CompNet The Competitiveness Research Network

User Guide for the 6th Vintage of the CompNet Dataset

This Version: 28.11.2018

This User Guide has been written by Marco Christophori, Fabian Marder, Vera Schramm, Dimitri Trofimenkov (all IWH), Daniele Aglio, Jan Paul van de Kerke, Matteo Sartori and Alessandro Zona Mattioli (all ECB). It has been commented by Richard Bräuer, Matthias Mertens (IWH), Filippo di Mauro (NUS), Paloma López-Garcia (ECB) and approved by Peter Haug.

Table of Content

List of Tables

List of Figures

1. Introduction

This user guide aims at providing the interested researcher with all information necessary to have an easy start with the usage of the 6th Vintage of the CompNet dataset. The user guide is part of a group of complementary documents for the newest vintage of the CompNet dataset. The other two are the 6th Vintage CompNet Cross-Country Report (López-Garcia et al., 2018) and the CompNet Comparability Report (Altomonte et al., forthcoming). To get the most out of the CompNet dataset the review of all three documents is highly recommended. They are available on CompNet's webpage: www.comp-net.org.

The user guide is structured as follows: Chapter 1 gives a short summary of the history of CompNet and the CompNet dataset. Chapter 2 introduces the 6th Vintage of the CompNet dataset, including the structure and the range of indicators available. Chapter 3 highlights the caveats and possible limitation of the dataset. Chapter 4 illustrates how to use the dataset by analysing a joint distribution. The user guide includes an extensive appendix with information on data providers and data sources, constructed individual indicators, sector coverage and an overview of all available indicators.

1.1 The Competitiveness Research Network

The Competitiveness Research Network (CompNet) provides a forum for high level research and policy analyses in the area of competitiveness and productivity. One of the main activities of the network is the regular release of a firm-level based dataset which covers, in its last vintage, 18 European countries with a strong focus on providing cross-country comparable information to enable frontier research and policy work.

The network was founded by the European System of Central Banks (ESCB) in 2012 with the objective of analysing competitiveness from a more comprehensive perspective encompassing a micro, macro and cross-border dimension. The ultimate goal was and remains to be identifying a robust, theoretical and empirical link between drivers of competitiveness and macroeconomic performance.

Today, CompNet is a standalone network funded by the following institutions (in alphabetical order):

- European Bank of Reconstruction and Development (EBRD);
- European Central Bank (ECB);
- European Commission (EC);
- European Investment Bank (EIB);
- Halle Institute for Economic Research (IWH);
- Tinbergen Institute (TI).

CompNet is the coordinator of a network which brings together data providers and data users.[1] The interplay between actors enables CompNet to: (1) include in the dataset indicators demanded by policy makers and researchers; (2) address the two most common problems regarding cross-country firm-level data, namely, lack of comparability across countries and confidentiality concerns. Comparability is achieved by actively working together with data providers to improve ex-ante sample and variable harmonization, as well as by applying ex-post state-of-the-art techniques to improve sample representativeness.[2] Confidentiality is addressed by following the so called "distributed micro-data approach" (Bartelsman et al., 2004). In this approach a common code is used to extract relevant information from existing firm-level datasets available within each National Central Bank (NCB) or National Statistical Institute (NSI). The protocol computes indicators at the firm-level and then collapses the information to a given level of aggregation. For each indicator, CompNet collects not only the mean but also the complete set of statistical moments of the distribution of firms operating at the given level of aggregation. Furthermore, CompNet provides the so-called "joint distributions" which allow assessing variables conditional on the distribution of another variable. That is, the joint distributions are an instrument to analyse the different characteristics of firms at both extremes of the labour productivity or size distribution in a given sector, or of exporting and non-exporting firms in the same sector or of credit constrained and unconstrained firms to give but a few examples of the immense possibilities.[3]

[1] See Table 14 in the appendix for an overview of data providers and their respective institutions.
[2] See for instance the CompNet Comparability Report, (Altomonte et al., forthcoming).
[3] Please refer to Section 2.2.2 for more information and examples of the joint distributions.

This approach allows keeping most of the richness of firm-level data while preserving confidentiality. The granularity of the resulting dataset helps researches to better interpret variations in economic competitiveness across industries, countries and time periods as well as to identify the impact of shocks. The variety of computed indicators, large granularity and harmonization efforts set the dataset apart from other micro-based cross-country datasets.[4]

Micro-based data provides crucial information for understanding the drivers of competitiveness: aggregate indicators alone, when interpreted as if they had been generated by the behaviour of a representative firm, may often become misleading. These interpretation problems arise because a persistent degree of firm level heterogeneity is present even in narrowly defined industries and size classes. For this reason, relying on micro-data for economic analysis and research guarantees a number of advantages[5]:

- Disentangling the heterogeneous responses of firms.
- Assessing the impact of macro shocks or policy impacts conditional on the underlying micro structure.

In fact, widespread heterogeneity at the firm level opens up the possibility that aggregate performance depends jointly on firm-level decisions (on factor inputs, innovation and technological capacity or export strategy) as well as on market environment (macro wage and price dynamics, structural economic conditions and strategic interactions) [6]. Thus, cross-country information on the underlying distribution of firms is essential in order to assess drivers of aggregate productivity, export performance, and competitiveness.

1.2 References

Altomonte, Carlo, Eric Bartelsman, Jan-Paul van de Kerke, Paloma López-Garcia, Filippo di Mauro, Marc Melitz, Michael Polder and Sebastien Roux. "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability." *The Competitiveness Research Network* (Forthcoming).

[4] Examples of these are the AMADEUS dataset from Bureau van Dijk and the COMPUSTAT dataset from Standard & Poor's.
[5] For a detailed overview see e.g. López-Garcia et al. (2015).
[6] See e.g. Caves (1998) or Bartelsman and Doms (2000).

Bartelsman, Eric J., and Mark Doms. "Understanding productivity: Lessons from longitudinal microdata." *Journal of Economic literature* 38.3 (2000): 569-594.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. "Microeconomic evidence of creative destruction in industrial and developing countries." (2004). *IZA Discussion Paper No. 1374; Tinbergen Institute Discussion Papers No. TI 2004-114/3; World Bank Policy Research Paper No. 3464.* Available at SSRN: https://ssrn.com/abstract=612230.

Caves, Richard E. "Industrial organization and new findings on the turnover and mobility of firms." *Journal of economic literature*36.4 (1998): 1947-1982.

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

López-Garcia, Paloma, and Filippo di Mauro. "Assessing European competitiveness: the new CompNet microbased database." (2015). *ECB Working Paper Series No 1764.*

2. The 6th Vintage CompNet Dataset

The CompNet firm-level-based dataset offers a wide range of indicators constructed on firm-level information. This chapter introduces the structure of the 6th vintage of the CompNet dataset and gives a description of the datasets within CompNet.[7]

2.1 Sample, Time Span & Levels of Aggregation

The 6th Vintage of CompNet dataset represents an unbalanced panel dataset which covers 18 European countries. This provides researchers with a dataset for cross-country studies that includes a rich set of indicators from five different fields: productivity, finance, labour, competition and trade. For more information on the data collecting process refer to Section 5.1 in the appendix or, for a detailed description, to the cross-comparability report (Altomonte et al., forthcoming).

CompNet variables and indicators are available for two samples: "full" and "20E". The full sample intended to cover the period 1999-2016 for most of the countries in the sample. However, actual data availability reduces this time span to 2003-2015 for the majority of the participating countries. In some countries firms are legally obliged to report their balance sheet data only when certain thresholds are met. For example, in Poland only firms with more than 10 employees have to report their accountings. For this reason, in order to provide a more homogeneous sample across countries, CompNet constructed also the 20E sample, including only firms that have at least 20 employees for the same time span.

Even if coverage is not complete, what matters is that available firms are representative of the underlying population. In this respect, CompNet applies a reweighting scheme based on inverse probability weights to the raw data so that the share of firms by macro-sector and size class is the same as in the population. This year, for the first time, reweighting is dependent on the specific indicator and it therefore provides more accurate results. The weighting procedure is applied to both the full sample and the 20E sample.[8] Table 1 shows the countries actual sample sizes and time spans available within CompNet dataset.

[7] A more detailed overview with all the indicators CompNet offers can be found in Table 16 of the appendix.
[8] More details about the exact reweighting procedure can also be found in Annex 2 of the cross-country report (López-Garcia et al., 2018).

Table 1: Countries, Samples and Time Span

Country	Full Sample	20E	Time Span
Belgium	**X**	**X**	2004-2015
Croatia	**X**	**X**	2002-2016
Czech Republic		**X**	2003-2015
Denmark	**X**	**X**	2000-2015
Finland	**X**	**X**	1999-2015
France	**X**	**X**	2004-2014
Germany		**X**[1)]	1999-2014
Hungary	**X**	**X**	1999-2015
Italy	**X**	**X**	2001-2014
Lithuania	**X**	**X**	2000-2015
Netherlands	**X**	**X**	2000-2014
Poland		**X**	2005-2015
Portugal	**X**	**X**	2006-2015
Romania	**X**	**X**	2005-2015
Slovakia		**X**	2000-2015
Slovenia	**X**	**X**	2005-2016
Spain	**X**	**X**	2009-2015
Sweden	**X**	**X**	2003-2015

Notes: 1) only manufacturing sector

Target Population:

CompNet covers non-financial corporations with at least 1 employee operating in the following NACE Rev 2 chapters: B-J and L to N and division 95.[9] This definition is consistent with the one of category S.11[10] in the European System of Accounts (that is, excluding sole proprietors). However, there are some exceptions across countries regarding the selected

[9] Please note that CompNet *input* includes all industries of the non-financial corporations (NFC) sector. The CompNet output however slightly deviates from this classification, excluding NACE rev.2 chapters B, D, E and division 95 and excluding sole proprietors. Please refer to Table 15 in the appendix for a detailed overview of the NFC sectors included in CompNet.

[10] Definition: the non-financial corporations' sector (S.11) consists of institutional units which are independent legal entities and market producers, and whose principal activity is the production of goods and non-financial services. The non-financial corporations sector also includes non-financial quasi-corporations.

target population.[11] Detailed information on sectors included in CompNet is provided in Table 15 of the appendix.

Levels of Aggregation (or Dimensions):

Indicators available in the CompNet dataset are aggregated to different levels, e.g. according to different sector definitions or firm size. The available levels of aggregation are: country level, macro-sector level (corresponding to NACE Rev. 2 sections), sector level (corresponding to 2-digit NACE Rev. 2 sectors), size-classes. For a detailed definition of the macro-sector and sector level aggregation, please refer to Table 15 in the appendix. The size class definitions follow the Eurostat classification system: firms with 1 to 9 employees (micro-firms) are in size class 1. Firms with 10 to 19 employees are in size class 2 and firms with 20 to 49 employees are in size class 3. In size class 4 researchers find all the firms with 50 to 250 employees and size class 5 indicates firms with more than 250 employees.

In a pilot way, the 6th vintage offers indicators aggregated to the regional level, defined at the NUTS2 level. This information is available for up to 10 of the 18 countries included in the dataset. The use of this new level of aggregation is experimental in that we are still not sure of whether the postal code indicated by firms in their balance sheets is a good proxy for the firm location, or rather refers to the headquarters. Moreover, the extent to which this is valuable information might differ across countries. Despite all these uncertainties, we have decided to make this information available. The purpose is to gather feedback from data users to either improve the collected information in future vintages or drop it all together.

[11] Germany: only manufacturing; Finland: real estate activities excluded; Portugal and Romania: manufacture of tobacco products, manufacture of coke and refined petroleum products, postal and courier activities and real estate activities excluded.

2.2 Structure of the Dataset

Figure 1: Structure of the Dataset

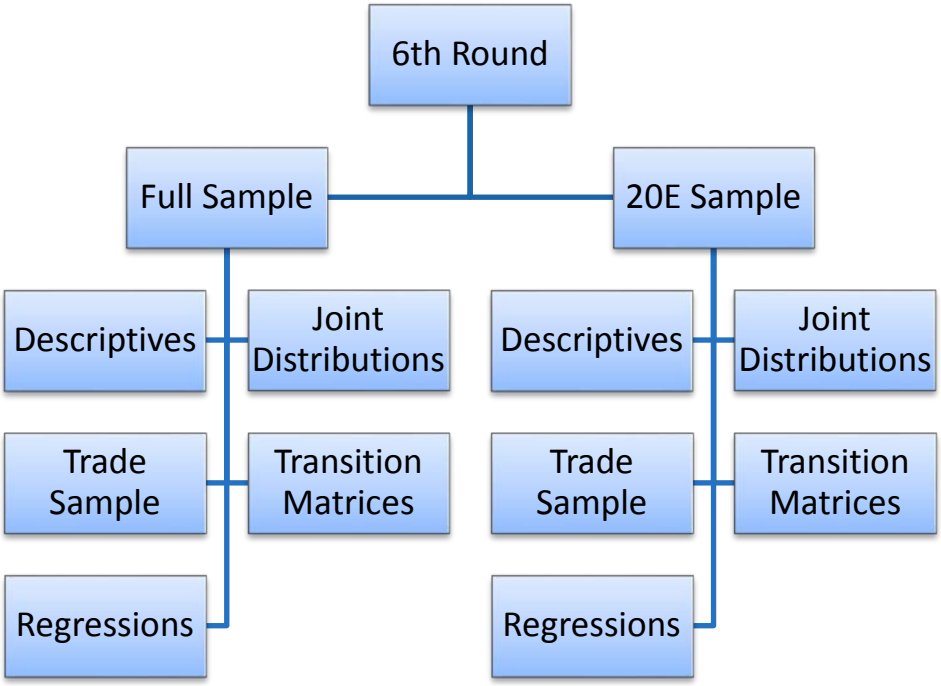


Figure 1 shows the folder structure of the dataset. Each folder contains different datasets or regression tables. In the following, the logic of the naming convention as well as the peculiarities of the different datasets will be described. At the end of the chapter the regression outputs and the structure of the variables will be presented.

Naming of the Datasets:

Within the different folders, the data files are only available in Stata format (.dta)[12] and named according to the following basic structure:

Content_dimension_sample_countries.dta

The three different labels represent different information regarding the data:

1. ***Content***: states the area of study covered by the dataset
2. ***Dimension***: specifies the level of aggregation of the dataset

[12] However, many statistical software packages are capable of importing and converting Stata files, e.g. in R by the R-package readstata13 or via RStudio.

3. *Sample*: indicates the sample on which the dataset is built:
    a. "full" includes all firms with at least 1 employee
    b. "20E" includes all firms with at least 20 employees

"countries" indicates that all countries are appended. This last word was introduced to distinguish the country-specific datasets we receive from the appended dataset, with all countries appended, we distribute to data users.

