition weighted with fixed input shares from 2009.
wf14	Dynamic OP decomposition weighted with fixed input shares from 2014.

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

The covariance between size and productivity provides a snapshot 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:

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

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 reliable 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 levels.

5.3.3 Further Decompositions

We apply the productivity decomposition of Olley & Pakes (1996), as explained above, to further variables, including, among others, labor shares and markups. This provides data users with size-weighted aggregates of these variables and allows them 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 that, 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. (2017). The CompNet dataset includes three different zombie firm dummy specifications to identify distressed firms. These different indicators have different rationales

for 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:	*FD07_zombie_negprof*
Description:	Dummy equal 1 if firm reports negative profit for three consecutive years and is not considered to be high labor growth firm,[36] and 0 otherwise.

Variable:	*FD05_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 labor 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:	*FD06_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 labor 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 levels. 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.

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

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 do not result from true productivity differences between industries, but rather from 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. We, therefore, recommend using non-production-function-based variables for comparing productivity levels across industries, like our labor productivity variables.

A way to mitigate these issues of comparing levels between industries is to rely on comparing percentage changes in productivity between industries. 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., 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 either using 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 extend 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.

Due to the comparability issues of production-function-based productivity variables, the 9th 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 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 of 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)[37] 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:

$$(1)\ Prob(credit\_constraint) = \alpha + \beta_1 \cdot finlev_i + \beta_2 \cdot ifp_i + \beta_3 \cdot pm_i + \beta_4 \cdot coll_i + \beta_5 \cdot cashH_i + \beta_6 \cdot \ln(TA_i) + \gamma \cdot Z_{t,d,f,c} + \varepsilon,$$

where $finlev_i$ 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, TA_i are the total assets for firm i, and $Z_{t,d,f,c}$ is a set of dummies to control for time t, industry d, firm size f, and country c -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:

$$(2)\qquad SAFE\_score_i = -5.47 + 0.07 \cdot finlev_i + 0.46 \cdot ifp_i - 0.50 \cdot pm_i - 0.09 \cdot coll_i - 1.14\ cashH_i - 0.05 \ln(TA_i)$$

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

[37] The SAFE is conducted by the European Central Bank (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).

survey. In the final step, the *safe* dummy variable for a given firm is assigned a value of 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 (2018) 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 labor 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.[38] 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 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 practice 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.

Labor Market Power

Following a recent stream of work (e.g. Dobbelaere & Mairesse (2013), Mertens (2020)), we measure labor market power, γ, by dividing the markup formulas from De Loecker and

[38] We rely on the intermediate input decision of the firms, since we are aware that different degrees of (in)flexibility of labor 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).

Warzynski (2012), based on firms labor 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, labor 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 analyze 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:

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

Where X_{it} is the firm average employment; E_{it} and E_{it-1} are the employment in current and previous time point for a particular firm, respectively; 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, country, firms' age, and technological intensity dimensions. For example, at the two-digit sector level the growth rate has to be weighted by a firm weight in the following way:

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

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).

5.3.9 Revenue Dynamism

To measure the extent to which the market share is redistributed between firms over time, we introduce new variables, at each aggregation level, to measure the change in nominal revenue in the same manner like JCR and JDR in equation (5) and follow a weighting procedure similar to equation (6).

Equation (7) shows the calculation of the revenue growth rate:

$$(7) \qquad Z_{it} = 0.5 \cdot (T_{it} + T_{it-1}) \text{ and } gr_{it} = \frac{(T_{it} - T_{it-1})}{Z_{it}},$$

Where Z_{it} is the firm average nominal revenue; T_{it} and T_{it-1} are nominal revenue in current and previous time point for a particular firm, respectively; and gr_{it} is the firm-level growth rate of nominal revenue. Furthermore, we separate positive growth rates and negative growth rates in their respective variables.

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, and the 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 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 can 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 that 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 9th Vintage of the CompNet dataset, we mostly follow the outlier routine of the 8th vintage, with adding only slight additions based on our experience from the 8th vintage and discussion with data providers. The outlier routine is based on five different procedures. Notably, we do not drop any firm observation, but rather replace outlier values in specific variables with missing values.

