EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1634 / FEBRUARY 2014

MICRO-BASED EVIDENCE OF EU COMPETITIVENESS

THE COMPNET DATABASE

CompNet Task Force

In 2014 all ECB publications feature a motif taken from the €20 banknote.

THE COMPETITIVENESS RESEARCH NETWORK

CompNet The Competitiveness Research Network

This paper presents research conducted within the Competitiveness Research Network (CompNet). The network is composed of economists from the European System of Central Banks (ESCB) - i.e. the 28 national central banks of the European Union (EU) and the European Central Bank – a number of international organisations (World Bank, OECD, EU Commission) universities and think-tanks, as well as a number of non-European Central Banks (Argentina and Peru) and organisations (US International Trade Commission).

The objective of CompNet is to develop a more consistent analytical framework for assessing competitiveness, one which allows for a better correspondence between determinants and outcomes.

The research is carried out in three workstreams: 1) Aggregate Measures of Competitiveness; 2) Firm Level; 3) Global Value Chains CompNet is chaired by Filippo di Mauro (ECB). Workstream 1 is headed by Chiara Osbat, Giovanni Lombardo (both ECB) and Konstantins Benkovskis (Bank of Latvia); workstream 2 by Antoine Berthou (Banque de France) and Paloma Lopez-Garcia (ECB); workstream 3 by João Amador (Banco de Portugal) and Frauke Skudelny (ECB). Julia Fritz (ECB) is responsible for the CompNet Secretariat.

The refereeing process of CompNet papers is coordinated by a team composed of Filippo di Mauro (ECB), Konstantins Benkovskis (Bank of Latvia), João Amador (Banco de Portugal), Vincent Vicard (Banque de France) and Martina Lawless (Central Bank of Ireland).

The paper is released in order to make the research of CompNet generally available, in preliminary form, to encourage comments and suggestions prior to final publication. The views expressed in the paper are the ones of the author(s) and do not necessarily reflect those of the ECB, the ESCB, and of other organisations associated with the Network.

Acknowledgements
Comments by Eric Bartelsmann and John Haltiwanger are gratefully acknowledged. Gaia Greta Dossi and Alessandro di Nola provided skilful research assistance. This paper does not represent neither the views of the ECB nor those of the participating NCBs.

Complete list of authors by institution
The paper was written and coordinated by Paloma Lopez-Garcia, Filippo di Mauro (both ECB) and Carlo Altomonte (Bocconi University) with input from co-authors at 13 national central banks and a number of other institutions.

European Central Bank	Paloma Lopez-Garcia, Filippo di Mauro, Nicola Benatti, Chiara Angeloni
Bocconi University and Bruegel	Carlo Altomonte
Banca d'Italia	Matteo Bugamelli, Leandro D'Aurizio
EFIGE	Giorgio Barba Navaretti, Emanuele Forlani
ISTAT	Stefania Rossetti, Davide Zurlo
Banque de France	Antoine Berthou, Charlotte Sandoz-Dit-Bragard
Banque Nationale de Belgique/Nationale Bank van België	Emmanuel Dhyne
Banco de Portugal	João Amador, Luca David Opromolla, Ana Cristina Soares
Banca Nationala a României	Bogdan Chiriacescu, Ana-Maria Cazacu
Národná banka Slovenska	Tibor Lalinsky
Deutsche Bundesbank	Elena Biewen, Sven Blank, Philipp Meinen
Narodowy Bank Polski	Jan Hagemejer
Banco de España	Patry Tello, Antonio Rodríguez-Caloca
Banka Slovenije	Urška Čede
Česká národní banka	Kamil Galuščák
Eesti Pank	Jaanika Meriküll
Magyar Nemzeti Bank	Péter Harasztosi

Address Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Internet http://www.ecb.europa.eu
Fax +49 69 1344 6000

ISSN 1725-2806 (online)
EU Catalogue No QB-AR-14-008-EN-N (online)

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2361366.

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ABSTRACT

Drawing from confidential firm-level balance sheets in 11 European countries, the paper presents a novel sectoral database of comparable productivity indicators built by members of the Competitiveness Research Network (CompNet) using a newly developed research infrastructure. Beyond aggregate information available from industry statistics of Eurostat or EU KLEMS, the paper provides information on the distribution of firms across several dimensions related to competitiveness, e.g. productivity and size. The database comprises so far 11 countries, with information for 58 sectors over the period 1995-2011. The paper documents the development of the new research infrastructure, describes the database, and shows some preliminary results. Among them, it shows that there is large heterogeneity in terms of firm productivity or size within narrowly defined industries in all countries. Productivity, and above all, size distribution are very skewed across countries, with a thick left-tail of low productive firms. Moreover, firms at both ends of the distribution show very different dynamics in terms of productivity and unit labour costs. Within-sector heterogeneity and productivity dispersion are positively correlated to aggregate productivity given the possibility of reallocating resources from less to more productive firms. To this extent, we show how allocative efficiency varies across countries, and more interestingly, over different periods of time. Finally, we apply the new database to illustrate the importance of productivity dispersion to explain aggregate trade results.

Keywords

cross country analysis, firm-level data, competitiveness, productivity and size distribution, total factor productivity, allocative efficiency.

JEL codes

L11, L25, D24, O4, O57

NON-TECHNICAL SUMMARY

The economic literature has long recognized that firm-level data delivers crucial information for understanding the drivers of competitiveness, as aggregate performance depends strongly on firm-level decisions and shocks might have a different macroeconomic impact depending on the underlying distribution of firms. For these reasons, one of the pillars of the Competitiveness Research Network of the EU System of Central Banks (CompNet) from the beginning has been the micro or firm-level based analysis of sector and country competitiveness. However, cross-country firm-level analysis is hindered in practice at least by two reasons. First, existing cross-country indicators based on firm-level data are often not comparable given that they refer to different periods, methodologies or use different variable definitions. Second, firm-level data is normally confidential. As a result, micro-based analysis of competitiveness remain mostly bounded at the national level, limiting therefore the scope for benchmarking analysis, that is, for looking for best practices in peer countries, and for better understanding the role of the regulatory framework and institutions on micro and macro developments.