An Example:

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

unconditional_country_full_countries.dta

Accordingly, the joint distribution of the investment ratio at the macro-sector for the 20E sample will be in:

jd_invest_ratio_mac_sector_20e_countries.dta

2.2.1 Descriptives

The descriptive section provides the user with a general overview of all indicators available in CompNet. Within the descriptives there are four different types of datasets, i.e. unconditional files, productivity decompositions, weighted labour shares, and misallocation measurements.

1. ***Unconditional Files:***

The unconditional files contain the unconditional distributions of all productivity, finance, labour and competition indicators available in CompNet's dataset and are traceable under the prefix *unconditional_*. They are available for the country, macro-sector, sector and size class dimension both for the full and the 20E sample.

2. ***Productivity Decomposition:***

The productivity decomposition dataset contains the estimates of different allocative and dynamic efficiency measures. The content part in the dataset's name is *prod_decomp_*. Three different levels of aggregation are available, which are the country, the macro-sector,

and the sector level. Among the indicators in this dataset, the researcher can find the decomposition of sector productivity in the unweighted mean and covariance term, also called the "OP gap" which captures the allocative efficiency as well as the decomposition of sector productivity growth in within and between components proposed by Foster et al. (2006) among others. Further information on the computation of these indicators can be found in Section 5.3.3 of the appendix. The indicators are named with either *_op* or *_foster* in the dataset's name. All datasets are available for both the full and 20E sample.

3. ***Labour Market Imperfections:***

The labour share weighted Dobbelaere-Mairesse (2013) indicator for labour market imperfections is based on macro-sector or sector level estimations. The corresponding data files are indicated through the prefix *weighted_dm_imp_* and are available for both the 20E and the full sample. Furthermore, the indicator is calculated based on either a Cobb-Douglas production function (*cd*) or a translog production function (*tl*). See further information in Section 5.3.6 of the appendix.

4. ***Misallocation:***

The misallocation data files are indicated through the prefix *misallocation_* and are estimated at the macro-sector and the sector level for both the 20E and the full sample. They contain the within-sector dispersion of 46 productivity and competition variables like labour productivity, capital productivity or De Loecker and Warzynski markups. The dispersion is calculated in the spirit of Kehrig (2011) to make dispersion within industries comparable among each other and over time. Differently from Kehrig (2011), who is using the mean of the variance, CompNet uses the median of the standard deviation in order to have more outlier robust results. The misallocation data file includes as well other allocative efficiency indicators such as the newly introduced (in the CompNet dataset) Petrin-Sivadasan indicator. More information can be found in Section 5.3.3 of the appendix.

Limitations of the Descriptives:

In some cases, misallocation measures and labour market imperfection indicators were constructed (Table 2), because the underlying production functions could not be estimated correctly due to data availability.

Table 2: Limitations of the Descriptives

Country:	Limitation
Belgium	no *weighted_dm_imp_sectl_20e*
Croatia	no *misallocation_macsector_full* no *misallocation_sector_full*
France	no *weighted_dm_imp_sectl_20e*
Lithuania	no *weighted_dm_imp_sectl_20e*
Netherlands	no *prod_decomp_sectorfoster_full* no *prod_decomp_sectorop_full* no *weighted_dm_imp_sectl_20e*
Portugal	no *weighted_dm_imp_sectl_20e*
Slovenia	no *misallocation_macsector_full* no *misallocation_sector_full*
Spain	no *weighted_dm_imp_sectl_20e*

2.2.2. Joint Distributions

Joint distributions are conditional distributions of CompNet's indicators.[13] There are about 20 different files, each containing a joint distribution based on the split of the sample according to a given indicators. The files include the distribution of all indicators for firms in each of the splits. The variable used as a basis to split the sample can be both discrete and continuous. An example of a discrete variable would be a dummy taking the value one for distressed firms and zero for non-distressed firms, in the same sector. The data file would include then all distributions of the rest of variables conditional on the firm being distressed, or not. If the split is done on the basis of a continuous variable, for example firm size measured by employees, we would split the sample in ten deciles and compute the distribution of all remaining variables for firms in each of the deciles of the distribution of firm size.

[13] Please refer to Section 3.3 of the Cross-Country Report for detailed information and more examples on joint distributions.

The content part of the dataset's name indicates which variable is used to split the sample. Furthermore, the datasets are available at the country, macro-sector, sector and macro-sector size class dimensions, as well as for both the full and 20E sample.

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

Example: *jd_l_country_20e_countries.dta*

The dataset provides all indicators conditional on the continuous variable *l* (number of employees of the firms), which is indicated in the content part of the data file *"jd_l"*. Furthermore, this dataset contains indicators for firms with 20 or more employees (*20e*) operating in all sectors of the economy, i.e. at the country level (*country*).

Figure 2: Labour Productivity Distribution by Size Centile in Poland

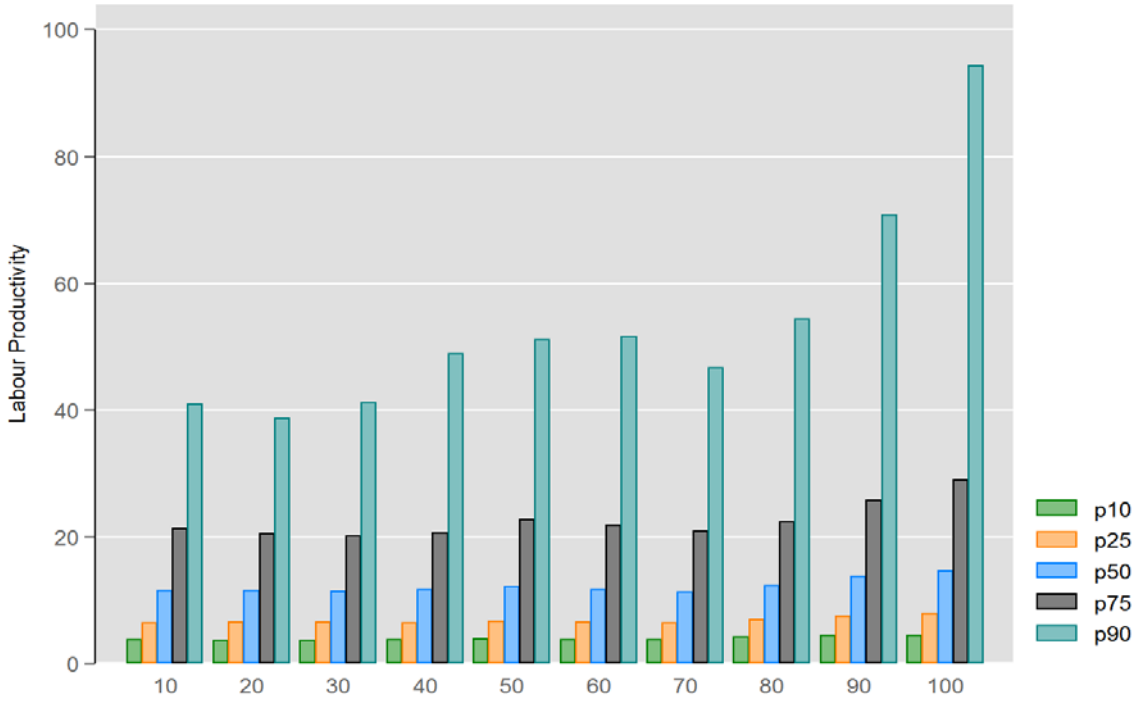


Figure 2 shows an example of a (continuous) joint distribution. In particular, it shows the labour productivity distribution of firms in different deciles of the firm size distribution in Poland in 2015. It becomes clear that there is a discontinuity in the productivity of firms by size. Up to the p70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

Table 3 gives an overview of the content and the conditional variable with description of the joint distributions

Table 3: Description of the Joint Distributions

Content	Conditional Variable	Description
jd_d_zombie_intcov	*D_Zombie_intcov*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: Interest > operating profits for 3 years (conditional on positive profits).
jd_d_zombie_mu_maccd	*D_Zombie_MU_macCD*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: firms with markup < 1 (markup estimate based on macro-sector Cobb-Douglas coefficient)
jd_d_zombie_mu_mactl	*D_Zombie_MU_macTL*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: firms with markup < 1 (markup estimate based on macro-sector translog coefficient).
jd_d_zombie_mu_seccd	*D_Zombie_MU_secCD*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: Firms with markup < 1 (markup estimate based on sector Cobb-Douglas coefficient).
jd_d_zombie_mu_sectl	*D_Zombie_MU_seccTL*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: Firms with markup < 1 (markup estimate based on sector translog coefficient).
jd_d_zombie_negprof	*D_Zombie_negprof*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: Firms with negative operating profits for three consecutive years.
jd_d_zombie_nothg	*D_Zombie_nothg*	Dummy variable which takes the value 1 if firms are zombie firms after the following definition: Firms with negative operating profits for 3 years and not a high-growth firms. Note: No high growth firms are defined as firms with a growth rate <= 3.03%
jd_dummy_exp	*Dummy_export*	Dummy variable which indicates if firms have positive exports.
jd_t10_exp_country	*t10_exp_country_jd*	Dummy variable which indicates if firms belong to the top 10 exporter in the respective country regarding export turnover.
jd_invest_ratio	*ct_invest_ratio*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of the invest ratio of the firms. The invest ratio is defined as follows: (Growth rate of capital + depreciation) divided by capital.
jd_l	*ct_l*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of number of employees.
jd_ln_prod	*ct_lnlprod*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of the labour productivity.
jd_lnsr	*ct_lnSR*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of Solow residuals (weights 1/3 and 2/3).
jd_lntfp_rev_maccd	*ct_lntfp_rev_macCD*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of total factor productivity with macro-sector revenue Cobb-Douglas coefficient.
jd_lntfp_rev_mactl	*ct_lntfp_rev_macTL*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of total factor productivity with macro-sector revenue translog coefficient.
jd_lntfp_rev_seccd	*ct_lntfp_rev_secCD*	Variable which indicates the centiles (10^{th}, 20^{th}, …, 100^{th}) of total factor productivity with sector revenue Cobb-Douglas coefficient.

Content	Conditional Variable	Description
jd_lntfp_rev_sectl	*ct_lntfp_rev_secTL*	Variable which indicates the centiles (10th, 20th, …, 100th) of total factor productivity with sector revenue translog coefficient.
jd_lntfp_va_maccd	*ct_lntfp_va_macCD*	Variable which indicates the centiles (10th, 20th, …, 100th) of total factor productivity with macro-sector value added Cobb Douglas coefficient.
jd_lntfp_va_seccd	*ct_lntfp_va_secCD*	Variable which indicates the centiles (10th, 20th, …, 100th) of total factor productivity with sector value added Cobb Douglas coefficient.
jd_safe	*SAFE*	Dummy variable which indicates the share of credit constrained firms according to estimated elasticities from SAFE. Note: For details on the SAFE estimation refer to Section 5.3.5 in the appendix.

Limitations of the Joint Distributions:

Due to limitations of the raw dataset or estimation problems of the underlying production function (total factor productivity measures), a few JD-files are missing for some countries:

Table 4: Country Specific Limitations of the Joint Distributions

Country:	Limitation
Czech Republic	no *jd_lntfp_va_maccd_macsec_szcl_20e*
Finland	no *jd_lntfp_va_maccd_macsec_szcl_20e*
Hungary	no *jd_d_zombie_intcov_…* (both samples and all dimensions) *no jd_d_zombie_negprof_…* (both samples and all dimensions) no jd_*d_zombie_nothg_…* (both samples and all dimensions) no *jd_l_macsec_szcl_20e* no *jd_safe_…_20e* (all dimensions)
Poland	no *jd_lntfp_rev_maccd_…full* (all dimensions) no *jd_lntfp_va_maccd_…full* (all dimensions) no *jd_lntfp_va_maccd_sector_20e*
Slovakia	no *jd_lntfp_va_maccd_sector_20e*
Spain	no *jd_lntfp_va_maccd_macsec_szcl_20e*

2.2.3 Transition Matrices

Transition matrices aim to show the share and characteristics of firms transitioning from a certain size quintile to another one, within the same level of aggregation, in a 3-year window. The data file follows a similar structure as the joint distributions given that the variable used to split the sample is the indicator describing the transition of the firm along the size distribution.

Indeed, the variable *TRmat_l_country* is the name of the split variable. For example, for all firms that moved from the second labour quantile to the fifth are the variable takes the value "*from_q2_to_5*". Thereby researchers have the opportunity to investigate the share and characteristics of firms which had made different transitions regarding their number of employees.

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

Here the datasets are indicated with the prefix *transition_matrix* and the available dimension are country, macro-sector and sector.

2.2.4 Regressions

In addition to the above introduced datasets, CompNet's dataset provides the user with a series of firm level regression output, available as excel and text files. For each country there are regression outputs where the user can find the results of the export probabilities[14], export premia[15], and production function estimations. We discuss each of them in the following subsections.

Export-Probability:

The export related regression outputs give the user the results of a probit estimation of the probability to export on the productivity and size of firms. There is a separate regression for every independent variable and for every country. The independent variables are deciles of the different productivity and size variables. All regressions control for year dummies. Formally, the coefficients of the following model are estimated by maximum likelihood:

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

with **ct_x** as the productivity centile dummies and **szclass** as the macro-sector size class dummies and $\Phi()$ as the standard normal cumulative distribution function. The observations are weighted by their inverse sampling probability defined as the theoretical number of firms of a certain macro-sector size class divided by the actual number of sample firms in that macro-sector size class.

[14] Only available for the trade module; see Section 2.2.5.
[15] Only available for the trade module; see Section 2.2.5.

Table 5: Probability to Export (Weighted Probit Regression)

Dependent Variable	Independent Variables
Dummy_exp	*ct_lnkprod*
	ct_lnlprod
	ct_lnlprod_rev
	ct_lnSR
	ct_lntfp_rev_macCD
	ct_lntfp_rev_secCD
	ct_lntfp_rev_macTL
	ct_lntfp_rev_secTL
	ct_lntfp_va_macCD
	ct_lntfp_va_secCD
	ct_ lntfp_rev_adj_macCD
	ct_ lntfp_rev_adj_secCD
	size class
	year

Export Premium:

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

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

with **prod** as the vector of the chosen productivity measure (in logs), **X** as the matrix of various export dummies, **SZ** containing the size class dummies, the two-digit sector dummies **SEC, CR** for the crisis dummies and **IT** containing the interaction terms of **CR** and the export dummy Dummy_exp, and finally the error term **e**. The number of independent variables used depends on the model specification.