In the 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 labor variable for firms that exhibit extraordinary growth rates in the labor variable. Here, extraordinary growth is defined as growth that violates the following condition:

$$\left(\frac{headcount(t)}{headcount(t-1)}-1\right)\sqrt[3]{\left(headcount(t-1)\right)}>80$$

$$\left(\frac{headcount(t-1)}{headcount(t)}-1\right)\sqrt[3]{\left(headcount(t)\right)}>80$$

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.

In the fourth part, we flag capital, turnover, intermediate input expenditure, labor cost, energy expenditures, and employment (headcounts and full-time equivalents) for the top and bottom two percent values in the distribution of various ratios within size classes and sectors separately. If values for variables are flagged in both dimensions, we replace these values with missing values. Specifically, we flag turnover with respect to the distributions of turnover over headcounts, turnover over capital, turnover over labor costs, and turnover over intermediates. We flag capital with respect to the distributions of turnover over capital and capital over headcounts. We flag headcounts with respect to the distributions of turnover over capital, capital over headcounts, and intermediates expenditures over headcounts. We flag full-time equivalents with respect to the distributions of turnover over full-time equivalents, capital over full-time equivalents, and intermediate expenditures over full-time equivalents. We flag intermediate expenditure with respect to the distribution of turnover over intermediate expenditures and headcounts over intermediate expenditures. We flag labor

costs with respect to the distributions of turnover over labor costs. Finally, we flag energy expenditures with respect to the distribution of turnover over energy expenditures.

Finally, we also flag outliers in total assets, cash and cash equivalents, cash flow, EBIT, interest paid, long-term debt, total debt, inventories, depreciation, accounts payable, accounts receivable, current liabilities, non-current liabilities, equity, profits and losses before taxes, other current liabilities, other non-current liabilities, other current assets, other fixed assets, intangible fixed assets, current assets, total fixed assets, dividends, research and development expenditures, and investment with respect to the distributions of turnover divided by these variables. We again flag values within sectors and size classes and only replace values with missing values that are flagged in both dimensions.

5.4.3 Weighting Procedure

The 9th Vintage of the CompNet dataset uses a weighting procedure which includes population weights from Eurostat. Weights are based on the number of firms in a given year, two-digit industry and employment size class.

To illustrate the general weighting procedure[39], 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

$$(1) \qquad v_i = \frac{firms_{tyz}}{m_{yz}} \ ,$$

where $firms_{tyz}$ is the number of firms, at a certain year t, of size class y, and industry z in the total population and m_{yz} is 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[40] in the output dataset) is then

[39] The Stata command *summarize* with option "aweights" is applied. For further details, please refer to *summarize* documentation. One has to take into account that "aweights" is not, strictly speaking, correct. However, according to Stata's resources and support (FAQs), they produce the correct population variance, mean, and percentiles. This allows using *sum* though "pweights" are not available.

[40] By definition *V* should be equal to $\sum firms_{tyz}$ if m_{yz} >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.

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

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

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

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

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

The weighted sample variance s^2 is

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

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

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

Consequently, weighted skewness is defined as

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

and the weighted kurtosis as

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

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 p^{th} percentile x_p, define $p = \frac{np}{100}$ and $W_i = \sum_{j=1}^{i} w_j$. Then one has to find the first index i for $W_i > p$.

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

5.4.4 Data Sources

Table 22: Country-specific Data Sources

Country	Data Source	Acronym	Institution Responsible for Source	Data Provider	Firms Included in Dataset*	Source Specific Information
Belgium	European Central Bank - Bank for the Accounts of Companies Harmonized	ECB - iBACH	Nationale Bank van België	European Central Bank		
Croatia	The Croatian Business Registry (Annual financial statements)	FINA	Financial Agency Croatia	Croatian National Bank	All (census)	Almost all raw data are from FINA
	Court Registry	FINA	Financial Agency Croatia			
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	Danmarks Nationalbank	Link	NACE2 Classification provided by Statistics Denmark
	General enterprise statistics	FIRM	Statistics Denmark		Link	
Finland	Structural business and financial statement statistics data	FSS	Tax Administration	Statistics Finland	Enterprise register (total population)	Breaks in 2006, 2013 Own NACE2 concordance used
	International trade statistics data	ITS	Finnish Customs	Statistics Finland	Enterprises trading goods	Intrastat thresholds
	Employment statistics data	FOLK	Finnish Centre for Pensions	Statistics Finland	All employers	
France[41]	Élaboration des statistiques annuelles d'entreprises	Esane	Statistics France	Statistics France	Around 4 million firms in source, 2 million in sample.	2008-2020
	Système Unifié de Statistiques d'Entreprises	Suse	Statistics France		Around 2 million firms in source.	2004-2007
	Base Tous Salariés		Statistics France		Around 2 million firms in source.	2004-2020
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 annual turnover