One way to tackle the confidentiality and comparability issues associated to firm-level analysis is to start from firm-level data and generate customized indicators of firm dynamics at the industry level. This approach is known as "distributed micro-data analysis" and it has been followed by CompNet to set a new research infrastructure able to deliver cross-country firm-based indicators. The research infrastructure involves the ECB as well as 13 NCB, one National Statistical Institute (ISTAT) and the EFIGE team, resulting in a current coverage of 11 EU countries which together represent about two-thirds of European Union's GDP: Belgium, Czech Republic, Germany, Estonia, France, Hungary, Italy, Poland, Spain, Slovakia and Slovenia. With the input of all parties, a harmonized protocol or set of commands was put together to construct firm-level indicators on competitiveness-related variables such as productivity or unit labor costs. Special care was taken to ensure that the protocol harmonized cross-country data-management on a number of crucial areas, including the industry classification, use of deflators, outlier treatment and variable definition and computation.

The output of the exercise is a database featuring 11 EU countries and 58 NACE Rev.2 industries over the period 1995-2011 with comparable information on productivity performance and dynamics of underlying heterogeneous firms. The advantage of CompNet's research infrastructure with respect to existing information included in aggregate statistics is that it enables to keep much of the richness of firm-level data in terms of full distribution of variables or joint correlations computed at the firm-level. This very rich set of indicators will set the stage for devising better informed policy decisions.

Some of the preliminary findings emerging from this new database are as follows:

- There is large heterogeneity in terms of firm performance within narrowly defined sectors, even more than across sectors.
- Within sectors, firm performance cannot be proxied by a normal probability distribution. Instead the distribution is highly skewed, with a large number of poorly performing firms and few "champions". The fact that firm performance distribution is not symmetric implies that the "average" firm is not representative of the underlying distribution of firms.
- Firms at different ends of the performance distribution, that is, bottom and top performing firms within a given sector, behave very differently in terms of size, productivity dynamics or unit labor cost dynamics. More concretely, we find that more productive firms are systematically larger than low productive firms (top performing firms can be up to 10 times larger as the firm in the median of the productivity distribution in the same sector). Additionally, top productive firms, within a given sector, feature a more dynamic productivity evolution and, above all, contain to a larger extent labor costs.
- Firm performance dispersion within a sector can lead to higher sector productivity because it enables the reallocation of production inputs (resources) from low to high productive firms. The higher the covariance between productivity and size, at the firm level, the more efficient is resource allocation.
- The extent to which an efficient allocation of resources across firms operating within the same sector contributes to aggregate productivity varies widely across countries and sectors. We find, however, that in all countries the distribution of resources, labor in this case, is worse in non-tradable sectors than in the tradable one. This means that in non-tradable sectors the largest firms are not necessarily the most productive ones.
- The importance of considering the dispersion rather than merely the average of the productivity distribution is confirmed when we use our dataset to explain sector export performance across Europe. In sectors where the labor productivity distribution is more highly skewed (fatter and longer right tail), it appears more likely to observe higher volume of exports. This would support the claim that the performance of the top percentiles drives aggregate trade outcomes, independently of the average productivity level.

Overall, this paper is intended as a solid documentation of the newly constructed CompNet firm-level indicators database. It also uncovers some interesting regularities across countries as well as country-specific developments which sets the stage for future research and policy-oriented work.

1 INTRODUCTION

"A competitive economy, in essence, is one in which institutional and macroeconomic conditions allow productive firms to thrive. In turn, the development of these firms supports the expansion of employment, investment and trade."

Mario Draghi, President of the European Central Bank (2012)

The economic literature has since long recognised that firm-level data delivers crucial information for understanding the drivers of competitiveness, as aggregate performance depends strongly on firm-level decisions (on labour and capital markets as well as innovation and technological capacity). Moreover, widespread heterogeneity in firm's behaviour has been well documented (Caves 1998, Bartelsman and Doms 2000), thus highlighting the limits of models based on the representative agent hypothesis. These findings also suggest that better knowledge of the underlying distribution of size and productivity might be required in order to assess aggregate productivity growth, and thus competitiveness.

In policy terms, an important implication of the existing firm heterogeneity is that a similar policy shock might yield different results on (aggregate) competitiveness measures across countries or industries, with important consequences for welfare and distribution, depending upon the specific firm configuration prevailing at any moment in time.[1] It then follows that we need not only to improve on firm-level indicators, moving from averages to the knowledge of the entire distribution of firm performance, but also to incorporate in a more systematic way the impact of firm heterogeneity on 'standard' assessments of competitiveness.

As such, the analysis of the micro (firm-level) dimension is one of the key areas of work of the Competitiveness Research Network of the EU System of Central Banks (CompNet).[2]

However, a major limitation in this regard is that firm-level analysis in Europe is currently hampered by a lack of sufficient and comparable data across countries. More precisely, while significant research has already been initiated by a number of National Central Banks (NCB) also in the area of firm heterogeneity, for the time being these studies remain mostly bound at the national level, due to a number of issues (confidentiality, different methodologies, non-overlapping years) which limit the comparability across existing datasets. In other words, there is no such a thing as yet as a "Single Market for firm-level data" in the European Union.

1 Melitz and Redding (2013) provide a comprehensive summary of the different channels through which a trade shock interacts with firm heterogeneity in driving aggregate productivity.