Table 6 shows the dependent and independent variables:

Table 6: Export Premium (Weighted OLS-Regressions)

Dependent Variables	Independent Variables
lnkprod	*Dummy_exp_3y*
lnlprod	*Dummy_exp_new*
lnlprod_rev	*Dummy_stop_exp*
lnSR	*Dummy_exp_switch*
lntfp_rev_macCD	*Dummy_exp*
lntfp_rev_secCD	*sector*
lntfp_rev_macTL	*size class*
lntfp_rev_secTL	*Dummy_financial_crisis*
lntfp_va_macCD	
lntfp_va_secCD	

Production Functions:

Production functions were estimated by pooling all firms operating in a given macro-sector or sector level, and by assuming either a Cobb Douglas or translog production function. The output measure of the firm used in the regression was either real value added or real turnover. See Section 5.3.2 in the appendix for more details in the estimation procedure.

Table 7: Production Function (Weighted Two-Step Instrumental Variable Regression)

Dependent Variables	Independent Variables	Production function
Real revenue (rturnover) **Real value added (rva)**	*lnl* *lnm* *lnrk* *ln_K_l1* *ln_M_l1* *k2_l1* *m2_l1* *k3_l1* *m3_l1* *km_l1* *k2m_l1* *km2_l1*	Translog Cobb Douglas

Limitations of the Regressions:

Table 8: Country Specific Limitations of Regression Availability

Country	Limitation
Belgium*	no export premium/deciles regression
Portugal*	no export premium/deciles regression
Spain*	no export premium/deciles regression

Note: *based on BACH data which does not include export information.

2.2.5 The Trade Module

Export activity and its two-way link with firm productivity is at the core of competitiveness analysis. For this reason moments of the distribution of a set of variables related to export activities have been computed at various levels of aggregation (country, macro-sector, sector and size class). Moreover, to better understand what makes a firm an exporter, the dataset contains joint distributions in which the split variable is a dummy taking the value one if the firm is an exporter (also considering different types of exporters – see below) and 0 otherwise. Once again, the joint distributions provide the distribution of all indicators computed by CompNet for both exporting and non-exporting firms in the same sector.

Only firms with an export value of at least 0.5% of their total turnover are considered as exporters. Furthermore, several types of exporters have been defined: Exporters in the current year, new exporters, switching firms (firms that are exporters in year t, but not in t-1 and t+1), firms that had stopped exporting and two-way traders (firms that are exporters as well as importers). Some of these export dummies are available for two and three

consecutive years. See table 10 for some variable definitions and table 16 in the appendix for the complete list and definitions of trade-related indicators.

Please keep in mind that all trade indicators, are available only for the manufacturing sector and that they are also available only for a smaller range of countries than for the "Full" and "20E" modules.[16]

To ensure comparability and usability, these indicators can be found in the separate trade module of the CompNet dataset.[17] The trade indicator files were produced by the same Stata code as the "normal" output files, although with some notable differences: 1) All observations outside the manufacturing sector (NACE 2 rev. 10-33) are dropped from the dataset and 2) several additional trade indicators are calculated and 3) the number of joint distribution files has been reduced.

The trade datasets follow a similar structure as the regular datasets, with a descriptive part and joint distribution, transition matrices and regressions. However, the range of variables and data files is limited to those files and indicators that are most relevant for trade analysis. E.g., the descriptive subfolder contains only the unconditional files, the number of joint distribution files is limited and we have included only the export regressions in the regression folder.

Descriptives

The descriptive section of the trade module provides the user with the unconditional distributions of all indicators available in CompNet. In addition to the financial, productivity, labour and competition indicators, a wide range of trade indicators can be found in the unconditional datasets for the dimensions country, sector and size class. Please see table 16 in the appendix for the complete list of trade-related indicators. Furthermore, you can find additional information about the trade module in Dhyne et al. (2015). Data are available for both the 20E and the full sample for most countries (exceptions see Table 5). As described in Section 2.2.1 the prefix *unconditional_* indicates the unconditional files of the trade module.

[16] Full sample is available for: Croatia, Finland, Italy, Lithuania, Slovenia, and Sweden. 20E sample is available for Croatia, Czech Republic, Finland, France, Germany, Hungary, Italy, Lithuania, Poland, Romania, Slovakia, Slovenia, and Sweden.

[17] See Figure 1 for an overview.

Limitations of the Unconditional Files:

Table 9: Country Specific Limitations of the Unconditional Files (Trade Module Only)

Country:	Limitation
Belgium*	no full sample no 20E sample
France	no full sample
Germany	no full sample
Hungary	no full sample
Netherlands	no full sample no 20E sample
Portugal*	no full sample no 20E sample
Romania	no full sample
Slovakia	no full sample
Spain*	no full sample no 20E sample

Note: *based on BACH data which does not include export information.

Joint Distributions[18]

In the joint distributions produced by the trade module the user finds distributions conditional on the firm being an exporter or other export-related variables, described in table 10, analogously to Section 2.2.2, the content part of the dataset's name indicate which variable is used as split variable. The column variables of the joint distributions include the complete set of indicators. Furthermore, the datasets are available at the country, macro-sector, sector and macro-sector size class dimensions, as well as for both the full and 20E sample.

[18] For the description of the joint distributions (for example: *jd_dummy_exp* or *jd_lnsr)*, see Section 2.2.2.

Table 10: Description of the Joint Distributions (Trade Module)

Content	Conditional Variable	Description
jd_dummy_exp_new	*Dummy_exp_new*	Dummy variable which takes the value 1 if firms have positive exports in the observed year as well as the next one but not in the previous year
jd_dummy_exp_new2	*Dummy_exp_new2*	Dummy variable which takes the value 1 if firms have positive exports in the observed year but not in the previous year.
jd_dummy_exp_switch	*Dummy_exp_switch*	Dummy variable which takes the value 1 if firms have positive exports in the observed year but not in the previous and next year.
jd_dummy_stop_exp	*Dummy_stop_exp*	Dummy variable which takes the value 1 if firms have positive exports in the observed year as well as in the previous one but not in the next year.
jd_dummy_stop_exp2	*Dummy_stop_exp2*	Dummy variable which takes the value 1 if firms have positive exports in the observed year but not in the next year.
jd_dummy_trader_2way	*Dummy_trader_2way*	Dummy variable which takes the value 1 if firm is considered both an exporter and an importer in the observed year.
jd_lnlprod_rev	*ct_lnlprod_rev*	Variable which indicates the centiles (10^{th}, 20^{th},… , 100^{th}) of number of the labour productivity (Turnover based).

Limitations of the Joint Distributions:

Table 11: Country specific limitations of the Joint Distributions (Trade Module)

Country:	Limitation
Croatia	no *jd_trader_2way_..._20e* (all dimensions)
France	no *jd_trader_2way_..._20e* (all dimensions)
Poland	no *jd_trader_2way_..._full* (all dimensions)
Slovenia	no *jd_trader_2way_..._full* (all dimensions) no *jd_trader_2way_..._20e* (all dimensions)

Regressions

See Section 2.2.4 for details on the regressions. However, estimations of production functions are not included in the trade module because they would be identical with the 20E or full sample regressions for the manufacturing business.

Transition Matrices

Analog to the transition matrices explained in Section 2.2.3. However, for the trade module the transition matrices contain additional trade-related variables and are only referring to the manufacturing business.

2.3 Statistics: Name Variable Convention

Most variables delivered in the CompNet dataset follow the same convention. By adding the following prefixes/suffix to the variable names, the user will be able to find for example the growth rate, all percentiles, and the total number of firms upon which the indicators have been computed in a given year, country or sector. Dummy variables are an obvious exception from this structure.

Table 12: Variable Structure

Prefix/Suffix	Meaning
_tot_mark	Number of firms in the sample with available information to compute the relevant indicator (i.e. does not include observations with missing values)
_p1, _p5, _p10, _p25, _p50, _p75, _p90, _p95, _p99	Percentiles of the considered variable
_mean	Mean of the considered variable
_sd	Standard deviation of the considered variable
_skew	Skewness of the considered variables
_kurt	Kurtosis of the considered Variable
_sum_weights	Number of firms in the population

2.4 References

Altomonte, Carlo, Eric Bartelsman, Jan-Paul van de Kerke, Paloma López-Garcia, Filippo di Mauro, Marc Melitz, Michael Polder and Sebastien Roux. "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability." *The Competitiveness Research Network* (Forthcoming).

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

Dhyne, Emmanuel; Demian, Calin-Vlad; Harasztosi, Péter; Meriküll, Jaanika; Lalinsky, Tibor; Cazacu Ana-Maria; Bugamelli, Matteo; Soares, Ana Cristina; Berthou, Antoine; Oropallo, Filippo. "Assessing European firms' exports and productivity distributions: the CompNet trade module." (2015). *ECB Working Paper Series No 1788.*

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

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

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

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

3. Limitations and Caveats of the Dataset

Over the years, progress has been made in making the CompNet dataset more intuitive, practical and easy to use. However, data users must still be careful when using a micro-aggregated dataset that allows for cross-country comparisons. This chapter summarizes the main issues which are addressed in depth by the CompNet Comparability Report (Altomonte et al., forthcoming).

3.1 Cross-Comparability in CompNet 6th Vintage

The data collection process discussed in Section 5.1 of this user guide has three main advantages: (i) the dataset uses existing national datasets, with no need to undertake new and costly data collection efforts, (ii) confidentiality of the micro information is fully protected by using the micro-aggregation technique and (iii) member institutions participate actively in improving and using the dataset. By using existing national data sources, the data collection process is less cost intensive but has a considerable downside: there is no control over 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 introduce threats to cross-country comparability, which we define as incomparable samples or variables due to differences in the underlying data sources. In these cases, it is of paramount importance to document the remaining differences in order to help analysts deciding on their relative importance, as well as suggesting strategies to mitigate the potential biases of own estimations based on CompNet data. For that reason, CompNet has invested a great deal of effort to produce detailed meta-data and to analyse strengths and weaknesses of the data in terms of cross-country comparability. This documenting effort sets our dataset apart from other sources of granular data.

Before we turn to the sources of incomparability, three alleviating remarks have to be made at the outset: i) member institutions and data providers participate actively in improving the

datasets. [19] This means that they are aware of all comparability issues and the skills and experience for solving these problems for future vintages are present within the network. ii) As a result of the micro-aggregation technique, the dataset compares distributions across countries rather than actual values, which are thus less prone to change due to definitional differences. iii) Quality issues have, to a considerable extent, been addressed by CompNet ex-post through the use of weighting. Econometric modelling techniques can also be used by the analyst.

We divide the causes of possible incomparability into country and source specific comparability issues as well as variable and indicator specific incomparabilities. The following sections discuss these causes and provide some apparent examples drawn from the CompNet Comparability Report (Altomonte et al., forthcoming).

3.1.1 Country and Source Specific

The country and source specific causes of incomparability refer to the fundamentals of the different data sampling methodology in each country. Some exemplary questions here are at what level of aggregation the information is captured, what industries are covered, whether firms are representative of the population in terms of macro-sector and size classes and whether there are significant breaks or changes affecting the quality of the underlying source. Note, however, that data providers update the whole time series every time they run the code (therefore, not only one extra year is added) in order to minimize breaks in the dataset resulting from the addition of new indicators, change in sector classification or improvement of the underlying methodology. We discuss these three topics referring to information drawn from the CompNet Comparability Report (Altomonte et al., forthcoming).

Units of Observations

In a dataset containing micro information, firm data can be gathered at different levels of aggregation, the so-called units of observation. Eurostat uses the enterprise level of observation which is defined as "*an organizational unit producing goods or services which have a certain degree of autonomy in decision-making. An enterprise can carry out more than one economic activity and it can be situation at more than one location. An enterprise*

[19] Currently the CompNet code is adjusted based on the recommendation following from the CompNet Comparability Report: "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability".

may consist out of one or more legal units." The enterprise level is used by a selected number of data providers of CompNet as well but the majority use the legal unit, which is a lower level of aggregation (Altomonte et al., forthcoming). The usage of different levels of aggregation matters because different data sources across country will target a different 'slice' of the economy. Consolidation of the balance sheets also plays a role here; unconsolidated information at the enterprise level could inflate economic activity relative to consolidated enterprise information.

Representativeness

On a more fundamental note, it is important to have representative data for all different countries. Enough firms should be covered by the domestic data sources and more importantly, these firms should be representatively distributed across different size classes and macro-sectors. Although the coverage rates differ between countries, the overall assessment of the sample representativeness is positive (Altomonte et al., forthcoming).

Breaks and Changes

The data sources used by CompNet develop over time. Thus, changes may also occur within the time span the indicators have been collected. The Comparability Report features extensive country-specific tables to show exactly which changes to sample composition, taxation systems or variables definitions occurred over time. Changes in methodology or indicator definitions do not cause breaks since, as mentioned before, the whole time-series is recalculated for the same raw data set every time we ask data providers to run the CompNet code.

3.1.2 Variable and Indicator Specific

The variable- and indicator specific sources of incomparability refer to possible differences of raw variable definitions. The common code sent out to data providers calculates the output indicators from the underlying raw variables. Hence, differences between the definitions of the input may cause large differences in the output of the code. We discuss the observed definitions of the employment variable and the valuation of the output and labour cost variables.

Employment

The number of employees is crucial for estimating labour productivity, total factor productivity and numerous other indicators. To be able to compare these indicators, they must be based on the same underlying variables, but the measurement of the number of employees is not as straightforward as one may think. In fact, it can refer to either the headcount or to full time equivalents (fte). Moreover, in terms of inclusion either only employees or the total employed personnel can be taken into consideration. We observe differences in measurement and in inclusion across countries, with headcount and employees only being the most frequent common denominator among the data sources (Altomonte et al., forthcoming).

Valuation of Output Variables and Labour Costs

Data providers sourced information often from financial statement data. In these databases variables can be valuated differently. For instance: are the output variables valued at factor prices or at market prices? Do output variables (measured in monetary units) or labour costs include taxes or subsidies? Zooming in on the labour cost, the inclusion of taxes or subsidies is important, especially in European countries where social security contributions can amount to a large share of the gross wages. When labour costs are valued differently across countries, output measures using this variable might not be fully comparable. The choice for the method of valuation depends on the ability to add or subtract taxes or subsidies, but also on the type of research one intends to do. When analysing profit-maximizing firms for instance, the researcher wants to use the complete labour costs rather than those costs net of taxes. In the CompNet 6th vintage dataset considerable efforts have been made to mitigate the impact of valuation differences even further. If these differences could not be fully accounted for, the analyst should take appropriate actions.

This section discussed a few concerns of cross-country comparability in the CompNet dataset. For in-depth information please refer to the comparability report (López-Garcia et al., 2018). Most of the issues that the user might encounter while using the CompNet data arise from these comparability matters. To be aware of these issues is the first step in order to overcome them and optimize the quality of the analysis carried out on the data. The main asset that has been made available to the researcher for this purpose is the Comparability Tool. This tool enables researchers to compare the metadata underlying any computed

indicator across countries and across sources to learn to what extent they are comparable. Moreover, it becomes possible to distinguish between those differences imputable to real economic, social or institutional variables and those which are instead due to other causes.