[41] Sample composition changed in 2017 and 2018

	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 Authorities	Central Bank of Hungary	230,000 – 420,000 depending on the year, all double entry bookkeeping firms	Non-mandatory variables for tax-records are underreported. (e.g., 30% of firms do not report the number of employees). 2003-2020
	Business Registry	VR	Central Statistical Office (KSH)		all	2003-2020
	Pension Payment data, including the work history		Pension Payment Directorate (ONYF)		Firms having employees on the 15th of March were included in the sample.	2003-2019

Italy	European Central Bank - Bank for the Accounts of Companies Harmonized	ECB - iBACH	Cerved / Banca d'Itália	European Central Bank		
Latvia	Central Statistical Bureau of Latvia	CSP	Central Statistical Bureau of Latvia	Central Statistical Bureau of Latvia	1,424,029	2007-2018 (excluding 2014 for financial data).
Lithuania	Statistical Survey on the Business Structure (Annual questionnaire F-01)	F01	Statistics Lithuania	Central Bank of Lithuania	NFC drawn from total economy. Unbalanced panel data, from around 5000-70,000 firms.	2000-2020
	Business Register	BR	Centre of Register			
	Customs declaration	CU	Customs of the Republic of Lithuania			
Netherlands	General Business Register	ABR	Statistics Netherlands	Statistics Netherlands	1- 2 million annually (enterprise groups).	2006-2020. Link
	Business Demographic Framework	BDK	Statistics Netherlands		1.3-2.1 million annually (enterprises).	2007-2020. Link
	The Statistics on Non-Financial Enterprises NFO: A survey for large non-financial enterprises (SFGO) and of tax	SFGO; SFKO	Statistics Netherlands		123,000-262,000 annually (enterprise groups).	2000-2020. Link

	data for small non-financial enterprises (SFKO)					
	International trade in goods	IHG	Statistics Netherlands		205,000-548,000 annually (enterprises).	2010-2020. Link
Poland	Report on revenues, costs and financial result as well as on expenditure on fixed assets	F-01	Statistics Poland	Central Bank Poland	139,310 unique firms with varied time coverage (on average 53,500 firms a year).	Some variables are available only from one database.
	Annual enterprise survey	SP	Statistics Poland			
Portugal	Integrated Business Accounts System	SCIE	Statistics Portugal	GEE - Office for Strategy and Studies - Ministry of Economy.	1,301,000	The national statistical system in Portugal produces the Microdata database through administrative procedures, thus this dataset encompasses the total population of firms. The national statistical system has at its disposal basic statistical data exhaustively covering all Portuguese companies, encompassing a wide range of statistical variables originating in a single source, with some

						administrative data for individual enterprises (sole proprietors or independent workers) received on the basis of a protocol with Tax Administration. Statistics Portugal conducts a set of validations and quality control rules. Statistics Portugal makes a top-down analysis, aggregating the data by the level of the required analysis, making data comparisons with previous years and within the year with other companies in the same strata. With this process, outliers are identified and microdata is corrected whenever it's needed. Aside from the validations and quality control rules that are activated by Statistics Portugal, there is no further pre-treatment of missing values.
Romania	Balance sheet information on non-financial enterprises	Bal. Sheet	Ministry of Public Finances	National Bank of Romania	384,511 observations of	

					firms with 20 or more employees.	
			The National Institute of Statistics (NIS).			
Slovakia	Annual report on production industries	Reports	Statistics Slovakia	National Bank of Slovakia	Population of firm with 20 and more employees (number increasing in time, up to around 10,000 in 2020).	2000-2020. True period covered differs across variables, only for firms with 20 and more employees.
	Statistical register of organizations	Register			Restricted to firms included in the core dataset (Reports).	2000-2020.
	Foreign trade statistics	Customs			Restricted to firms included in the core dataset (Reports).	2004-2020. Only flows above official export and import thresholds.
	Bisnode database	Bisnode	Bisnode Slovakia		Restricted to firms included in the core dataset (Reports).	2020.