2 ESCB CompNet website: http://www.ecb.europa.eu/home/html/researcher_compnet.en.html

CompNet has thus dedicated part of its first year's activities to overcome this serious limitation in the existing toolbox. In particular, CompNet has adopted a common methodology to analyse existing firm-level datasets available within each NCB, and to collect indicators based on the firm-level data while preserving at the same time their confidentiality. To this end, CompNet national teams have used state-of-the art statistical and econometric methodologies, also to ensure cross country comparability of the indicators. The ultimate objective of the project is to develop a micro-founded analysis of competitiveness' dynamics across countries which could be systematically used for policy purposes. With respect to existing work based of firm-level information, but reported in aggregated fashion (e.g. Eurostat), we are therefore able to exploit the information content coming not only from averages, but also from the distribution of firms across several dimension, e.g. productivity and size.

The aim of this paper is threefold: first, the paper documents the data collection process, second, it describes in detail the database, and, third, it presents some of the preliminary results emerging from this brand new source of information. Ultimately, the paper is intended as a solid documentation of the indicators database, to be used as a reference for forthcoming applications to be developed by smaller research teams of CompNet.

In terms of results, the paper confirms that there is high heterogeneity of firm performance within narrowly defined sectors across EU countries for what concerns productivity and costs, and discusses the ensuing policy implications in terms of competitiveness analysis. In particular, the findings of this paper are consistent with the idea that, in a setting characterized by firms which are heterogeneous in productivity (or costs), the average performance (value added, productivity, …) in a given country or industry, i.e. what is generally considered a proxy of 'competitiveness', depends not only on the productivity of the average firm, but also on the extent to which factors of production can be reallocated towards the most productive firms. In other words, aggregate performance (e.g. productivity) of an industry or a country correlates not only with the performance of the average firm, but also with the variance of the underlying distribution of firms, as a higher dispersion generates more scope for reallocating resources within the same industry/country towards better firms, thereby increasing observed aggregate performance.[3]

[3] The role of dispersion in the underlying firm-level distribution of productivity/costs as a component (together with the behaviour of the average firm) of aggregate performance can be proved theoretically through an appropriate aggregation, at the industry or country level, of the basic results developed within the Melitz-Ottaviano (2008) framework, an exercise presented in Annex 9. Note how this result depends on the shape of the demand function that is linear with variable markups; in a CES setting with constant markups, only the performance of the average firm would matter, with no role for dispersion. The empirical evidence presented in the paper, above it in its application to explain trade developments, supports however the idea that aggregate performance is (also) correlated to the dispersion of the underlying indicator at the firm level.

After having provided evidence across countries and industries which is consistent with the above intuition, we also show how the contribution of reallocation has changed during the crisis, declining in some countries and increasing in others.

As a further illustration of the importance of acknowledging the underlying distribution of firm's performance to explain aggregate dynamics, the paper uses the newly developed database to explain aggregate trade performance. In this application sector-level exports across CompNet countries are related to total factor and labour productivity within-sector dispersion, as well as to the higher moments of both distributions. The main finding is that current trade performances are positively, and significantly, correlated with two year-lagged productivity dispersion.

Following a description of data and methodology used to collect and construct the indicators (Section 2), the paper analyses firm productivity across the EU. Section 3 dwells first upon the dispersion of productivity and size distributions across EU countries and sectors and then looks at differences across heterogeneous firms, analysing some of the features of firms residing in different segments (e.g. bottom vs. top) of the productivity distribution. Section 4 examines how - in our sample - productivity distributions translate into indicators of allocative efficiency, both static and dynamic. Section 5 presents one application of the database whereby different moments of the productivity distribution are used to explain sector-specific trade developments. Section 6 concludes.

2 THE COMPNET DATABASE

2.1 CROSS-COUNTRY ANALYSIS

Cross-country analysis is crucial for policy-making. Benchmarking, looking for best practices in peer countries, has been widely used by international institutions such as the OECD or the World Bank. Furthermore, cross-country regressions have allowed researchers to better understand the impact of the regulatory framework and institutions on micro and macro developments. They are also very useful to investigate the impact of similar shocks on different economies based on their specific economic institutions and market structures. Those cross-country analyses have been based on National Account data, if focused on macro aggregates, OECD-STAN, EUROSTAT industry statistics or EU KLEMS if based on industry aggregates and, finally, Amadeus (Bureau van Dijk) or, more recently, survey-based data like EFIGE, if the analysis is based on firm-level dynamics.[4] Each of these levels of analysis is informative in some sense, although it also has some drawbacks. For example, cross-country comparisons using macro data might overlook the role that the specific sector structure of a country could play to explain observed differences. Industry-based comparisons, on the other hand, have allowed researchers to compare productivity trends across countries, taking into account their particular sector and size structures, as well as to account for observed growth differences. But this approach misses the information provided by the dynamics of the underlying firms. Moreover, industry aggregates are not so informative when evaluating the impact of policies given that one cannot disentangle whether policy shocks affect firms' decisions, market dynamics, or both.

Firm-level cross-country analyses are still rare due to the confidential nature of the data as well as lack of comparability across countries. For those reasons, the commercial databases compiled by Bureau van Dijk (Amadeus is the European version of it) with information from the firm registries have been widely used. The drawback of these databases, as it will be shown in section 2.7 below, is that some variables, like employment, basic for productivity analysis, are not compulsory in all countries so actual firm coverage is drastically reduced. Moreover, firm sampling is not random which implies that samples are very biased to large firms. Hence this type of commercial databases can be useful for some analyses (those sampling only listed companies, for example), but are not very informative if one wants to compare underlying distributions of firms. EFIGE[5], on the other hand, has proven very useful to undertake cross-

[4] The Community Innovation Survey has also been widely used in firm-level studies related to innovation and R&D activities.

[5] The EU-EFIGE/Bruegel-UniCredit dataset (in short the EFIGE dataset) is a database collected within the EFIGE project (European Firms in a Global Economy: internal policies for external competitiveness) supported by the Directorate General Research of the European Commission through its 7th Framework Programme and coordinated by Bruegel. For more information please refer to Altomonte and Aquilano (2012).

country studies of competitiveness. However, only one wave of data is available up to now which restricts the type of questions that can be addressed with the survey.