Asymmetries between data sources that do not change over time could be controlled for by the inclusion of country dummies in a regression specification or with other indicators related to the type of source, for instance. In this way, the researcher could track or test possible biases or include them in a model specification.

3.2 Comparison with National Accounts

CompNet indicators are aggregated from firm-level sources where the information is based on national taxation legislation, European legislation and accounting principles (e.g. GAAP). The national accounts aggregated data differ significantly from the CompNet indicators, first and foremost because the data stems from a wide variety of sources which also cover other sectors than the CompNet database with its focus on non-financial corporations. These different sources are then consolidated into the national accounts according to the national accounts standards like the European System of Accounts (ESA). Hence, the two datasets will show similar patterns but are vastly different because they measure different slices of economic activity.

3.3 Other General Issues

Sample and Population Figures

One obvious question that occurs to the applied researcher is how to compute aggregate statistics by using the 6th Vintage CompNet dataset. The problem is that the applied weighting procedure gears the descriptive statistics of the CompNet indicators towards describing the total population, and not the underlying sample. The sole deviation from this principle is the “tot_mark” column variable, which shows the number of firms in the cell with available information to construct the indicator. If the researcher wants to retrieve the total employment, value added or any other indicator in a given cell, it is important to use the “sum_weights” variable when referring to the population and the “tot_mark” version when referring to the sample. Therefore, in order to know , e.g., the total employment for the *population* of a given cell of the dataset it will be sufficient to multiply the average

employment “l_mean” by the column variable “l_sum_weights” (i.e. the number of firms in the *population*). This line of reasoning applies to all variables.

If the researcher wants to collapse the dataset to a higher level of aggregation (for example, from the sector to the macro-sector level), one needs to use the “sum_weights” variable to have population representative weights. This is an important difference with respect to previous CompNet vintages which had implemented the reweighting procedure only for the 20E sample, but not for the full sample.

How to Deal with Dummy Variables

The CompNet dataset contains many dummy variables referring to a wide variety of topics, from distressed firms to trade, etc. The CompNet code is built to store in each output file an array of the percentiles and moments of the distribution of each variable. However, this operation does not make much sense for dummies since their distribution is binary. For these variables, the only descriptive statistic in the output files that provides useful information is the mean, i.e. the percentage of observations (firms) for which the variable is equal to 1. For example, the mean of “Dummy_exp” equals the share of exporters in the given cell and will therefore take values between 0 and 1.

3.4 References

Altomonte, Carlo, Eric Bartelsman, Jan-Paul van de Kerke, Paloma López-Garcia, Filippo di Mauro, Marc Melitz, Michael Polder and Sebastien Roux. "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability." *The Competitiveness Research Network* (Forthcoming).

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

López-Garcia, Paloma, and Filippo di Mauro. "Assessing European competitiveness: the new CompNet microbased database." (2015). *ECB Working Paper Series No 1764.*

4. Use Case: The "Cleansing Effect"

This use case aims at showcasing the relevance of using the CompNet dataset for cross-country comparison and serves as a hands-on introduction to the structure of the dataset. In this example, we use the dataset to test whether the financial crisis of 2008-2009 had a "cleansing effect" as Caballero and Hammour (1994) called it in their seminal paper. This means that the share of resources occupied by unproductive firms decreases; either through the reallocation of factors of productions (in our case labour) or through (unproductive) firms completely exiting the market. Consequently, the economy is "cleaned", as the allocation is becoming more efficient, leading to a rise in average productivity.

However, in contrast to Caballero and Hammour, using data for the US, Foster et al. (2016) concluded that the financial crisis did not have the expected "cleansing effect".[20] Therefore, we use the CompNet data to analyse employment flows before and after the financial crisis of 2008-2009 for a set of EU countries.[21] To do so, we look at the joint moment of employment and productivity which enables researchers to analyse the employment changes of firms across productivity deciles. In particular, we analyse the efficiency of labour allocation over time in several EU countries, and especially around the time of the great recession of 2008-2009.

With the CompNet dataset, this cannot only be illustrated on the country level, but also on the macro-sector level. To showcase this, the use case will also present an application of the above described analyses on the construction and manufacturing sector.

4.1 Technical Description

As we intend to use the joint moments of employment and productivity for our analysis the joint distribution files of the CompNet dataset are utilized. To analyse the flow of the labour force across firms which exhibit different levels of productivity, a researcher can choose between the joint distributions of a wide range of productivity indicators. For this use case we choose labour productivity as our productivity measure and consequently need to pick the file "*jd_lnlprod_country_full_countries.dta*" for the analysis at the country level. This data file contains the values of numerous firm characteristics given the position of the firm

[20] See also Bartelsman et al. (2013).
[21] This use case draws heavily from chapter 4 of López-Garcia and di Mauro (2015).

within the productivity distribution.[22] As indicated by the name, the productivity distribution in this file always refers to the country dimension.

The 6th vintage of the CompNet dataset also includes joint distributions for labour productivity for the macro-sector, macro-sector size class and sector level. At a later point in this chapter we will show this for a macro-sector analysis of the manufacturing and construction sector (and consequently we will use the file "*jd_lnlprod_mac_sector_full_countries.dta*").[23]

For this use case we will look at the average number of employees employed by firms in the different percentiles of the productivity distribution and calculate the respective net labour force growth rates within time periods of three to four years.[24] While doing so, it is important to be mindful about the group of interest. In our case we are not interested in the development of employment within our sample, but in the overall population. Consequently, just as suggested in Section 3.3, we calculate the level of employment within the respective labour productivity decile via the mean of the number of employees for firms ("l_mean") and multiply it with the respective summed weights ("l_sum_weights").

4.2 The Country Dimension

To analyse the effect of the great recession we compare the relative change in employment across the different productivity deciles between a "normal" time span (2002-2004/2005-2007) and the developments during the crisis period (2004-2007/2008-2012). The results for two countries, classified as "stressed countries"[25], i.e. countries that have been seriously affected by the financial crisis, are shown in Figure 3.

The left-hand graph shows the change in labour force in "normal times" in the stressed countries, that is the percentage change in employment between the period 2002-2004 and 2005-2007. Here, we observe that employment has been growing through all parts of the economy, with the most significant increases among the most productive firms. During the crisis, however, the data shows that employment has decreased or stagnated more or less for all but the top productive firms where we can still observe a significant increase in

[22] For a detailed description of joint distributions review Section 2.2.2.
[23] For a complete overview of available dimensions review Section 2.1.
[24] For studies depending on labour statistics it is also important to understand how labour was measured. For a complete overview of this issue please review Chapter 3.3.2 of the Comparability Report.
[25] Hungary, Italy.

employment. This result suggests the existence of a cleansing effect of the recession. The most productive firms seemed to be much better able to cope with the stressed economic environment than the less productive firms.

Figure 3: Changes in Employment Split by Productivity, "Normal" Times vs. Crisis Period on the Country Level; Stressed Countries

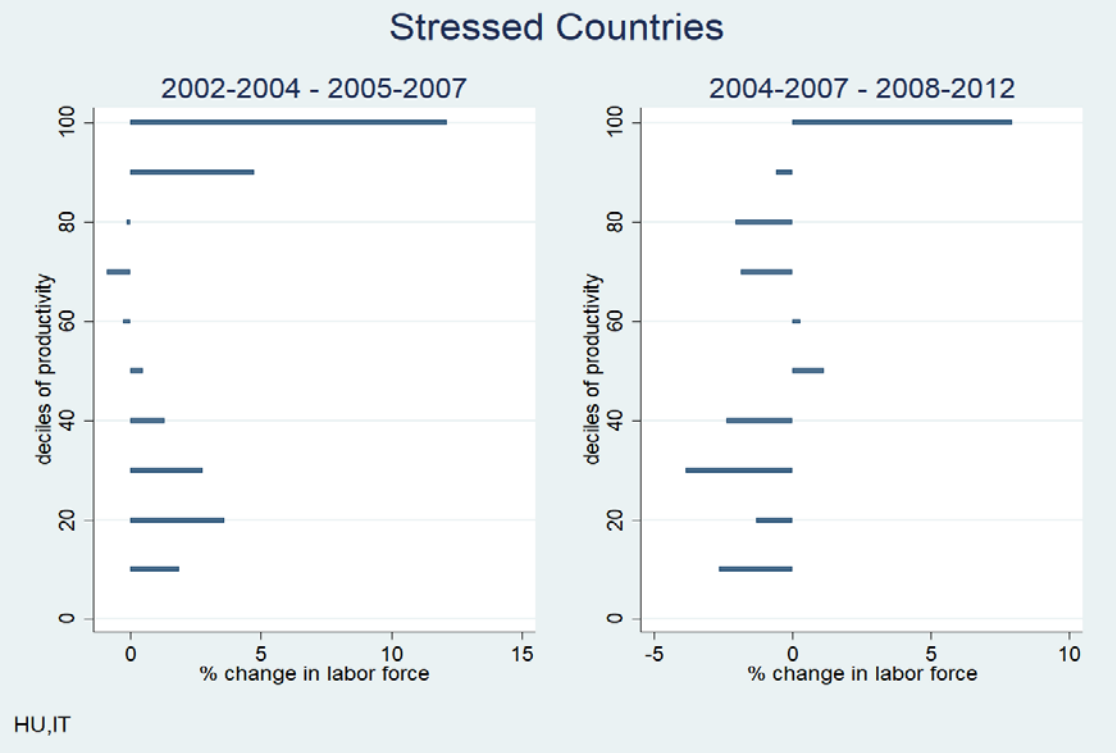
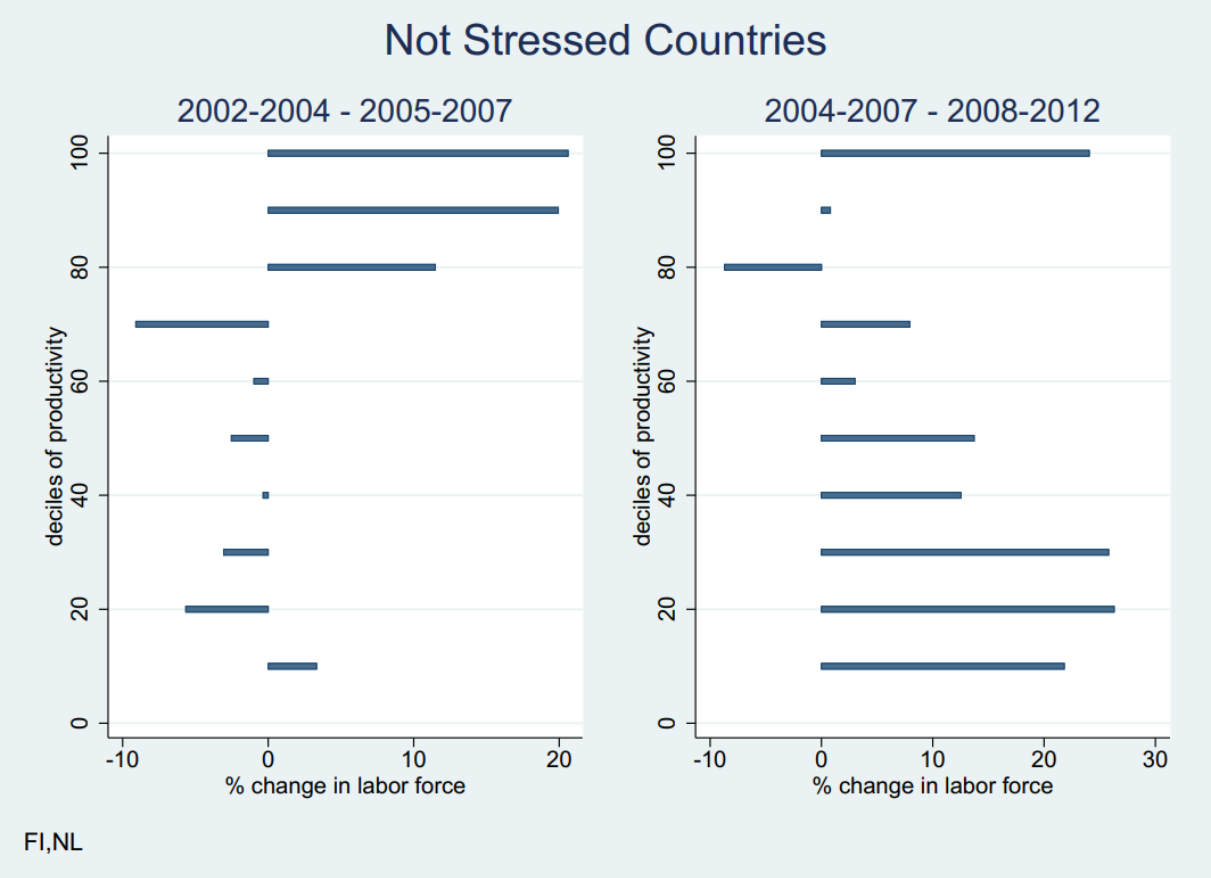


Figure 4 enables us to compare the result for the stressed countries with the development for two countries which were less affected by the financial crisis than the "stressed" countries.[26] Here, we can observe that the difference between the "normal" period and the crisis period is that while the top productive firms (10th centile) have been keeping up their employment growth, the employment in the 8th or 9th centile has decreased or stagnated, whereas employment has increased for the mid- to low productive firms (first to 7th productivity centile). Obviously, there are no indications for a "cleansing effect" and at this stage we do not intend to speculate why the employment reduction has primarily affected the third most productive centile.

[26] Finland, Netherlands.

Figure 4: Change in Employment Split by Productivity, "Normal" Times vs. Crisis Period on the Country Level; Not Stressed Countries



4.3 The Macro-Sector Dimension

Besides the cross-country dimension of the joint distribution indicators, the CompNet dataset also allows for analysing the sector and macro-sector dimensions. In addition to examining the labour reallocation due to a "cleansing effect" on the aggregate level, it is important to understand the potential drivers of such an aggregate pattern. To investigate this question, the joint distribution of labour productivity and employment on macro-sector level of the full sample is used (*jd_lnlprod_mac_sector_full_countries.dta*). As indicated in the beginning, we will repeat this exercise now for the construction (macro-sector number 2) and manufacturing (macro-sector number 1)[27] sectors.

Figure 5 shows the changes in employment for the two periods in the construction sector by deciles of labour productivity. In the pre-crisis period (left panel) almost all firms across all

[27] See for a complete overview of the sector definitions Table 15 in the appendix.

deciles of productivity show an increase in employment. The right panel shows that the crisis period is characterized by a significant negative change for stressed countries as e.g. Italy. Firms in non-stressed countries, however, only increased their employment, for all productivity deciles. In general, the "cleansing effect" can be observed in the stressed countries, where low productive firms significantly reduced their labour force, while the most productive firms either gained or lost significantly less (e.g. Hungary).