						Used only for imputation of 2020 values of fixed assets and related variables[42].
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: 2002-2021.	
Spain	European Central Bank - Bank for the Accounts of Companies Harmonized	ECB - iBACH	Banco de España / Mercantile Registries	European Central Bank		
Sweden	Structured business statistics	SBS	Tax Authority	Tax Authority	14,816,959	2003-2020. 2003-2007: All non-financial enterprises that, in the reference year, conducted their main activities in NACE 1.1 Sections A-O excluding J and L. From 2008 and onwards: All non-financial enterprises that, in the

[42] See section 3.3.2.

						reference year, conducted their main activities in NACE Rev 2 sections A-S, excluding K and O.
	International trade in goods	ITGS	Statistics Sweden and Custom Authority	Statistics Sweden and Custom Authority	251,600	2003-2020. The threshold in the Intrastat survey has been changed a few times during the period covered. More information: See paragraphs A.14 and B.4 in the Quality documentation: link. ITGS statistics are based on Regulation (EC) No 223/2009 of the European Parliament and of the Council on European statistics.
	Business register	BR	Statistics Sweden/Tax Authority	Statistics Sweden/Tax Authority	32,249,454	2003-2020. Contains all administrative information on all enterprises registered in each reference year.
	Labor statistics based on administrative sources	RAMS	Statistics Sweden/Tax Authority	Statistics Sweden/Tax Authority	3,743,394	2003-2020. RAMS population consist of all active enterprises which uses at

						least one person employed at 31 of December in the reference year.
Switzerland	Value added statistic (Wertschöpfungsstatistik)	(Link) (Link)	Federal statistical office	Federal statistical office	Around 20,000 yearly.	2009-2020.
United Kingdom						

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 as shown below in Table 23.

Table 23: 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, licenses, 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, licenses, 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 realized 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.		
Labor 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 labor, 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	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 - labor cost - depreciation	

Variable	First Best	Second Best	Third Best
	expense - Depreciation and amortization 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	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	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	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 2-digit	Two-digit division number according to NACE Rev. 2		
Industry 3-digit	Three-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.		
Labor	Headcounts of the number of employees (yearly average) with employed shareholders/owners excluded	Headcounts at a certain date	Full time equivalent
Labor – FTE	Number of employees in full time equivalents (FTE)		
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. nationalized 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 shares of the firm directly or indirectly; 4 = otherwise (private firm)		
Share of skilled labor	Share of employees having post-secondary (tertiary) education. Tertiary education is the educational level following		

Variable	First Best	Second Best	Third Best
	the completion of a school providing a secondary education. It includes universities as well as trade schools, colleges and vocational training.		

CompNet dataset includes deflated variables created based on deflators derived from Eurostat National Accounts (nama_10_a64, nama_10_a64_p5, and prc_hicp_aind for the consumer price index). Table 23a provides a list of deflated variables and the deflators that have been used. However, for some countries in the CompNet dataset, for which data for output deflator and intermediate input deflator are not available from Eurostat, we use the output deflator and intermediate input deflator from the EUKLEMS national accounts data[43] . Those EUKLEMS price indices are in national currencies and had to be converted to Euro by using the official exchange rates.

- We use the output deflator and intermediate input deflator from the EUKLEMS dataset for 12 countries: Croatia, France, Italy, Lithuania, Malta, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and United Kingdom.
- Switzerland and Croatia do not publish any fixed capital consumption deflator. Instead, the Eurostat gross investment deflator from at the two-digit sector level was used.
- Malta does not publish any output or intermediate input deflator. Hence, the Eurostat value-added deflator had to be used.
- All deflators were recalculated for the base year 2005.
- All deflators – except for the consumer price index – were applied at the two-digit sector level. Missing sector values had to be substituted by neighboring sectors (e.g. for the missing intermediate input deflators of sectors 11 and 12 the intermediate input deflator of sector 10 was used.)

[43] The EUKLEMS data can be downloaded here https://euklems-intanprod-llee.luiss.it/download/ .