2.2 DISTRIBUTED MICRO-DATA ANALYSIS

One way to tackle the confidentiality and comparability issues associated to firm-level analysis is to start from firm-level data and generate customised indicators of firm dynamics at the industry level. This is the approach taken by the World Bank and the OECD project on firm dynamics (see Bartelsman et al. 2004 and Bartelsman et al. 2009), known as "distributed micro-data analysis." Following this type of methodology and taking advantage of the links between the European Central Bank (ECB) and the National Central Banks (NCB), CompNet has set a new research infrastructure able to deliver cross-country firm-based indicators.

The research infrastructure involves the ECB as well as 13 NCB, one National Statistical Institute (ISTAT) and the EFIGE team, resulting in a current coverage of 11 EU countries: Belgium, Czech Republic, Germany, Estonia, France, Hungary, Italy, Poland, Spain, Slovakia and Slovenia, which together represent about two-thirds of European Union's GDP. Portugal and Romania participate as well in the project although their data could not be included in this version of the paper. With the input of all parties, a harmonised protocol or set of commands was put together to construct firm-level indicators on competitiveness-related variables such as productivity or unit labour costs (more details are provided in the next section). Special care was taken to ensure that the protocol harmonized cross-country data-management on a number of crucial areas, including the industry classification, use of deflators, outlier treatment and variable definition and computation.

More concretely, industries were classified in all countries at the 2-digit NACE REV.2 level and deflators were obtained generally from EUROSTAT National accounts, at NACE rev2 64 sector level. These deflators were actually sent to each national team who checked them in detail. Belgium, France, the Czech Republic, Slovakia, Slovenia and Spain decided to provide their own set of deflators, constructed by their respective Central Banks or National Statistical Offices.

The correction for outliers was centrally implemented in the set of codes, using a standardised method in order to ensure cross-country comparability. The data cleaning exercise is twofold, being applied both to the initial variables and to the estimated productivity measures. The correction is applied on the growth rates of the variables and follows three steps. First, the observations with negative value-added are replaced as missing values. The second step creates the growth rates of the variables and identifies the 1st and the 99th percentiles of their distributions. Third, the values of variables in level and in logarithms are considered as missing values if the corresponding growth rates belong to the 1st percentile and the 99th percentile.[6] It is important to note that a very thorough sensitivity analysis has been carried out to test for the impact of different outlier treatments on the main results of the paper. The conclusion is that results are robust. [7]

Much care was devoted to ensure that variables were equally defined across all countries. A table with first and second best definitions was circulated well in advance. Indicators and estimations were computed following the exact same set of commands, which were executed by each national team on their respective firm-level samples. National teams were also in charge of providing detailed information on the metadata, that is, the description of their databases, sources, existing thresholds and comparisons with EUROSTAT data (details included in Annex 1).

The output of the exercise undertaken by each country is a set of indicators aggregated to the pre-specified industry level, which ensures confidentiality of the data. The indicators computed in the exercise provide, for each country, comparable key stylized facts at the industry/year level about productivity performance and dynamics of underlying heterogeneous firms. The advantage of CompNet's research infrastructure with respect to existing information included in aggregate statistics is that it enables to keep much of the richness of firm-level data in terms of full distribution of variables or joint correlations computed at the firm-level. This additional information will enable researchers to correctly interpret variations in productivity performance across countries, industries, and time periods, as well as to better forecast the impact of policies, given the underlying distribution of firms. Such a wealth of information should, therefore, set the stage for devising better informed policy decisions.

2.3 DESCRIPTION OF THE DATA

The unit of analysis is the firm, as opposed to the establishment, defined as a legal entity. This includes limited liability companies, limited partnerships, traded companies and the like.[8] Self-

---

6 The TFP, which is quite noisy, was also corrected in levels, excluding top and bottom 1%, in addition as in growth rates.

7 This sensitivity analysis is available upon request. It was performed on raw Estonian data from Amadeus, for the years 2003-2010, and included several methods to detect outliers, both in growth rates, level of key ratios or a combination of both. Different thresholds to define an outlier were considered, based on the p1 and p99 of the distribution and also on relative measures like the median at the industry/year level plus/minus 3 or 5 times the interquartile range. Results were very robust to all these different treatment of outliers, even if very few, or very many, observations were dropped. Note, however, that the robustness exercise was done using a very specific sample of firms (Estonian firms from Amadeus) so the results of the exercise could be slightly different if other sample had been taken instead.

8 Affiliates of foreign businesses are included as long as they are legally incorporated in the country.

employed (physical persons with economic activity) are not included. It is important to highlight from the outset that, currently, there is no information on the age of firms, so we cannot distinguish between incumbent and new firms. This is important when it comes to the analysis of resource reallocation given that one important channel of reallocation within sectors is the entry and exit of firms.

The underlying sources of the national firm-level data is diverse, although in most countries the information comes from business registries and/or balance sheets offices of the Central Banks or finance ministries. Fiscal sources are used by France, Belgium (for small firms) and Hungary.[9] The National Statistical Offices are the source of the Czech Republic, Poland and Slovakia. With respect to size thresholds, Poland and Slovakia have samples restricted to firms with more than 20 employees or more than 5 million euros of turnover (in the case of Slovakia). The rest of countries cover all size classes although, as it can be seen in Table 2, the coverage of the smallest size class is very different across countries. For this reason, we have defined two samples which are used in different parts of the paper, depending on the indicator analysed. The first one is the *full sample*, including all firms for which data is available in the different countries. The second sample is the *restricted sample* and includes only firms with more than 20 employees. The latter is more homogenous across countries so it will be used extensively throughout the paper.

Table 1 below provides general information on the full samples of countries. For a more detailed account of the country-specific data and sources, please refer to Annex 1.

9 In the case of France, the use of "Béneficés Réels Normaux" data was granted by the "Comité du Secret Statistique" to the authors of the paper at the Banque de France.