Figure 5: Change in Employment of Firms Split by Productivity in Construction Sector, "Normal" Times vs. Crisis Period

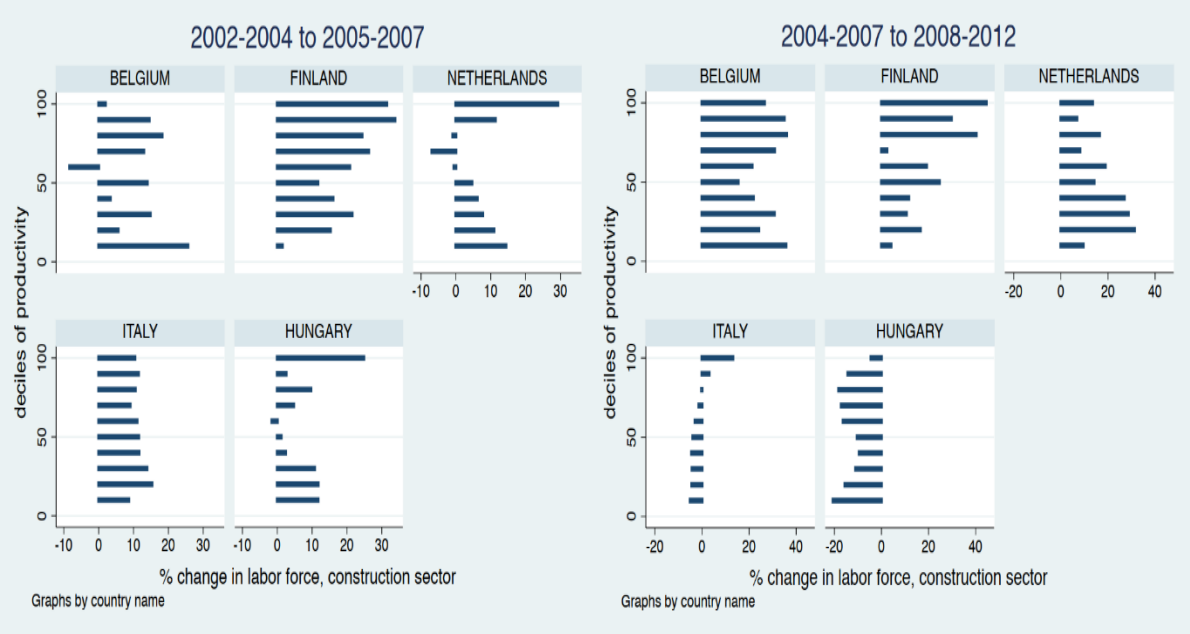
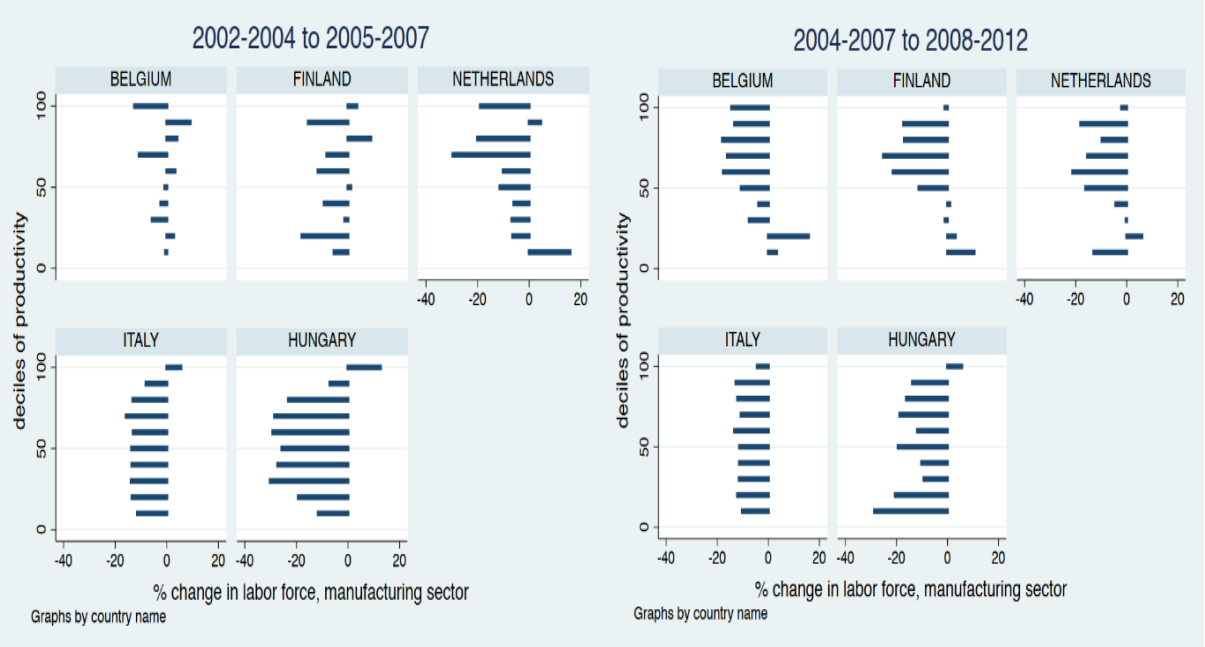


Figure 6 shows the manufacturing sector. In comparison to the construction sector we can observe a different trend in the "normal" period. In the left panel of figure 6 we can observe a decrease in labour employed for most productivity levels. In the crisis period the negative trend is even stronger and only the less productive firms in non-stressed countries show a slight increase in their employment. Overall, we can observe a considerable drop in employment in every country in the manufacturing sector.

Figure 6: Change in Employment of Firms Split by Productivity in Manufacturing Sector, "Normal" Times vs. Crisis Period



This use case highlighted the usefulness of one of the joint distributions of the 6th CompNet dataset to understand the underlying dynamics of labour flows before and after the financial crisis of 2008-2009. The results are mostly in line with the literature and the existence of a cleansing effect in stressed countries during the crisis period was observed.

In contrast to this brief analysis focusing on only one indicator, the interested researcher is referred to the new CompNet Cross-Country Report (López-Garcia et al., 2018) for more comprehensive information about the wide range of indicators and research opportunities of the 6th CompNet database vintage.

4.4 References

Altomonte, Carlo, Eric Bartelsman, Jan-Paul van de Kerke, Paloma López-Garcia, Filippo di Mauro, Marc Melitz, Michael Polder and Sebastien Roux. "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability." *The Competitiveness Research Network* (Forthcoming).

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. "Cross-country differences in productivity: The role of allocation and selection." *American Economic Review* 103.1 (2013): 305-34.

Caballero, Ricardo J., and Mohamad L. Hammour. “The Cleansing Effect of Recessions.” *The American Economic Review*, vol. 84, no. 5, 1994, pp. 1350–1368. *JSTOR*, www.jstor.org/stable/2117776.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. "Reallocation in the Great Recession: cleansing or not?" *Journal of Labor Economics* 34.S1 (2016): S293-S331.

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

López-Garcia, Paloma, and Filippo di Mauro. "Assessing European competitiveness: the new CompNet microbased database." (2015). *ECB Working Paper Series No 1764.*

Lucas Jr, Robert E. "On the size distribution of business firms." *The Bell Journal of Economics* (1978): 508-523.

5. Appendix

5.1 Data Collection

Firm-level data provides crucial information about a wide range of economic phenomena. CompNet works bilaterally with many institutions in several European countries to create its dataset. This allows immediate feedback from and to data providers to solve any problem that may arise quickly and efficiently. Two important concerns regarding firm-level data are confidentiality and the treatment of outliers. The following two subsections elaborate on the way CompNet deals with these concerns.

5.1.1 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 actually handling any confidential micro data at firm level, but only the anonymized and harmonized output delivered by the individual country teams. The code produces descriptive statistics and regression results at different levels of aggregation (while keeping the rich information of the underlying distributions) and ensures that the user of the final data will not be able to uniquely identify individual firms. The result is the micro-aggregated data provided in the CompNet dataset.

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

The CompNet code includes a specific routine, which is ran in the final stage of the computation that checks the eventual output cells. This routine includes thresholds for the minimum number of observations to guarantee that no individual firm can be identified and tests for statistical dominance. If a cell is based on a limited amount of underlying micro-observations, which might make the identification of individual firms possible, the cell will be dropped. However, this dropped information is still accounted for in the total distribution to maintain a high level of representativeness. The second test is the test for statistical

dominance. It includes thresholds for the largest permissible size share a single observation takes on in a given cell.

These thresholds can be set a-priori by the data providers to satisfy their country or institution specific conditions. These are the parameters which have been used by most of the data providers:[28]

1. Overall minimum number of observations for all statistics.
2. The minimum number of observations for all descriptive statistics (all percentiles mean, extrema, and moments) can be fixed ex ante.
3. The minimal number of observations for the 1% and 99% percentiles can be adjusted separately.
4. The minimal number of observations for the 5% and 95% percentiles can also be adjusted separately.
5. 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.

5.1.2 Outlier Dropping

The second routine focusses on miss-measured values and categorizes them as outliers. Previous vintages of data collection have revealed a trade-off. On the one hand, the outlier procedure must not affect or distort aggregate results by limiting the number of observations used for the indicator calculation. On the other hand, it must be strict enough to correctly filter out values that can be identified as outliers.

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

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

[28] Two exceptions are the Netherlands and Finland: Whereas the Dutch data providers did not allow the publication of percentiles for certain “unprocessed” variables (e.g. labour or capital stock) or the respective growth rates derived from these variables, Finland only wanted to include one additional confidentiality threshold for 10%, 25%, 75% and 90% percentile.

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

5.2 How to Apply for the Data

The application process for the CompNet dataset is designed to be as simple as possible. To receive access, you have to fill in the online request form in the data section on the CompNet-homepage[29]. The Executive Committee of CompNet will review the request and will inform you about their decision. The processing time can be reduced if you provide sound information about yourself and your research project. Additional information (such as a CV) can be very helpful. The applicant will normally be informed about the decision within a week. Please note the terms and conditions and other important regulations regarding the usage of the data, which are described in detail on the application page.

In case of acceptance, you will receive an email with the necessary credentials to log in the system and full access to the 4th (up to 2012), 5th (up to 2013) and 6th (up to 2016) data collection rounds for a period of six months. The renewal is possible at the end of this period by sending a request to the secretariat (no application form necessary at this stage). In case anything is unclear all questions can be directed to secretariat@comp-net.org.

5.3 Derivation of Indicators

This section discusses the estimation and theoretical background of a selected number of indicators. Specifically the labour productivity and total factor productivity indicators, the zombie dummy variables, the SAFE-indicator, the Dobbelaere-Mairesse indicator of labour market imperfections, the Petrin-Sivadsan gap indicators, the markup estimation and the job creation and destruction rate indicators.

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

5.3.1 Productivity

Before we proceed to the actual estimation of the various productivity measures included in the dataset, we would like to refer to a particular statistic computed for the productivity indicators: the weighted average in a given level of aggregation.

Weighted averages are named according to the following convention: *aggr_lntfp_rev_secCD* where *aggr* indicates that the average is weighted; *lntfp* refers to the indicator and thus to the concept; *rev* the weighting of the estimates and *secCD* the aggregation level at which the estimation is ran combined with the functional form used. We use different weights, shown in column 2 of Table 13.

Table 13: Weighting of the Productivity Indicators

Indicator	Weights
aggr_lnkprod	Labour
aggr_lnlprod	
aggr_lnlprod_rev	
aggr_lnSR	
aggr_lnlprod_rev	Revenue
aggr_lntfp_rev_adj_secCD	
aggr_lntfp_rev_macCD	
aggr_lntfp_rev_macTL	
aggr_lntfp_rev_secCD	
aggr_lntfp_rev_secTL	
aggr_lntfp_va_macCD	Value added
aggr_lntfp_va_secCD	

As the table shows depending on the indicator the estimates are weighted by the firm's share of value added, revenue or labour.

5.3.2 Production Function Estimation, TFP, and Marginal Products

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

We estimate several different types of production functions at the two-digit sector, the macro-sector, and the country level. In most cases, we rely on a Cobb-Douglas specification. For markups, however, we also estimate a translog type of production function as this allows for time varying output elasticities. Besides that, we apply gross-output and value-added production models. However, for the sake of brevity we cover only the value-added production function estimation in this section.

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

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

where rva_{it} is real value added, k_{it} is the real book value of net capital, l_{it} is total employment, and a_{it} *denotes hicks-neutral productivity*. ε_{it} is an i.i.d. error term and θ^x denotes the output elasticity of input $x = \{k_{it}, l_{it}\}$ To control for productivity and potential endogeneity concerns, we apply a control function approach as in Olley and Pakes (1996) and Levinsohn and Petrin (2003). Assuming that productivity evolves according to a Markov process and plugging in our control function for productivity, we can rewrite equation (3) as:

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

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

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

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

$$(3) \quad TFP_{it} = a_{it} = rva_{it} - (\theta^k k_{it} + \theta^l l_{it}).^{30}$$

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

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

[30] We also provide a markup adjusted TFP following Gal (2013).

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

5.3.3 Allocative Efficiency

- 5.3.2.1 Static Allocative Efficiency (Olley and Pakes, 1996)

Olley and Pakes introduced a very simple-to-compute indicator of allocative efficiency measured by the industry-level 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:

$$(6) \qquad y_{si} = \sum_{i \in S} \theta_{it} \omega_{it} = \bar{\omega}_{st} + \sum_{i \in S} (\theta_{it} - \bar{\theta})(\omega_{it} - \bar{\omega}_{st})$$

where *S* is the set of firms belonging to industry *s*, θ_{it} and ω_{it} represent size and productivity of firm *i* at time *t*, respectively, ϑ_{st} and ω_{st} represent the unweighted mean size and productivity of industry *s* at time *t*, respectively.

The decomposition splits the weighted average of firm productivity in two components: the unweighted industry mean and the covariance between productivity and size. The latter captures allocative efficiency in industry *s* at time *t* since it reflects the extent to which firms with higher than average productivity have a greater market share in terms of size.

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

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

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

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

$$(7) \qquad \Delta y_{st} = \sum_{i\in C} \theta_{i,t-k}\Delta\omega_{it} + \sum_{i\in C} \omega_{i,t-k} - \widehat{\omega}_{s,t-k})\Delta\theta_{it} + \sum_{i\in C} \Delta\theta_{it}\ \Delta\omega\_it + \sum_{i\in N} \theta_{it}\left(\omega_{it} - \widehat{\omega}_{s,t-k}\right) - \sum_{i\in X} \theta_{i,t-k}\left(\omega_{i,t-k} - \widehat{\omega}_{s,t-k}\right)$$

Where Δ is the differential operator between *t-k* and *t*; *C* denotes continuing firms, *N* denotes entering firms, and *X* denotes exiting firms; ϑ_{it} and $\omega_{i,t}$ represent size and productivity of firm *i* at time *t*, respectively, ϑ_{st} and ω_{st} represent the weighted mean size and productivity of industry *s* at time *t*, respectively.