Table 23: List of deflated variables

Deflated variable	Definition	Original Deflator Used
FV13_rifa	Real intangible fixed assets	Implicit price deflator base year 2010 EUR for fixed capital consumption (=depreciation); unit PD10_EUR, indicator P51C
FV14_rk	Real capital	Implicit price deflator base year 2010 EUR for fixed capital consumption (=depreciation); unit PD10_EUR, indicator P51C
FV15_rlc	Real labor costs	Average harmonized consumer price index, annual data, base year 2015 EUR; indicator CP00, unit INX_A_AVG
FV16_rm	Real intermediate inputs	Implicit price Deflator in EUR base year 2010 intermediate consumption; unit PD10_EUR, indicator P2; KLEMS-only countries: II_PI intermediate inputs, price index base year 2015 (NAC), converted to EUR
FV17_rrev	Real revenue	Implicit Price Deflator in EUR base year 2010 output; unit PD10_EUR, indicator P1; KLEMS-only countries: GO_PI gross output, price index base year 2015 (NAC), converted to EUR.
FV18_rva	Real value-added	Implicit Price Deflator in EUR base year 2010 total gross value added; unit PD10_EUR, indicator B1G
FV19_rva_pos	Real value-added, only positive values	Implicit Price Deflator in EUR base year 2010 total gross

		value added; unit PD10_EUR, indicator B1G
FV29_rinvest	Real investment	Implicit price deflator in EUR base year 2010 gross fixed capital formation – total fixed assets; unit PD10_EUR, asset N11G, indicator P51G
FV30_rrd	Real R&D expenditure	Implicit price deflator in EUR base year 2010 gross fixed capital formation – total fixed assets; unit PD10_EUR, asset N11G, indicator P51G
LV24_rwage	Real wage	Average harmonized consumer price index, annual data, base year 2015 EUR; indicator CP00, unit INX_A_AVG

Table 24: Country-specific Definitions of Input Variables

Variable / Country	BE*	HR*	CZ	DK	FI	FR	DE	HU	IT*	LV	LT	MT	NL	PL	PT	RO	SK	SI	ES	SE	CH	UK
Fixed assets	1	2	1	2	1	1	2	2	1	2	2	2	1	1	2	1	0[44]	2	1	2	2	N/A
Capital	2	2	1	2	0[45]	1	2	3	1	2	2	2	1	1[46]	2	2[47]	2	2	1	2	2	
Intangible fixed assets	1	2	1	2	1	1	2	2	1	2	2	2	1	1[46]	2	1	2	2	1	2	2	N/A
Other fixed assets	1	2	1	2	1	1	2	2	1	2	2	2	0[48]	1[46]	2	1	N/A	2	1	2	2	N/A
Current assets	1	2	1	2	1	1	2	2	1	2	2	2	3	1	2	1	0[49]	2	1	2	2	
Cash and cash equivalents	1	2	0[50]	2	1	1	2	2	1	2	2	2	1	1	2	2[51]	N/A	2	1	2	2	
Total inventories	2	2	1	2	1	1	0	2	1	2	2	2	1	1	2	1	2	2	1	2	2	N/A
Accounts receivable	1	2	1	2	1	1	0	2	1	2	2	2	1	1	2	1	2	2	1	2	2	N/A
Other current assets	1	2	1	2	1	1	2	2	1	2	2	2	0	1	2	1[52]	N/A	2	1	2	2	
Total assets	1	2	1	2	1	1	2	2	1	2	2	2	1	1	2	1[53]	2	2	1	2	2	N/A
Total shareholder funds (equity)	1	2	1	2	1	1	2	2	1	2	2	2	1	1[46]	2	1	2	2	1	2	2	N/A
Non-current liabilities	0	2	1	2	1	1	2	2	1	2	2	2	3	1[46]	2	0[54]	N/A	2	0	2	2	N/A
Long-term debt	0	2	1	2	N/A	1	2	2	1	2	2	2	2	1[46,55]	2	N/A	N/A	2	0	2	2	

[44] Tangible fixed assets plus intangible fixed assets (end of the year value).
[45] Tangible fixed assets (yearly average).
[46] Available only from the SP survey. Smaller number of observations compared to the full database.
[47] Tangible fixed assets: beginning- and end-year values are averaged.
[48] The value of the participation in the capital of other companies plus the total of receivables and investments with a (remaining) term of more than one year.
[49] Total assets minus total fixed assets (end of the year value).
[50] Cash plus bank accounts only (yearly average).
[51] Includes cash: beginning- and end-year values are averaged.
[52] Calculated as total current assets minus accounts receivable minus total inventories.
[53] Calculated as total fixed assets plus current assets.
[54] Long-term debt: beginning- and end-year values are averaged.
[55] Prior to 2011, the variable covered only loans without bonds. Since 2011, the variable covers all components of the first-best definition.