TABLE 1 COUNTRY SAMPLES

	Average No. of Firms	Average No. of Firms with more than 20 employees	Total Coverage/Eurostat [1] (Full sample)				Year Coverage	Sector Coverage [2] (Full sample)			Overlapping years with Eurostat
			Firms	N. of Employees	Turnover	Labour Costs		Mining and Quarrying	Manufacturing	Services [3]	
BELGIUM	66,842	7,757	33%	73%	86%	101%	1996-2011	3/5	full	full	2008-2010
CZECH REP.	21,156	11,470	4%	64%	78%	62%	2002-2011	full	full	full	2005-2010
ESTONIA	12,186	1,849	66%	85%	78%	62%	1995-2011	4/5	full	full	2005-2010
FRANCE	348,179	55,004	26%	84%	100%	87%	1995-2007	full	23/24	full	2007;2010
GERMANY	30,688	22,401	2%	38%	66%	55%	1997-2010	1/5	23/24	27/29	2008-2010
HUNGARY	13,683	5,368	4%	47%	73%	53%	2003-2010	full	full	full	2005-2010
ITALY	53,054	21,476	2%	17%	22%	24%	2001-2008	2/5	23/24	full	2008
POLAND	6,250	6,250	1%	14%	80%	86%	2002-2011	full	full	28/29	2005-2010
SLOVAKIA	3,954	3,446	8%	67%	89%	80%	2000-2011	full	23/24	full	2008-2010
SLOVENIA	16,676	2,123	33%	78%	92%	94%	1995-2011	4/5	23/24	full	2005-2010
SPAIN	245,121	22,770	24%	42%	58%	60%	1995-2011	full	full	full	2008-2010

[1] Average across comparable years (shown in the last column) and sectors in Eurostat and CompNet. Note that there are not overlapping years between CompNet and Eurostat for France. Hence coverage for France has been computed comparing 2007 CompNet data with 2010 Eurostat data, which could bias upward the comparison.
[2] Sector coverage should be read as follows: 3/5 means that 3 out of the 5 mining and quarrying industries are covered.
Similarly, 27/29 means that 27 out of 29 services industries are covered in that particular country. Full coverage means that all industries are covered.
[3] Financial services and trade are excluded from services

The time coverage of the sample is generally the period 1995-2011[10] although important differences across countries remain given that some countries like Czech Republic, Hungary, Italy, Poland, or Slovakia have data only from the early 2000s. Year 2011 is either not covered at all or only partially covered in some countries. For these reasons the analysis undertaken in the following sections of the paper will consider only the period 2002-2010. Sector coverage is quite complete, comprising up to 58 2-digit manufacturing and non-manufacturing NACE REV.2 industries (the detailed list of industries is provided in Annex 2).

Table 1 shows the coverage of the full sample versus the population, in terms of firms, number of employees, turnover, and labour costs.[11] Coverage rate in terms of firms varies widely, ranging from about 1% in Italy and Poland to more than 65% in Estonia. The coverage in terms of turnover or number of employees is higher, due to the fact that in countries with low coverage typically larger firms are sampled. To illustrate this, Table 2 below shows the sample representativeness in each country, in terms of size and broad sectors, as compared to the population of firms provided by Eurostat. Out of the 11 listed countries, 5 have reasonably representative samples in terms of size distribution and sector whereas 6 have an over-representation of large and manufacturing firms (Slovakia, Poland, Italy, Hungary, Germany and Czech Republic).

[10] The database is, however, expected to be updated regularly.

[11] The population is assumed to be the one provided by the Structural Business Statistics of Eurostat. However, Eurostat data are retrieved from surveys with incomplete coverage of the population of firms for Belgium and Estonia, which can explain ratios even above 100%. On the other hand, please be aware that in some countries like Spain and Slovenia (details in the notes of Table 2), CompNet data includes only companies whereas Eurostat data considers a much larger set of business entities which also includes sole proprietors.

TABLE 2 SAMPLE REPRESENTATIVENESS

	Size Distribution						Sector Distribution				Period
	0 to 19 employees		20 to 249 employees		250 or more employees		Manufacturing		Non-Manufacturing		
	Compnet	Eurostat	Compnet	Eurostat	Compnet	Eurostat	Compnet	Eurostat	Compnet	Eurostat	
BELGIUM	89.2%	96.0%	10.0%	3.8%	0.8%	0.2%	19.3%	14.4%	80.7%	85.6%	2008-2010
CZ	49.4%	97.4%	45.9%	2.4%	4.7%	0.2%	44.0%	25.6%	56.0%	74.4%	2005-2010
ESTONIA	88.3%	90.4%	11.2%	9.1%	0.5%	0.4%	23.4%	20.6%	76.6%	79.4%	2005-2010
FRANCE	86.3%	96.4%	12.8%	3.4%	0.9%	0.2%	19.2%	14.9%	80.8%	85.1%	2007;2009
GERMANY	22.7%	91.6%	61.7%	7.8%	15.6%	0.6%	51.2%	16.3%	48.8%	83.7%	2008-2010
HUNGARY	58.5%	97.2%	37.1%	2.6%	4.3%	0.2%	70.5%	16.2%	29.5%	83.8%	2005-2010
ITALY	56.5%	97.4%	42.5%	2.5%	1.0%	0.1%	58.2%	19.5%	41.8%	80.5%	2008
POLAND	0.0%	94.9%	95.4%	4.6%	4.6%	0.5%	63.3%	35.9%	36.7%	64.1%	2005-2010
ROMANIA	90.8%	91.3%	8.5%	8.0%	0.7%	0.7%	25.0%	25.5%	75.0%	74.5%	2005-2010
SLOVAKIA	14.9%	89.6%	76.0%	9.5%	9.1%	0.9%	50.4%	29.7%	49.6%	70.3%	2008-2010
SLOVENIA	88.6%	96.1%	10.5%	3.5%	0.8%	0.3%	25.6%	24.4%	74.4%	75.6%	2005-2010
SPAIN	93.0%	96.4%	6.5%	3.4%	0.5%	0.2%	19.4%	17.2%	80.6%	82.8%	2008-2010

Notes: Average across all common available sectors and years, specified in the last column.