- 5.3.3.3 Petrin-Sivadasan Gap (Petrin and Sivadasan 2013)

Petrin and Sivadasan (2013) proposed a measure of labour's allocative efficiency, based on the absolute difference between the value of labour's marginal product and its marginal cost. Following Petrin and Sivadasan (2013) closely, we approximate the marginal cost of a labour input with the average wage. Hence, the absolute gap between the value of the marginal product of labour and its wage can be written as:

$$(8) \qquad |G_{it}| = |VMPL_{it} - w_{it}|,$$

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

- 5.3.3.4 Hsieh-Klenow Indicator (Hsieh and Klenow 2009)

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

5.3.4 Distressed Firms

"Distressed firms", sometimes also called "zombie firms", are often described in the literature as firms who in a perfectly competitive market would have been forced to exit the market already. There are many ways of defining zombie firms, see for example Caballero et al. (2008) or McGowan et al. (2013). The CompNet dataset includes 4 different zombie firm dummy specifications to identify distressed firms. These different indicators have different rationales of defining a zombie firm to mirror the variety present in the literature. In the

following all four types of zombie firm indicators are discussed: markup, negative profits, "not-high-growth" and interest coverage based indicators:

Variable: *D_Zombie_MU_macCD*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with Markup < 1 (Markup estimate based on Macro-sector Cobb-Douglas coefficient).

This variable takes the value 1 (which identifies the firm as a zombie) if the markup estimate of a firm is smaller than 1. The markup estimate is based on the **Cobb-Douglas production function** estimated at the **macro-sector level**.

Variable: *D_Zombie_MU_macTL*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with Markup < 1 (Markup estimate based on Macro-sector translog coefficient).

This variable takes the value 1 if the markup estimate of a firm is smaller than 1. The markup estimate is based on a **translog production function** estimated at the **macro-sector level**.

Variable: *D_Zombie_MU_secCD*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with Markup < 1 (Markup estimate based on sector Cobb-Douglas coefficient).

Variable: *D_Zombie_MU_secTL*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with Markup < 1 (Markup estimate based on sector translog coefficient).

Similar to the variables above these two variables also use the markup estimate as an identifier for zombie firms. The only difference is that the markup estimate is based on the Cobb-Douglas/translog production function at the **sector level.**

Variable: *D_Zombie_negprof*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with negative operating profits for three consecutive years.

In this definition, a firm is flagged as a zombie if its operating profits are negative for three consecutive years.

Variable: *D_Zombie_nothg*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Firms with negative operating profits for 3 years and are not categorized as high-growth firms.

This variable uses an extended definition of a zombie firm to the previous one. Firms are flagged as zombies if they have negative operating profits for three consecutive years **and** do not show high growth. High growth firms are defined as firms with an employment growth rate higher than 33.1% over the last three years.

Variable: *D_Zombie_intcov*

Description: Dummy variable which takes the value 1 if firms are zombie firms according to the following definition: Interest > oper.profits for 3 years (conditional on positive profits).

In this definition firms with positive operating profits are nevertheless flagged as zombies if their interest payments have exceeded their operating profits for three consecutive years.

5.3.5 Indicators of Credit Constraints

- ICC Indicator (SAFE):

CompNet has constructed a firm-level "indicator of credit constraints" (ICC), defining firms that can be considered credit-constrained based on their financial situation.

The first step to construct the ICC indicator is to match firms' responses about binding credit constraints from the Survey on Access to Finance of Enterprises (SAFE) with their financial characteristics available in the AMADEUS database from Bureau van Dijk.

The SAFE is conducted by the ECB jointly with the European Commission twice per year. The survey intends to assess the financial conditions of firms in the Euro area (the survey is also conducted for some countries outside the Euro zone). 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).

After matching the firms' responses to survey with their financial statements available in the AMADEUS database from Bureau van Dijk, the second stage of the process is to estimate the impact of several indicators of the financial position of a firm on its probability to be credit constrained. More specifically, the regression equation is the following:

$$(9)\ Prob\,(credit\_constraint) = \alpha + \beta_1 \cdot finlev + \beta_2 \cdot ifp + \beta_3 \cdot pm + \beta_4 \cdot coll + \beta_5 \cdot cashH + \beta_6 \cdot lnTA + \gamma \cdot control\ var + \varepsilon$$

where $finlev$ is the financial leverage, ifp is the index of financial pressure, pm is profit margin, $coll$ is collateral, $cashH$ is cash holding and TA are the total assets. The control variables are time, sector, firm size and country-specific effects. For a more detailed explanation of the variables used in the regression, see Ferrando et al. (2015).

The third step is to use the coefficients of the estimated 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 for each firm *i* as:

$$(10)\quad SAFE\_score_i = -1.88 + 0.71 \cdot finlev_i + 0.28 \cdot ifp_i - 0.51 \cdot pm_i - 0.21 \cdot coll_i - 1.2 \cdot cashH_i - 0.05 \ln(TA_i)$$

Once the firms are ranked according to the SAFE score, the next step is to set a threshold value of the SAFE score above which we can define firms in a given level of aggregation as being credit constrained. The value of the threshold is time-varying and country-specific and

is set so that the share of firms above this threshold at the country level is the same as the share of credit constrained firms for a given country-year reported in the SAFE survey.

Last, we set $ICC_i = 1$ if the estimated SAFE score index is above the threshold we obtained from the before mentioned exercise, and $ICC_i = 0$ otherwise. The SAFE dummy variable in the CompNet database reflects the ICC values and the mean of the SAFE dummy consequently reports the share of credit constrained firms in any given level of aggregation.

5.3.6 Indicators of Market Imperfection

- 5.3.6.1. Labour Market Imperfection (Dobbelaere and Mairesse, 2013)

This indicator is designed to capture labour market imperfections and is derived according to Dobbelaere and Mairesse (2013). It captures the difference between the markup formulas from De Loecker and Warzynski (2008), based on intermediate input and labour decisions of the firm:

$$(11) \quad \mu^M = \theta^M * \frac{P_{it}Q_{it}}{z_{it}M_{it}}$$

$$(12) \quad \mu^L = \theta^L * \frac{P_{it}Q_{it}}{w_{it}L_{it}}$$

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.

Dobbelaere and Mairesse (2013) show that on perfect markets it holds that: $\mu^M = \mu^L$. Using the intermediate input market as a competitive benchmark, one can recover labour market frictions ϕ by substracting both markup expressions from each other: $\phi = \mu^M - \mu^L$.

- 5.3.6.2 Markup Estimation

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:

$$(13) \quad \mu_{it} = \alpha_{it}^{M} * \frac{P_{it}Q_{it}}{P_{it}^{M}M_{it}},$$

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

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

To analyse job flows at a given level of aggregation, we follow the seminal paper of Davis et al. (1996). We know by now that job flows are much larger, and can give a more accurate and pre-dated picture of labour market developments, than net employment growth recorded by aggregate statistics. Despite their importance, there are few sources providing comparable job flows at the highly disaggregated sector level. The measures are based on the firm-level growth rate of employment, which is computed in the following way:

$$(14) \qquad X_{it} = 0.5 \cdot (E_{it} + E_{it-1}) \text{ and } g_{it} = \frac{(E_{it} - E_{it-1})}{X_{it}}$$

Where X_{it} is the firm average employment (E_{it} and E_{it-1} are the employment in current and previous time point for a particular firm) and g_{it} is the firm-level growth rate of

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

employment. Since the growths rate incorporates both entry and exit, it also accounts for the creation and destruction respectively.

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

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

where X_{st} is the average employment for a particular sector. Therefore, at the 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.

5.4 Tables

Table 14: Data Sources and Data Provider

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
Belgium	-	-	-	National Bank of Belgium (BACH)	-	Micro-information underlying the financial ratio's in the BACH database*
Croatia	Yearly financial statements of firms	FINA	Financial Agency Croatia	Croatian National Bank	NFC drawn from total economy	
Czech Republic	Annual report of economic units in selected production industries P5-01	P501	Statistics Czech Republic	Czech National Bank	NFC drawn from total economy	full coverage for firms with >50 employees, stratified survey for smaller firms
	Extrastat/Intrastat foreign trade transaction data	TRADE	Statistics Czech Republic		NFC drawn from total economy	
	Business Register	RES	Statistics Czech Republic		NFC drawn from total economy	
Denmark	Accounts statistics - non-agricultural industries	Acc. Stat.	Statistics Denmark	Central Bank of Denmark	NFC	

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
	General enterprise statistics	Gen. Stat.	Statistics Denmark		NFC drawn from total economy	
Finland	Structural business and financial statement statistics data	SBS	Statistics Finland	Statistics Finland	NFC drawn from total economy	
	International trade statistics data	ITS	Finnish Customs		NFC drawn from total economy	
France	Regime of Normal Real Profits	BRN	Statistics France	Statistics France	NFC drawn from total economy	Complementing sources with RSI. BRN covers large firms +788K
	Simplified Regime for the Self-Employed	RSI	Statistics France		NFC drawn from total economy	Complementing sources with BRN. RSI covers small firms below 788K
Germany	administrative firm-level data	AFiD	German Statistics	German Statistics	Manufacturing	German sample only covers the manufacturing sector, subsequently only firms with more than 20 employees are included
Hungary	Tax registry database of National Tax and Customs Administration	NAV	National Tax and Customs Administration	Central Bank Hungary	NFC drawn from total economy	Non-mandatory variables for tax-records are underreported. E.g. 30% of Firms do not report the number of employees.

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
	Business Registry	VR	Statistics Hungary and Central Bank of Hungary		NFC drawn from total economy	
	Export-Import data of Hungarian Enterprises	Külker	Statistics Hungary		NFC drawn from total economy	
Italy	Statistical Business Register	ASIA	Statistics Italy	Statistics Italy	NFC drawn from total economy	
	Balance Sheets of non-financial companies	BIL	Statistics Italy		NFC drawn from total economy	
	Large enterprise survey	SCI	Statistics Italy		NFC drawn from total economy	
	Foreign Trade Statistics based on custom data	COE	Statistics Italy		NFC drawn from total economy	Complementary source which is targeted at large firms (+ 100 employees)
Lithuania	Statistical Survey on the Business Structure (Annual questionnaire F-01)	F01	Statistics Lithuania	Central Bank Lithuania	NFC drawn from total economy	
	Business Register	BR	Centre of Registers		NFC drawn from total economy	

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
	Customs, Customs declarations	CU	Customs of the Republic of Lithuania		NFC drawn from total economy	
Netherlands	Statistics finances of non-financial enterprises	SFO	Statistics Netherlands	Statistics Netherlands	NFC drawn from total economy	Full coverage for small firms (< 40 Mln balance sheet total); Large firms (> 40 Mln balance sheet totals) are surveyed
	Business register	ABR	Statistics Netherlands		NFC drawn from total economy	
Poland	Reports on revenues, costs, profit and outlays on fixed assets	F01	Statistics Poland	Central Bank Poland	NFC	Exclusion of firms with less than 10 employees
	Stat. financial report	F02	Statistics Poland		NFC	Exclusion of firms with less than 10 employees
Portugal	Central balance sheet database, annual survey	CBSD	Central Bank of Portugal	Banco de Portugal (BACH)	NFC	Micro-information underlying the financial ratio's in the BACH database*
	Simplified corporate information	IES	Statistics Portugal and Central Bank of Portugal		NFC drawn from total economy	Micro-information underlying the financial ratio's in the BACH database*
Romania	Balance sheet information on non-	Bal. Sheet	Ministry of Public finances	National Bank Romania	NFC drawn from total economy	

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
	financial enterprises					
	Exports and imports of goods, firm-level data	TRADE	Statistics Romania		NFC drawn from total economy	
Slovakia	Annual report on production industries	Reports	Statistics Slovakia	National Bank of Slovakia	NFC drawn from total economy	Exclusion of firms with less than 20 employees
	statistical register of organizations	Register	Statistics Slovakia		NFC drawn from total economy	Exclusion of firms with less than 20 employees
	foreign trade statistics	Customs	Statistics Slovakia		NFC drawn from total economy	Exclusion of firms with less than 20 employees
Slovenia	Slovenia Public and Legal Records and Related Services	AJPES	Agency for Public Legal Records and Related Services	Univ. of Ljubljana	NFC drawn from total economy	
Spain	CBSO voluntary survey	CBA	Central Bank of Spain	Banco de España (BACH)	NFC	Micro-information underlying the financial ratio's in the BACH database*
	Spanish mercantile register	CBB	Mercantile registry		NFC	Micro-information underlying the financial ratio's in the BACH database*
Sweden	Structured business statistics	SBS	Statistics Sweden	Statistics Sweden	NFC drawn from total economy	

Country	Data source name	Acronym	Institution responsible for source	Data provider	Firms included in dataset*	Source specific Information
	International trade in goods	ITG	Statistics Sweden		NFC drawn from total economy	
	Business register	BR	Statistics Sweden		NFC drawn from total economy	

Notes: NFC = non-financial corporations

Table 15: Macro-Sectors and Two-Digit NACE Rev. 2 Sectors Covered by the CompNet Database

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
C	1	Manufacturing	10	Manufacture of food products
			11	Manufacture of beverages
			12	Manufacture of tobacco products
			13	Manufacture of textiles
			14	Manufacture of wearing apparel
			15	Manufacture of leather and related products
			16	Manufacture of wood and of products of wood and cork, except furniture
			17	Manufacture of paper and paper products
			18	Printing and reproduction of recorded media
			19	Manufacture of coke and refined petroleum products
			20	Manufacture of chemicals and chemical products
			21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
			22	Manufacture of rubber and plastic products
			23	Manufacture of other non-metallic mineral products
			24	Manufacture of basic metals
			25	Manufacture of fabricated metal products, except machinery and equipment
			26	Manufacture of computer, electronic and optical products
			27	Manufacture of electrical equipment
			28	Manufacture of machinery and equipment n
			29	Manufacture of motor vehicles, trailers and semitrailers
			30	Manufacture of other transport equipment
			31	Manufacture of furniture
			32	Other manufacturing
			33	Repair and installation of machinery and equipment
F	2	Construction	41	Construction of buildings
			42	Civil engineering
			43	Specialised construction activities
G	3	Wholesale and retail	45	Wholesale and retail trade and repair