Variable / Country	BE*	HR*	CZ	DK	FI	FR	DE	HU	IT*	LV	LT	MT	NL	PL	PT	RO	SK	SI	ES	SE	CH	UK
Other non-current liabilities	0	2	1	2	N/A	1	2	N/A	1	2	2	2	0[56]	1[46]	2	N/A	N/A	2	1	2	2	N/A
Current liabilities	0	2	1	2	1	1	2	2	1	2	2	2	3	1[46]	2	0[57]	N/A	2	0	2	2	N/A
Total debt	0	2	1	2	1	1	2	2	1	2	2	2	1	1[46,55]	2	N/A	0[58]	2	0	2	2	N/A
Accounts payable	1	2	1	2	1	1	2	2	1	2	2	2	1	1[46]	2	1	2	2	1	2	2	N/A
Other current liabilities	0	2	1	N/A	N/A	1	2	2	1	2	2	2	1	1[46]	2	N/A	N/A	2	1	2	2	N/A
Gross Output	2	1[59]	1	2	1	1	1	3	1	2	2	2	1	1[59]	1	2	0[60]	1	1	1	2	
Labor cost	1	1	1	1	1	1	2	1	1	1	1	1	2	1	0[61]	1	1	1	1	1	1	
Intermediate inputs	2	1	1	2	1	1	2	1	1	2	2	2	1	1	1	2	1	1	1	1	1	
Energy Input	N/A	1	N/A	1	1	N/A	2	N/A	N/A	N/A	1	1	N/A	1	1	N/A	1	1	N/A	1	N/A	
R&D expenditures	N/A	1[62]	N/A	N/A	N/A	N/A	2	N/A	N/A	1	N/A	1	N/A	1[63]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Operating profit/loss (EBIT)	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	N/A
Interest paid and financial charges	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	N/A
Depreciation	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	N/A
Profits and losses before taxes	2	1	1	2	1	1	1	1	2	1	2	2	N/A	1	1	1	2	1	2	1	1	
Cash flow (from profit/loss statement)	3	3	1	3	1	1	1	4	3	3	3	3	N/A	3	0[64]	3	3	1	3	3	2	N/A
Dividends	1	N/A	1	1	1	1	0	1	1	0	N/A	1	1	N/A	N/A	N/A	1	N/A	1	1	1	

[56] Includes provisions.
[57] Short term debt: beginning- and end-year values are averaged.
[58] Bank loans plus financial assistance (end of the year value).
[59] The variable does not include subcomponents 2 and 3 of the first-best definition, see table 23.
[60] Turnover has to be valued at factor cost: gross sales revenues minus customer discounts, returns and allowances; excluding indirect taxes but including subsidies on products and production.
[61] The variable is calculated with all the accounting definitions defined in the first-best definition but estimated for total headcounts equivalent to employees plus employed shareholders/owners.
[62] Available only since 2016.
[63] Available only from the F-01 survey. Smaller number of observations compared to the full database.
[64] The variable is defined as the first-best definition except for (increase in liabilities minus decrease in liabilities).