Data of France from CompNet in 2007 is compared with that of 2009 from Eurostat, given the lack of overlapping years.

Data for Czech Rep., Spain, Slovenia, France, Hungary, Italy, Slovakia on self-employed persons are included in Eurostat, but excluded in CompNet.

To partially mitigate the bias towards large and manufacturing firms of some country samples, a common set of sector weights were created to compute country aggregates. The sector weights are computed as the average value added share of each industry across the 11 countries (more details in Box 1). However, a full robustness exercise has been carried out in order to explore the role of the different sector weighting systems on the main results of the paper. This exercise can be found in Annex 3 and shows that, generally, results hold also with country-specific or time-invariant weights.

BOX 1 COMMON SECTOR WEIGHTS COMPUTATION

The aggregation of sector data at the country level is done throughout the paper using a common set of sector weights for all countries. The reason is that, in some countries, there might be some small sectors which have outliers. Additionally, some countries have an over-representation of manufacturing firms. Country-specific weights might be highly affected by those issues. To partially mitigate these potential sources of bias we generate a set of common sector weights, computed as the share of value added each sector has in the whole economy (defined as the sum of all countries' values) in a specific year. More concretely:

Let s denote the sector, i the country and t the year. The sector weight is defined as:

$$\theta_{st} = \sum_i VA_{sit} / \sum_s \sum_i VA_{sit}$$

In this case the weight does not depend on the specific country and it is allowed to differ across time. Hence given a certain variable ϖ_{sit}, defined for country i, sector s and time t, we would compute the corresponding country aggregate y_{it} (or across-sectors average) as follows:

$$y_{it} = \sum_s \theta_{st} \varpi_{sit}$$

where θ_{st} is the sector weight.

Given the fact that most figures shown are based on sector aggregation, a robustness check to the use of alternative weighting schemes has been undertaken. The results of the exercise are extremely reassuring, and shown in Annex 3. The Annex shows Figures 1 and 4 redone using 4 alternative weighting schemes: (1) country-specific, time-variant sector weights; (2) common to all countries, time-invariant weights; (3) country-specific, time-invariant weights; (4) common to all countries, time-variant weights based on turnover, instead of on the value added share of the sector.

2.4 VARIABLES AND INDICATORS

The set of indicators available in the dataset (for details, see Annex 4) is broadly organised around three topics: (1) inputs and output of the production function, including value added, turnover, employment, fixed tangible assets, intermediate inputs and wages; (2) productivity-related indicators such as labour productivity, total factor productivity (TFP) and unit labour cost (ULC); and (3) allocative efficiency indicators, such as the Olley-Pakes (1996) covariance between market share and relative productivity and the Foster-Haltiwanger-Krizan (2006) accounting decomposition terms of aggregate productivity growth.

For each of the listed indicators, the dataset contains a number of descriptive statistics. In particular, exploiting the availability of firm-level data, we report not only the weighted average of a given indicator within each country/industry/year, but also different moments of the distribution like percentiles 1, 10, 25, 50, 75, 90 99, the maximum and minimum as well as the standard deviation, interquartile range and skewness. In addition, potentially meaningful correlations between indicators such as productivity and size are computed in each industry/year, as well as joint moments of the distribution of different pair of indicators. This implies that the database allows access to information on the average size, value added, total and average labour cost, tangible assets or turnover of firms in the various percentiles of the distribution of labour productivity, TFP and ULC. This type of information enables researchers to learn about dynamics and characteristics of firms located at the different tails of the performance distribution, which, as it will be shown in the next sections, are quite heterogeneous across countries and sectors. Understanding the underlying distribution of firms within each country/industry/year cell, as well as the particular firm dynamics in different parts of the distribution can be extremely useful for policy-makers in order to evaluate the impact of particular policies and shocks, or target particular groups of firms.

Inputs and output of the production process

Table 3 below includes the detailed definition of variables related to the inputs and output of the production process. Although a large effort has been made to harmonise definitions, country variation still remains, as can be seen in the third column of the table.

Employment, for example, is captured generally by full-time employment although part-time employment is also included in Slovenia, Spain, Poland, Germany, Belgium and Italy. Capital stock, on the other hand, is approximated by tangible fixed assets, at book value, included in the balance sheets of firms,[12] although intangible assets are also included in Poland and Hungary. Capital is deflated by the GDP deflator, obtained generally from EUROSTAT National accounts, at NACE rev2 64 sector level (see country details in Annex 1). Intermediate inputs include in most cases energy expenditures and are deflated with the deflator corresponding to sector 35 in NACE REV.2 (electricity, gas, steam and air conditioning supply). The cost of employees is proxied by wages and salaries paid by the firm, including compulsory employers' contributions to the social security. Finally, value added is generally measured as turnover (total sales net of the value added tax) minus intermediate costs, although some countries use the

[12] Note that this proxy of capital stock, based on balance-sheet data, might lead to measurement error in capital. See Galuscak and Lizal (2011) for discussion on measurement error in capital and results showing how the coefficient estimates change when the measurement error in capital is accounted for.

difference between production value, which includes sales and stocks, and intermediate inputs. Turnover is deflated using the GDP deflator.