NACE Rev. 2 Section	Macro-sector in CompNet	Description	Sector in CompNet	Description
		trade; repair of motor vehicles and motorcycles		of motor vehicles and motorcycles
			46	Wholesale trade, except of motor vehicles and motorcycles
			47	Retail trade, except of motor vehicles and motorcycles
H	4	Transportation and storage	49	Land transport and transport via pipelines
			50	Water transport
			51	Air transport
			52	Warehousing and support activities for transportation
			53	Postal and courier activities
I	5	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
			79	Travel agency, tour operator and other reservation service and related activities
			80	Security and investigation activities
			81	Services to buildings and landscape activities
			82	Office administrative, office support and other business support activities

Table 16: List of Indicators

Indicator	Definition
aggr_lnkprod	Log of the kprod aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lnlprod	Log of the lprod aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lnlprod_rev	Log of the lprod_rev aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lnSR	Log of the SR aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_rev_adj_secCD	Log of the tfp_rev_adj_secCD aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_rev_adj_macCD	Log of the tfp_rev_adj_macCD aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_rev_macCD	Log of the tfp_rev_macCD aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_rev_secCD	Log of the tfp_rev_secCD aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_va_macCD	Log of the tfp_va_macCD aggregated using market share weights, used for the Foster decomposition of productivity growth
aggr_lntfp_va_secCD	Log of the tfp_va_secCD aggregated using market share weights, used for the Foster decomposition of productivity growth
cash_holdings	Cash divided by total assets
collateral	Capital divided by total assets
country	Country of the observed firms
ct_inte_to_op	Decile of the variable inte_to_op, used in the joint distributions
ct_invest_ratio	Decile of the variable invest_ratio, used for the joint distributions
ct_l	Decile of the variable l, used for the joint distributions
ct_lnlprod	Decile of the log of the variable lprod, used for the joint distributions
ct_lnlprod_rev	Decile of the log of the variable lprod_rev, used for the joint distributions
ct_lnSR	Decile of the log of the variable SR, used for the joint distributions
ct_lntfp_rev_macCD	Decile of the log of the variable tfp_rev_macCD, used for the joint distributions
ct_lntfp_rev_secCD	Decile of the log of the variable tfp_rev_secCD, used for the joint distributions
ct_lntfp_va_macCD	Decile of the log of the variable tfp_va_macCD, used for the joint distributions
ct_lntfp_va_secCD	Decile of the log of the variable tfp_va_macTL, used for the joint distributions
D_Zombie_intcov	Dummy equal to 1 if interest payments > operating profits for 3

Indicator	Definition
	consecutive years, conditional on positive profits
D_Zombie_MU_macCD	Dummy equal to 1 if Markup < 1; Markup estimate based on mac-sector Cobb-Douglas coefficients
D_Zombie_MU_macTL	Dummy equal to 1 if Markup < 1; Markup estimate based on mac-sector translog coefficients
D_Zombie_MU_secCD	Dummy equal to 1 if Markup < 1; Markup estimate based on sector Cobb-Douglas coefficients
D_Zombie_negprof	Dummy equal to 1 if operating profits are negative for 3 consecutive years.
D_Zombie_MU_secTL	Dummy equal to 1 if Markup < 1; Markup estimate based on sector translog coeff.
D_Zombie_nothg	Dummy equal to 1 if negative operating profits for 3 consecutive years - conditional that the firm is not experiencing high growth
debt_burd	Interest paid over total assets
dev_ag_lnkprod	Mean-deviation from log of the kprod aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lnlprod	Mean-deviation from log of the lprod aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lnlprod_rev	Mean-deviation from log of the lprod_rev aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lnSR	Mean-deviation from log of the SR aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_rev_adj_macCD	Mean-deviation from log of the tfp_rev_adj_macCD aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_rev_adj_secCD	Mean-deviation from log of the tfp_rev_adj_secCD aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_rev_macCD	Mean-deviation from log of the kprod aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_rev_secCD	Mean-deviation from log of the kprod aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_va_macCD	Mean-deviation from log of the kprod aggregated by market share, used for the Foster decomposition of productivity growth
dev_ag_lntfp_va_secCD	Mean-deviation from log of the kprod aggregated by market share, used for the Foster decomposition of productivity growth
domestic_sales	Total turnover minus turnover from sales abroad
DispMac_Imp_lab_macTL	Dispersion of Imp_lab_macTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_Imp_lab_secCD	Dispersion of Imp_lab_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends,

Indicator	Definition
	as in Kehrig (2011)
DispMac_lmp_lab_secTL	Dispersion of lmp_lab_secTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_kprod	Dispersion of kprod at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lnkprod	Dispersion of the log of kprod at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lnlprod	Dispersion of the log of lprod at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lnlprod_rev	Dispersion of the log of lprod_rev at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lnSR	Dispersion of the log of SR at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_adj_macCD	Dispersion of the log of tfp_rev_adj_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_adj_secCD	Dispersion of the log of tfp_rev_adj_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_macCD	Dispersion of the log of tfp_rev_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_macTL	Dispersion of the log of tfp_rev_macTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_secCD	Dispersion of the log of tfp_rev_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_rev_secTL	Dispersion of the log of tfp_rev_secTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_va_macCD	Dispersion of the log of tfp_va_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lntfp_va_secCD	Dispersion of the log of tfp_va_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lprod	Dispersion of lprod at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_lprod_rev	Dispersion of lprod_rev at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_markup_M_macCD	Dispersion of markup_M_macCD at macro-sector level,

Indicator	Definition
	computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_markup_M_macTL	Dispersion of markup_M_macTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_markup_M_secCD	Dispersion of markup_M_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_markup_M_secTL	Dispersion of markup_M_secTL at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_k_macCD	Dispersion of mp_r_k_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_k_secCD	Dispersion of mp_r_k_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_l_macCD	Dispersion of mp_r_l_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_l_secCD	Dispersion of mp_r_l_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_m_macCD	Dispersion of mp_r_m_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_r_m_secCD	Dispersion of mp_r_k_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_va_k_macCD	Dispersion of mp_va_k_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_va_k_secCD	Dispersion of mp_va_k_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_va_l_macCD	Dispersion of mp_va_l_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_mp_va_l_secCD	Dispersion of mp_va_l_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_pcm_Kfix	Dispersion of pcm_Kfix at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_pcm_Kvar	Dispersion of pcm_Kvar at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_PS_gap_rev_macCD	Dispersion of PS_gap_rev_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)

Indicator	Definition
DispMac_PS_gap_rev_secCD	Dispersion of PS_gap_rev_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_PS_gap_va_macCD	Dispersion of PS_gap_va_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_PS_gap_va_secCD	Dispersion of PS_gap_va_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_SR	Dispersion of SR at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_rev_adj_macCD	Dispersion of tfp_rev_adj_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_rev_adj_secCD	Dispersion of tfp_rev_adj_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_rev_macCD	Dispersion of tfp_rev_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_rev_secCD	Dispersion of tfp_rev_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_va_macCD	Dispersion of tfp_va_macCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispMac_tfp_va_secCD	Dispersion of tfp_va_secCD at macro-sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_Imp_lab_macTL	Dispersion of Imp_lab_macTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_Imp_lab_secCD	Dispersion of Imp_lab_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_Imp_lab_secTL	Dispersion of Imp_lab_secTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_kprod	Dispersion of kprod at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lnkprod	Dispersion of the log of kprod at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lnlprod	Dispersion of the log of lprod at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lnlprod_rev	Dispersion of the log of lprod_rev at sector level, computed cleaning the firm-level marginal productivity from sector trends,

Indicator	Definition
	as in Kehrig (2011)
DispSec_lnSR	Dispersion of the log of SR at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_adj_macCD	Dispersion of the log of tfp_rev_adj_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_adj_secCD	Dispersion of the log of tfp_rev_adj_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_macCD	Dispersion of the log of tfp_rev_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_macTL	Dispersion of the log of tfp_rev_macTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_secCD	Dispersion of the log of tfp_rev_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_rev_secTL	Dispersion of the log of tfp_rev_secTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_va_macCD	Dispersion of the log of tfp_va_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lntfp_va_secCD	Dispersion of the log of tfp_va_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lprod	Dispersion of lprod at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_lprod_rev	Dispersion of lprod_rev at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_markup_M_macCD	Dispersion of markup_M_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_markup_M_macTL	Dispersion of markup_M_macTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_markup_M_secCD	Dispersion of markup_M_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_markup_M_secTL	Dispersion of markup_M_secTL at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_k_macCD	Dispersion of mp_r_k_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_k_secCD	Dispersion of mp_r_k_secCD at sector level, computed cleaning

Indicator	Definition
	the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_l_macCD	Dispersion of mp_r_l_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_l_secCD	Dispersion of mp_r_l_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_m_macCD	Dispersion of mp_r_m_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_r_m_secCD	Dispersion of mp_r_k_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_va_k_macCD	Dispersion of mp_va_k_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_va_k_secCD	Dispersion of mp_va_k_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_va_l_macCD	Dispersion of mp_va_l_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_mp_va_l_secCD	Dispersion of mp_va_l_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_pcm_Kfix	Dispersion of pcm_Kfix at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_pcm_Kvar	Dispersion of pcm_Kvar at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_PS_gap_rev_macCD	Dispersion of PS_gap_rev_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_PS_gap_rev_secCD	Dispersion of PS_gap_rev_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_PS_gap_va_macCD	Dispersion of PS_gap_va_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_PS_gap_va_secCD	Dispersion of PS_gap_va_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_SR	Dispersion of SR at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_tfp_rev_adj_macCD	Dispersion of tfp_rev_adj_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)

Indicator	Definition
DispSec_tfp_rev_adj_secCD	Dispersion of tfp_rev_adj_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_tfp_rev_macCD	Dispersion of tfp_rev_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_tfp_rev_secCD	Dispersion of tfp_rev_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_tfp_va_macCD	Dispersion of tfp_va_macCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
DispSec_tfp_va_secCD	Dispersion of tfp_va_secCD at sector level, computed cleaning the firm-level marginal productivity from sector trends, as in Kehrig (2011)
dummy_euro	Dummy for EURO Countries (Italy, Germany, France, Belgium, Spain, Portugal, Finland, Lithuania, Slovakia, Slovenia, Netherlands)
dummy_west	Dummy for West European Countries (Italy, Germany, France, Belgium, Spain, Portugal, Finland, Sweden, Denmark, Netherlands)
Dummy_exp	Dummy equal to 1 if the firm is exporting; minimum 0.05% of total turnover from sales abroad
Dummy_exp_3y	Dummy equal to 1 if the firm is exporting for more than 3 years
Dummy_exp_new	Dummy equal to 1 if the firm is exporting and it was not in the previous year
Dummy_exp_new2	Dummy equal to 1 if the firm is exporting and was not in the previous year
Dummy_exp_no2y	Dummy equal to 1 if the firm is not exporting for the last 2 years
Dummy_exp_no3y	Dummy equal to 1 if the firm is not exporting for the last 3 years
Dummy_exp_switch	Dummy equal to 1 if the firm is exporting but it was not in the previous neither in the following year
Dummy_stop_exp	Dummy equal to 1 if the firm is not exporting and it was in the previous year
Dummy_stop_exp2	Dummy equal to 1 if the firm is not exporting and was not in the previous year
equity_debt	Equity divided by debt
exp_ratio	turnover from sales abroad over total turnover
exp_share_sector	Firm's export value over the total export value of sector
exp_vad	exports in value added
financial_gap	Growth in fixed capital plus depreciation of capital minus cash flow, divided by turnover
gr_rate_l1	Annual employment growth rate
gr_rate_l3	Employment growth rate over 3 years
gr_rate_rk1	Annual real capital growth rate
gr_rate_rk3	Real capital growth rate over 3 years
hhi_rev_nom	Hirschman-Herfindahl index of market concentration

Indicator	Definition
Imp_lab_macCD	Dobbelaere-Mairesse (2013) indicator of market imperfection, using a revenue-based Cobb-Douglas production function estimation at mac-sector level
Imp_lab_secCD	Dobbelaere-Mairesse (2013) indicator of market imperfection, using a revenue-based Cobb-Douglas production function estimation at sector level
inte_to_op	Interest over operating profits
invest_ratio	Growth rate of capital plus depreciation, divided by Capital
jcr	Weighted average of positive growth rates of number of employees
jdr	Weighted average of negative growth rates of number of employees
kprod	Real value added divided by capital
l	Average of the number of workers in in terms of classification used by data provider over the year
lc	Nominal labour costs, constructed using labour cost which vary slightly in terms of including wages and employers' social security contributions across countries
lc_l	Nominal labour costs divided by the number of employees
leverage	Debt divided by total assets
LP_lnkprod	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of kprod
LP_lnlprod	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of lprod
LP_lnlprod_rev	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of lprod_rev
LP_lnSR	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of SR
LP_lntfp_rev_adj_macCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_macCD
LP_lntfp_rev_adj_secCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_secCD
LP_lntfp_rev_macCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of tfp_rev_macCD
LP_lntfp_rev_secCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth based on the log of tfp_rev_secCD
LP_lntfp_va_macCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth, based on the log of tfp_va_macCD
LP_lntfp_va_secCD	Sum of within, between and covariance terms of the Foster decomposition of productivity growth, based on the log of tfp_va_secCD

Indicator	Definition
lprod	Ratio of real value added over employees
lprod_rev	Ratio of real turnover added over employees
mac	Mac-sector of the observed firms. It could be: (1) Manufacturing; (2) construction; (3) wholesale and retail trade; (4) accommodation and food services; (5) transport and storage; (6) information and communication; (7) real state; (8) professional, scientific and technical activities; (9) administrative and support service activities
macsec_szcl	Mac-sector and size class of the observed firms. Size class could be: (1) 1-9 employees; (2) 10-19 employees; (3) 20-49; (4) 50-249; (5) 250 and more employees
markup_M_macCD	Markup à la De Loecker-Warzynski (2012) with materials as input, from a Cobb-Douglas estimation at mac-sector level using revenue
markup_M_macTL	Markup à la DeLoecker-Warzynski (2012) with materials as input, from a Translog estimation at sector level using revenue
markup_M_secCD	Markup à la De Loecker-Warzynski (2012) with materials as input,from a Cobb-Douglas estimation at mac-sector level using revenue
markup_M_secTL	Markup à la DeLoecker-Warzynski (2012) with materials as input, from a Translog estimation at sector level using revenue
mp_r_k_secCD	Marginal revenue product of capital, from a Cobb-Douglas estimation at sector level using revenue
mp_r_k_macCD	Marginal revenue product of capital, from a Cobb-Douglas estimation at mac-sector level using revenue
mp_r_l_macCD	Marginal revenue product of labour, from a Cobb-Douglas estimation at mac-sector level using revenue
mp_r_l_secCD	Marginal revenue product of labour, from a Cobb-Douglas estimation at sector level using revenue
mp_r_m_macCD	Marginal revenue product of materials, from a Cobb-Douglas estimation at mac-sector level using revenue
mp_r_m_secCD	Marginal revenue product of materials, from a Cobb-Douglas estimation at sector level using revenue
mp_va_k_macCD	Marginal revenue product of capital, from a Cobb-Douglas estimation at mac-sector level using value added
mp_va_k_secCD	Marginal revenue product of capital, from a Cobb-Douglas estimation at sector level using value added
mp_va_l_macCD	Marginal revenue product of labour, from a Cobb-Douglas estimation at mac-sector level using value added
mp_va_l_secCD	Marginal revenue product of labour, from a Cobb-Douglas estimation at sector level using value added
obs_lnSR	Obs. for Foster decomp based on: logarithmic Solow residual; weights 1/3 and 2/3
obs_lnkprod	Obs. for Foster decomp based on: ln capital productivity -va definition
obs_lnlprod	Obs. for Foster decomp based on: ln of labour productivity -va definition
obs_lnlprod_rev	Obs. for Foster decomp based on: ln of labour productivity - turnover definition
obs_lntfp_rev_adj_macCD	Obs. for Foster decomp based on: ln TFPR adjusted by mark-