Variable / Country	BE*	HR*	CZ	DK	FI	FR	DE	HU	IT*	LV	LT	MT	NL	PL	PT	RO	SK	SI	ES	SE	CH	UK
Gross Investment	0	1	1	1	1[65]	1	1	N/A	1	N/A	1	1	N/A	1[63]	1	N/A	1	N/A	1	1	1	
Effective tax rate	1	1	1	N/A	1	1	1	1	1	1	1	1	1	1[66]	N/A	N/A	1	1	1	1	1	N/A
Export value	N/A	1	1	2	1	1	1	1	N/A	N/A	2	2	1	1	1	2	1	1	N/A	1	N/A	N/A
Exports to extra-EU	N/A	N/A	1	N/A	0[67]	N/A	0	N/A	N/A	N/A	2	2	1	N/A	1	2	1	3	N/A	1	N/A	N/A
Exports to intra-EU	N/A	N/A	1	N/A	0[67]	N/A	0	N/A	N/A	N/A	2	2	1	N/A	1	2	1	3	N/A	1	N/A	N/A
Import value	N/A	1	1	2	1	N/A	0	N/A	N/A	N/A	2	2	1	1[63]	1	2	1	N/A	N/A	1	N/A	N/A
Imports from extra-EU	N/A	N/A	1	N/A	0[67]	N/A	0	N/A	N/A	N/A	2	2	1	N/A	1	2	1	N/A	N/A	1	N/A	N/A
Imports from intra-EU	N/A	N/A	1	N/A	0[67]	N/A	0	N/A	N/A	N/A	2	2	1	N/A	1	2	1	N/A	N/A	1	N/A	N/A
Industry 2-digit	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	N/A
Industry 3-digit	1	1	1	1	1	1		N/A	1	1	N/A	N/A	N/A	1	1	N/A	1	1	1	1	1	N/A
NUTS2	N/A	1	1	1	1	1	1	1	N/A	1	1	1	N/A	1	1	1	1	1	N/A	1	1	N/A
Firm's birth year	1	1	1	1	N/A	1	1	1	1	1	1	1	1	N/A	N/A	1	1	1	1	1	N/A	
Firm's exit year	1	1	N/A	1	N/A	0	0	1	1	N/A	1	1	1	N/A	N/A	N/A	1	N/A	1	1	N/A	N/A
Foreign ownership	N/A	1	1	1	1	1	0	1	N/A	1	1	1	1	1	N/A	N/A	1	N/A	0	1	N/A	
Labor - headcounts	3	1	1	2	1[68]	1	2	1	1	1	1	1	1	1	0[69]	1	1	3	3	2	2	
Labor - FTE	1	1	N/A	1	1	1	N/A	N/A	N/A	N/A	N/A	N/A	1	1	N/A	N/A	1	1	1	1	1	N/A

[65] Intangible investments are more heavily imputed.
[66] Computed based on raw data. Negative values (such as paying income tax when having negative profit) were set to zero.
[67] Based on time-inconsistent EU definition (yearly changes).
[68] There is a break in 2006.
[69] Headcounts (employees plus employed shareholders/owners) at the end of calendar year.

Variable / Country	BE*	HR*	CZ	DK	FI	FR	DE	HU	IT*	LV	LT	MT	NL	PL	PT	RO	SK	SI	ES	SE	CH	UK
Legal form	1	1	1	1	1	N/A	0	1	1	1	1	1	N/A	N/A	1	N/A	1	N/A	1	1	N/A	
Public or non-profit enterprise	N/A	1	1	N/A	1	N/A	0	1	N/A	N/A	1	N/A	1	1	N/A	N/A	1	N/A	N/A	1	N/A	
Share of skilled labor	N/A	N/A	N/A	N/A	0[70]	N/A	0	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1	N/A	N/A

Notes: 1: "first-best definition" according to Table 21 – 2: "second-best definition" – 3: "third-best definition"[71] – 0: "definition available, but other definition than in table 21 – N/A: "not available."

[70] Share of personnel having post-secondary (tertiary) education.
[a] Second best definition (2) for manufacturing sectors

5.4.6 List of Macro Sectors and Industries

Table 25: List of Macro-Sectors and Industries Included in the 9th 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

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

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

Table 26: Categorization of industries by technological intensity

Aggregation	Category	2-digit industries in CompNet
Aggregation within Manufacturing sector	High-technology industries	21: Manufacture of basic pharmaceutical products and pharmaceutical preparations 26: Manufacture of computer, electronic and optical products
	Medium-high-technology industries	20: Manufacture of chemicals and chemical products 27: Manufacture of electrical equipment 28: Manufacture of machinery and equipment 29: Manufacture of motor vehicles, trailers and semitrailers 30: Manufacture of other transport equipment
	Medium-low-technology industries	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 33: Repair and installation of machinery and equipment
	Low technology industries	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 31: Manufacture of furniture 32: Other manufacturing
Aggregation within service sectors	Knowledge-intensive services	50: Water transport 51: Air transport 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 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 78: Employment activities 80: Security and investigation activities
	Less knowledge-intensive services	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 49: Land transport and transport via pipelines 52: Warehousing and support activities for transportation

		53: Postal and courier activities 55: Accommodation 56: Food and beverage service activities 68: Real estate activities 77: Rental and leasing activities 79: Travel agency, tour operator and other reservation service and related activities 81: Services to buildings and landscape activities 82: Office administrative, office support and other business support activities

6. References

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411-2451. https://doi.org/10.3982/ECTA13408