TABLE 3	INPUTS AND OUTPUT OF THE PRODUCTION PROCESS. VARIABLE DEFINITION	
Variable	Common variable definition	Deviations to common definiton (countries included only when variables not exactly defined as in column 2)
Number of employees	Full-time employment, yearly average	DE: Total full- and part-time employment BE: Average Full-Time Equivalent over the accounting period CZ: Average number of employees ES: Full-Time Equivalent average in the case of large firms. Average total employment in the case of SMEs HU: Full-time equivalent average for the whole economic year IT: Total employment, full-time and part-time PL: Total employment at the end of the year SI: Average number of employees based on the number of work-hours in the period RO: Total employment, including full-time and part-time. Average of monthly total employment.
Total assets (Capital)	Tangible fixed assets	DE: Tangible assets - financial assets HU: Tangible + intangible assets IT: Tangible assets PL: Tangible + intangible assets
Material costs	Intermediate inputs, including energy	CZ: Intermediate inputs, excluding energy DE: Intermediate inputs, excluding energy FR: Intermediate inputs, excluding energy SI: All operating expenses excluding labour costs and write-downs in value
Cost of Employees	Wages + employer's contributions to obligatory social security funds	CZ: Wages without employer’s contributions DE: It includes all benefits paid by the employer SI: Total labour costs, including employer’s and employee’s compulsary contribution to the social security
Value added	Turnover - intermediate inputs	CZ: Production (=sales + stocks + activation) - intermediate inputs ES: Production - intermediate inputs FR: Production - intermediate inputs HU: Production - intermediate inputs RO: Production + capitalised production - intermediate inputs SI: Gross operating returns - intermediate inputs SK: Production - intermediate inputs
Turnover	Total sales net of VAT	CZ: Total sales ES: Total sales net of discounts and VAT FR: Sales include services SK: Gross turnover

Productivity indicators

Labour productivity is measured as real value added divided by employment. An alternative productivity indicator, turnover-based, is also computed as turnover over employment. A measure of capital productivity, defined as real value added per unit of the real stock of capital is also estimated. Finally unit labour costs are computed as nominal labour costs over real value added.

Total Factor Productivity is estimated as the residual of an estimated production function whereby labour, capital and intermediate inputs are used to produce a given level of output. We follow the semi-parametric estimator or control function approach proposed by Olley and Pakes (1996), Levinshon and Petrin (2003), Ackenberg et al (2006) and Wooldridge (2009). This approach uses observed input choices to instrument for unobserved productivity in order to control for the so-called "simultaneity bias". This bias emerges from the fact that despite firm-level productivity is not observed by the econometrician, it is known by the firm. This means that the firm, every period, will choose optimally the set of inputs of the production function after observing its own productivity. Not taking into account the fact that labour, capital and intermediate inputs are correlated to unobserved productivity, at the firm level, will yield inconsistent OLS estimators of the production function coefficients.

Olley and Pakes (1996) proposed a structural solution to the simultaneity bias by instrumenting unobserved firm-level productivity with investment and capital stock. However, a later paper by Levinshon and Petrin (2003) observed that investment is quite bumpy and can contain a lot of zeros, with the consequent loss of efficiency in the estimation. For that reason, Levinshon and Petrin (2003) proposed to instrument unobserved productivity with capital stock and intermediate inputs, arguing that any active firm will demand a positive amount of energy and raw material to produce. Ackerberg et al. (2006) improved on both methodologies by allowing the freely available input, labour, to be correlated with investment/intermediate input choices and thus also instrumented with its past values. Wooldridge (2009) proposed some years later to implement these approaches in a GMM framework, proving that within this framework there are efficiency gains. All these technical details can be found in Annex 6.

For the construction of the CompNet database both the approaches of Levinshon and Petrin (2003) and Wooldridge (2009) were followed in order to make sure that TFP differences were not driven by the chosen estimation methodology. Both sets of estimated coefficients of the production function were, however, very similar. Therefore the exercise below will only show results using the Wooldridge (2009) approach as implemented in Petrin and Levinsohn (2012) and Galuscak and Lizal (2011). Table A6 in the Annex 6 shows the estimated coefficients of labour and capital for each of the countries in the sample, grouped in 9 broad industries. The low estimated coefficient of capital can be due to the above mentioned measurement issues. However, it could also be the results of an attrition bias in the country samples whereby only good (thus surviving) firms are selected into the sample. This kind of selection bias might introduce a correlation between inputs and the error term which biases both capital and labour coefficients. For these reasons our analysis will discuss the dynamics of labour productivity together with TFP.

Production functions were estimated at the country/industry level. Hence, all firms in the same industry, within a given country, are assumed to have the same marginal returns of labour and capital.[13] Given those technology coefficients, predicted value added for each firm was computed taking into account its level of capital and labour. Finally, the difference between the actual and predicted value added was assumed to be due to efficiency differences in the use of inputs by each firm, or in other words, assumed to be the estimated firm-level total factor productivity.[14]

Allocative efficiency indicators

In general terms, allocative efficiency refers to a situation where available resources are put to their best use. The literature on heterogeneous firms has identified two closely interlinked definitions of allocative efficiency, one static and another dynamic (Haltiwanger, 2011). By static allocative efficiency is meant the extent to which, in the cross-section, firms with higher than average productivity have a larger than average size in the sector. That is, the static concept of allocative efficiency provides a snap-shot of how resources are allocated at a certain moment in time. To measure this concept, we follow Olley and Pakes (1996), who decomposed an index of industry-productivity level into an unweighted average of labour productivity, of all firms in the industry, and a covariance term between relative labour productivity and relative size of the firm (see Annex 7 for a detailed account of the Olley-Pakes decomposition). Hence, the covariance term could be interpreted as the degree to which resources are allocated efficiently across firms within the same industry. A low covariance indicates that aggregate productivity can improve by reallocating resources towards the most productive firms.

The dynamic allocative efficiency refers instead to the process whereby outputs and inputs are being reallocated from lower to higher productivity units. In particular, dynamic allocative

[13] Altomonte et al (2012) and Fernandez and Lopez-Garcia (2013) show, on a different set of firms observed across countries, that the main differences in terms of the technology coefficients of the production function are across industries, more than across countries and even across size classes.