Indicator	Definition
	ups, with mac-sector revenue Cobb-Douglas coefficients
obs_lntfp_rev_adj_secCD	Obs. for Foster decomp based on: ln TFPR adjusted by mark-ups, with mac-sector revenue Cobb-Douglas coefficients
obs_lntfp_rev_macCD	Obs. for Foster decomp based on: ln TFPR, with mac-sector revenue Cobb-Douglas coefficients
obs_lntfp_rev_macTL	Obs. for Foster decomp based on: ln TFPR, with mac-sector revenue translog coefficients
obs_lntfp_rev_secCD	Obs. for Foster decomp based on: ln TFPR, with mac-sector revenue Cobb-Douglas coefficients
obs_lntfp_rev_secTL	Obs. for Foster decomp based on: ln TFPR, with sector revenue Cobb-Douglas coefficients
obs_lntfp_va_macCD	Obs. for Foster decomp based on: ln TFPR, with mac-sector value added Cobb-Douglas coefficients
OP_count_lnSR	Obs. count for OP-Decomp based on: logarithmic Solow residual; weights 1/3 and 2
OP_count_lnlprod	Obs. count for OP-Decomp based on: ln of labour productivity - va definition
OP_count_lnlprod_rev	Obs. count for OP-Decomp based on: ln of labour productivity - turnover definition
OP_count_lntfp_rev_adj_macCD	Obs. count for OP-Decomp based on: ln TFPR adjusted by mark-ups, with mac-sector revenue Cobb-Douglas coefficients
OP_count_lntfp_rev_adj_secCD	Obs. count for OP-Decomp based on: ln TFPR adjusted by mark-ups, with mac-sector revenue Cobb-Douglas coefficients
OP_count_lntfp_rev_macCD	Obs. count for OP-Decomp based on: ln TFPR, with mac-sector revenue Cobb-Douglas coefficients
OP_count_lntfp_rev_macTL	Obs. count for OP-Decomp based on: ln TFPR, with mac-sector revenue translog coefficients
OP_count_lntfp_rev_secCD	Obs. count for OP-Decomp based on: ln TFPR, with mac-sector revenue Cobb-Douglas coefficients
OP_count_lntfp_rev_secTL	Obs. count for OP-Decomp based on: ln TFPR, with sector revenue Cobb-Douglas coefficients
OP_count_lntfp_va_macCD	Obs. count for OP-Decomp based on: ln TFPR, with mac-sector value added Cobb-Douglas coefficients
OP_count_lntfp_va_secCD	Obs. count for OP-Decomp based on: ln TFPR, with mac-sector value added Cobb-Douglas coefficients
OP_lnlprod	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of lprod
OP_lnlprod_rev	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of lprod_rev
OP_lnSR	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of SR
OP_lntfp_rev_adj_macCD	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_adj_macCD
OP_lntfp_rev_adj_secCD	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_adj_secCD
OP_lntfp_rev_macCD	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_macCD
OP_lntfp_rev_macTL	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_macTL
OP_lntfp_rev_secCD	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_adj_secCD
OP_lntfp_rev_secTL	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_rev_adj_secTL
OP_lntfp_va_macCD	Covariance term of the Olley-Pakes decomposition of

Indicator	Definition
	productivity (OP gap) based on the log of tfp_va_macCD
OP_lntfp_va_secCD	Covariance term of the Olley-Pakes decomposition of productivity (OP gap) based on the log of tfp_va_secCD
pcm_Kfix	Price-cost margin: revenues minus materials minus labour cost, divided by revenues
pcm_Kvar	Price-cost margin: revenues minus materials minus labour cost minus (estimated) capital cost, divided by revenues
profitmargin	EBIT over turnover
PS_gap_rev_macCD	Petrin-Sivadasan gap (2013) measure of misallocation, using a revenue-based Cobb-Douglas production function estimation at mac-sector level
PS_gap_rev_secCD	Petrin-Sivadasan gap (2013) measure of misallocation, using a revenue-based Cobb-Douglas production function estimation at sector level
PS_gap_va_macCD	Petrin-Sivadasan gap (2013) measure of misallocation, using a value added-based Cobb-Douglas production function estimation at mac-sector level
PS_gap_va_secCD	Petrin-Sivadasan gap (2013) measure of misallocation, using a value added-based Cobb-Douglas production function estimation at sector level
rk	Capital deflated with GDP deflator
rk_l	real capital divided by the number of employees
roa	Operating profit-loss divided by total assets
rturnover	Turnover deflated with the GDP deflator
rva	Value added deflated with sector specific deflators
SAFE	Dummy variable indicating if a firm is credit constrained (1) or not (0) based on the methodology used in the Survey on Access to Finance of Enterprises, as in Ferrando et al. (2015)
sector	Sector of the observed firms, referring to the 2-digit industry number according to NACE rev.2 classification
SR	Non-parametric Solow's residuals, from the equation: logSR = log(real value added) - 1/3*log(real capital) - 2/3*log(number of employees)
t_Zombie_intcov	Number of consecutive years the firm has the variable D_Zombie_intcov equal to 1
t_Zombie_negprof	Number of consecutive years the firm has the variable D_Zombie_negprof equal to 1
t10_exp	Dummy equal to 1 if the firm is among top 10 exporters by export revenues
t5_exp	Dummy equal to 1 if the firm is among top 5 exporters by export revenues
tfp_rev_adj_macCD	TFP from a Cobb-Douglas estimation at mac-sector level using revenue; adjusted by mark-ups
tfp_rev_adj_secCD	TFP from a Cobb-Douglas estimation at sector level using revenue; adjusted by mark-ups
tfp_rev_macCD	TFP from a Cobb-Douglas estimation at mac-sector level using revenue
tfp_rev_secCD	TFP from a Cobb-Douglas estimation at sector level using revenue

Indicator	Definition
tfp_va_macCD	TFP from a Cobb-Douglas estimation at mac-sector level using value added
tfp_va_secCD	TFP from a Cobb-Douglas estimation at sector level using value added
top	Share of turnover of the top 10 firms by total turnover
TOT_betw_lnkprod	Between term of the Foster decomposition of productivity growth based on the log of kprod
TOT_betw_lnlprod	Between term of the Foster decomposition of productivity growth based on the log of lprod
TOT_betw_lnlprod_rev	Between term of the Foster decomposition of productivity growth based on the log of lprod_rev
TOT_betw_lnSR	Between term of the Foster decomposition of productivity growth based on the log of SR
TOT_betw_lntfp_rev_adj_macCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_macCD
TOT_betw_lntfp_rev_adj_secCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_secCD
TOT_betw_lntfp_rev_macCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_macCD
TOT_betw_lntfp_rev_macTL	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_macTL
TOT_betw_lntfp_rev_secCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_secCD
TOT_betw_lntfp_rev_secTL	Between term of the Foster decomposition of productivity growth based on the log of tfp_rev_secTL
TOT_betw_lntfp_va_macCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_va_macCD
TOT_betw_lntfp_va_secCD	Between term of the Foster decomposition of productivity growth based on the log of tfp_va_secCD
TOT_cov_lnkprod	Covariance term of the Foster decomposition of productivity growth based on the log of kprod
TOT_cov_lnlprod	Covariance term of the Foster decomposition of productivity growth based on the log of lprod
TOT_cov_lnlprod_rev	Covariance term of the Foster decomposition of productivity growth based on the log of lprod_rev
TOT_cov_lnSR	Covariance term of the Foster decomposition of productivity growth based on the log of SR
TOT_cov_lntfp_rev_adj_macCD	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_adj_macCD
TOT_cov_lntfp_rev_adj_secCD	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_adj_secCD
TOT_cov_lntfp_rev_macCD	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_macCD
TOT_cov_lntfp_rev_macTL	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_macTL
TOT_cov_lntfp_rev_secCD	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_secCD
TOT_cov_lntfp_rev_secTL	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_rev_secTL
TOT_cov_lntfp_va_macCD	Covariance term of the Foster decomposition of productivity

Indicator	Definition
	growth based on the log of lntfp_va_macCD
TOT_cov_lntfp_va_secCD	Covariance term of the Foster decomposition of productivity growth based on the log of lntfp_va_secCD
TOT_diff_lnkprod	The variable LP_lnkprod minus the variable dev_ag_lnkprod, used for the Olley-Pakes decomposition of productivity
TOT_diff_lnlprod	The variable LP_lnlprod minus the variable dev_ag_lnlprod, used for the Olley-Pakes decomposition of productivity
TOT_diff_lnlprod_rev	The variable LP_lnprod_rev minus the variable dev_ag_lnlprod_rev, used for the Olley-Pakes decomposition of productivity
TOT_diff_lnSR	The variable LP_lnSR minus the variable dev_ag_lnSR, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_adj_macCD	The variable LP_lntfp_rev_adj_macCD minus the variable dev_ag_lntfp_rev_adj_macCD, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_adj_secCD	The variable LP_lntfp_rev_adj_secCD minus the variable dev_ag_lntfp_rev_adj_secCD, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_macCD	The variable LP_lntfp_rev_macCD minus the variable dev_ag_lntfp_rev_macCD, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_macTL	The variable LP_lntfp_rev_macTL minus the variable dev_ag_lntfp_rev_macTL, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_secCD	The variable LP_lntfp_rev_secCD minus the variable dev_ag_lntfp_rev_secCD, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_rev_secTL	The variable LP_lntfp_rev_secTL minus the variable dev_ag_lntfp_rev_secTL, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_va_macCD	The variable LP_lntfp_va_macCD minus the variable dev_ag_lntfp_va_macCD, used for the Olley-Pakes decomposition of productivity
TOT_diff_lntfp_va_secCD	The variable LP_lntfp_va_secCD minus the variable dev_ag_lntfp_va_secCD, used for the Olley-Pakes decomposition of productivity
TOT_with_lnkprod	Within term of the Foster decomposition of productivity growth based on the log of kprod
TOT_with_lnlprod	Within term of the Foster decomposition of productivity growth based on the log of lprod
TOT_with_lnlprod_rev	Within term of the Foster decomposition of productivity growth based on the log of lprod_rev
TOT_with_lnSR	Within term of the Foster decomposition of productivity growth based on the log of SR
TOT_with_lntfp_rev_adj_macCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_macCD
TOT_with_lntfp_rev_adj_secCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_adj_secCD
TOT_with_lntfp_rev_macCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_macCD

Indicator	Definition
TOT_with_lntfp_rev_macTL	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_macTL
TOT_with_lntfp_rev_secCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_secCD
TOT_with_lntfp_rev_secTL	Within term of the Foster decomposition of productivity growth based on the log of tfp_rev_secTL
TOT_with_lntfp_va_macCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_va_macCD
TOT_with_lntfp_va_secCD	Within term of the Foster decomposition of productivity growth based on the log of tfp_va_secCD
trade_credit	Accounts payable over total assets
trade_debt	Accounts receivable over total assets
trader_2way	Firms that export and import
TRmat_l	Classification by the transition across size classes (based on l), used for the transition matrices
ulc	Nominal labour costs divided by real value added
UWM_lnlprod	Unweighted mean of the log of lprod, used for the Olley-Pakes decomposition of productivity
UWM_lnlprod_rev	Unweighted mean of the log of lprod_rev, used for the Olley-Pakes decomposition of productivity
UWM_lnSR	Unweighted mean of the log of SR, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_adj_macCD	Unweighted mean of the log of tfp_rev_adj_macCD, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_adj_secCD	Unweighted mean of the log of tfp_rev_adj_secCD, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_macCD	Unweighted mean of the log of tfp_rev_macCD, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_macTL	Unweighted mean of the log of tfp_rev_macTL, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_secCD	Unweighted mean of the log of tfp_rev_secCD, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_rev_secTL	Unweighted mean of the log of tfp_rev_secTL, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_va_macCD	Unweighted mean of the log of tfp_va_macCD, used for the Olley-Pakes decomposition of productivity
UWM_lntfp_va_secCD	Unweighted mean of the log of tfp_va_secCD, used for the Olley-Pakes decomposition of productivity
W_Imp_lab_macCD	Dobbelaere-Mairesse (2013) indicator of market imperfection weighted by employees, using a revenue-based Cobb-Douglas production function estimation at mac-sector level
W_Imp_lab_secCD	Dobbelaere-Mairesse (2013) indicator of market imperfection weighted by employees, using a revenue-based Cobb-Douglas production function estimation at sector level
wage_premium	Percentage difference of firm's level lc_l from the sector median
wageshare	Labour costs divided by value added
WM_lnlprod	Weighted mean of the log of lprod by marketshare, used for the Olley-Pakes decomposition of productivity

Indicator	Definition
WM_lnlprod_rev	Weighted mean of the log of lprod_rev by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lnSR	Weighted mean of the log of SR by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_rev_adj_macCD	Weighted mean of the log of tfp_rev_adj_macCD by market share, used for the Olley-Pakes decomposition of productivity
WM_lntfp_rev_adj_secCD	Weighted mean of the log of tfp_rev_adj_secCD by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_rev_macCD	Weighted mean of the log of tfp_rev_macCD by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_rev_macTL	Weighted mean of the log of tfp_rev_macTL by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_rev_secCD	Weighted mean of the log of tfp_rev_secCD by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_va_macCD	Weighted mean of the log of tfp_va_macCD by marketshare, used for the Olley-Pakes decomposition of productivity
WM_lntfp_va_secCD	Weighted mean of the log of tfp_va_secCD by marketshare, used for the Olley-Pakes decomposition of productivity
year	Year when the firms are observed

5.5 References

Altomonte, Carlo, Eric Bartelsman, Jan-Paul van de Kerke, Paloma López-Garcia, Filippo di Mauro, Marc Melitz, Michael Polder, and Sebastien Roux. "Assessing the reliability of the CompNet micro-aggregated dataset for policy analysis and research: Coverage, representativeness and cross-EU countries comparability." *The Competitiveness Research Network* (Forthcoming).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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