Bighelli, T., di Mauro, F., Melitz, M. J., & Mertens, M. (2022). European firm concentration and aggregate productivity. *Journal of the European Economic Association*. https://doi.org/10.1093/jeea/jvac040

Boisvert, R. N. (1982). *The translog production function: Its properties, its several interpretations and estimation problems* (A.E. Res. 82-28). Department of Agricultural Economics, Cornell University. https://dx.doi.org/10.22004/ag.econ.182035

Caballero, R. J., Hoshi, T., & Kashyap, A. K. (2008). Zombie lending and depressed restructuring in Japan. *The American Economic Review*, 98(5), 1943–1977. https://doi.org/10.1257/aer.98.5.1943

Council of the European Union. (1993). *Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community* (Official Journal L 076). Organization for Security and Cooperation in Europe (OSCE).

Davis, S. J., Haltiwanger, J., & Schuh, S. (1996). Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, 8(4), 297–315. https://doi.org/10.1007/BF00393278

De Loecker, J., & Warzynski, F. (2012). Markups and firm-level export status. *The American Economic Review*, 102(6), 2437–2471. https://doi.org/10.1257/aer.102.6.2437

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2), 445–510. https://doi.org/10.3982/ECTA11042

Dobbelaere, S., & Mairesse, J. (2013). Panel data estimates of the production function and product and labor market imperfections. *Journal of Applied Econometrics*, 28(1), 1–46. https://doi.org/10.1002/jae.1256

Ferrando, A., & Ruggieri, A. (2018). Financial constraints and productivity: Evidence from Euro area companies. *International Journal of Finance & Economics,* 23(3), 257-282. https://doi.org/10.1002/ijfe.1615

Ferrando, A., Matteo, L., Altomonte, C., Blank, S., Felt, MH., Meinen, P., Neugebauer, K., & Siedschlag, L. (2015). *Assessing the financial and financing conditions of firms in Europe: The financial module in CompNet* (ECB Working Paper Series, No 1836). European Central Bank. https://dx.doi.org/10.2139/ssrn.2641070

Foster, L., Haltiwanger, J., & Krizan, C. J. (2006). Market selection, reallocation, and restructuring in the U.S. retail trade sector in the 1990s. *The Review of Economics and Statistics*, 88(4), 748–758. https://doi.org/10.1162/rest.88.4.748

Gal, P. (2013). *Measuring total factor productivity at the firm level using OECD-Orbis* (OECD Economics Department Working Papers, No. 1049). OECD Publishing. https://doi.org/10.1787/5k46dsb25ls6-en.

Haltiwanger, J., Kulick, R., & Syverson, C. (2018). *Misallocation measures: The distortion that ate the residual* (National Bureau of Economic Research Working Paper series, No. 24199). National Bureau of Economic Research. https://doi.org/10.3386/w24199

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. *The Review of Economics and Statistics*, 95(2), 347–361. https://doi.org/10.1162/REST_a_00288

Hirschman, A. O. (1964). The paternity of an index. *The American Economic Review*, 54(5), 761–762.

Kehrig, M. (2011). *The cyclicality of productivity dispersion* (U.S. Census Bureau Centre for Economic Studies Working Paper No. CES-11-15). U.S. Census Bureau.

López-Garcia, P., Aglio, D., Bräuer, R., Haug, P., Kerke, J. P., Mertens, M., Sartori, M., Serafini, R., Soares, A. C., & Mattioli, A. Z. (2018, October 8). *CompNet's 6th vintage of data: Novelties and main stylised facts*. The Competitiveness Research Network. https://www.comp-net.org/fileadmin/_compnet/user_upload/Documents/Cross-country_report_updated_28.02.2018.pdf

McGowan, M. A., Andrews, D., & Millot, V. (2017). *The walking dead?: Zombie firms and productivity performance in OECD* countries (OECD Economics Department Working Papers, No. 1372). OECD Publishing. https://doi.org/10.1787/180d80ad-en.

Mertens, M. (2020). *Micro-mechanisms behind declining labor shares: Market power, production processes, and global competition* (IWH-CompNet Discussion Papers, 3/2019). Halle Institute for Economic Research and the Competitiveness Research Network (CompNet).

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263–1297. https://doi.org/10.2307/2171831

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

The Competitiveness Research Network (CompNet). (2020). *Firm productivity report.*

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters,* 104(3), 112-114. https://doi.org/10.1016/j.econlet.2009.04.026