[14] Please be aware that we do not have firm-specific prices for output or intermediate inputs, but rather industry deflators. This implies that within-industry price differences will be embodied in productivity measures. Hence we will compute a "revenue-based" productivity measure rather than a "physical output" one. See Foster et al. (2008) for a full discussion on this issue.

efficiency implies that firms that have gained market share in the sector between two periods of time are those with higher than average productivity. Dynamic resource reallocation is commonly measured within the framework of the accounting decomposition of aggregate productivity growth (see Foster, Haltiwanger and Krizan 2006 and Annex 7 for details). In this framework, productivity growth, at the industry level, is decomposed into four terms: the contribution of established firms, the contribution of new firms, the contribution of exiting firms, and the contribution of resource reallocation between incumbent firms within the sector. The latter is captured by the so-called "between term," which reflects to what extent resources (labour and capital) are reallocated from the least to the most productive established firms in the industry. A final term, the "covariance term," captures the correlation from changes in productivity and changes in size of firms.[15]

2.5 SUMMARY STATISTICS

Table 4 below shows summary statistics for the period 2003-2007 for some of the key variables, including turnover, labour cost, employment, capital intensity and labour productivity (defined as real value added per employee) by country and sector. Sectors are grouped in tradables and non-tradables (Box 2).[16]

BOX 2 TRADABLE AND NON-TRADABLE SECTORS IN COMPNET

Tradables

- Manufacturing (NACE Section C) with exclusion of "Manufacture of coke and redefined petroleum products" (NACE Industry 19)

Non-tradables

- Construction (NACE Section F)
- Accommodation and food service activities (NACE Section I)
- Information and communication (NACE Section J)
- Professional, Scientific and technical activities (NACE Section M)
- Administrative and support service activities (NACE Section N)

15 According to the exhaustive analysis of Bartelsman et al. (2009), the static indicator of allocative efficiency is more robust in terms of theoretical predictions and measurement problems than the dynamic one, so we have to take with caution the results concerning the latter, also because entry and exit are not identified properly in our samples. In any case, we will report averages over several years which contributes to mitigate the measurement bias.

16 The aggregation is done using a similar set of common sector weights introduced in Box 1.

For each variable of interest (turnover, labour cost, employment, capital intensity and labour productivity) we show the average, the median and the interquartile range (as a proxy of dispersion). As expected, the mean turnover, firm size and capital intensity in tradables is larger than in non-tradables in all countries. Significant differences across countries exist in terms of median firm size, with Estonia, Slovenia, Belgium and Spain featuring the smallest firms, which is partially capturing the existing bias towards large firms in certain country samples (particularly in Germany, Slovak R., Czech. R.). Dispersion is very large, especially within tradables. The mean is above the median in all countries, for all variables, reflecting the skewness of the distribution, characterised by a large group of low value-added, small firms and few champions.

TABLE 4 DESCRIPTIVE STATISTICS[17], AVERAGE OVER THE PERIOD 2003-2007

		Turnover			Labour Cost			Labour			K/L			Labour Productivity		
country		p50	mean	iqr	p50	mean	iqr	p50	mean	iqr	p50	mean	iqr	p50	mean	iqr
BELGIUM	tradables	2996	67514	16925	565	4354	2785	12	67	45	38	87	62	59	70	35
	non-tradables	696	4051	2347	157	1532	602	4	37	13	30	138	75	54	67	36
CZ	tradables	2571	16888	11213	347	1073	907	55	144	124	13	23	21	12	15	9
	non-tradables	1338	4283	2644	230	568	382	34	71	46	10	24	17	14	17	11
ESTONIA	tradables	442	2079	1588	69	257	216	12	37	31	7	13	13	12	16	12
	non-tradables	174	956	433	31	135	89	5	17	11	4	20	11	13	17	15
FRANCE	tradables	1790	28564	6482	510	3165	1477	14	66	37	33	69	53	47	53	26
	non-tradables	789	4293	1861	338	1589	704	10	44	20	19	65	30	48	58	32
GERMANY	tradables	17945	113527	50653	3888	15131	8599	94	279	176	22	95	36	79	92	47
	non-tradables	6365	23614	15854	1911	7663	4780	56	224	144	16	164	39	65	83	49
HUNGARY	tradables	2234	26284	15480	236	1420	938	30	107	83	10	21	18	12	16	11
	non-tradables	1147	5709	3777	109	725	481	15	67	52	13	96	28	18	23	19
ITALY	tradables	3657	10338	6509	640	1398	946	21	39	26	27	46	44	43	48	22
	non-tradables	1939	5541	3262	500	2167	1119	18	97	48	14	45	30	38	42	20
SLOVAKIA	tradables	3417	28358	34772	607	1989	1678	85	222	179	21	35	35	10	13	8
	non-tradables	2187	7777	6094	540	1115	1056	61	119	102	13	24	20	13	15	9
SLOVENIA	tradables	684	9975	9993	147	1807	2321	9	78	73	22	40	35	21	25	14
	non-tradables	311	1525	832	75	370	218	5	22	14	16	45	35	19	23	14
SPAIN	tradables	882	38894	48574	231	1651	1233	9	39	27	17	36	35	35	40	23
	non-tradables	306	2925	778	113	1071	297	5	46	13	11	75	30	31	38	22

17 Turnover and labour costs are expressed in thousands of euros. Capital intensity and labour productivity, in thousands of euros per employee and employment in physical persons.

2.6 DATA COMPARISON WITH EUROSTAT

Validation of the raw data has been undertaken with EUROSTAT Structural Business Statistics.[18] The exercise consists in computing the correlation over time between variables such as total turnover, value added, employment and labour cost aggregated at the country/industry/year level from CompNet with those from official statistics provided by EUROSTAT. To do the comparison, for each country and variable, all sector-years were pooled together and then compared with the corresponding vector, with the same sectors and years, from EUROSTAT. Please note that for some countries we only have overlapping information with EUROSTAT between 2008 and 2010. As a result, in those countries, growth rates' comparisons have to be taken with caution.

TABLE 5 CORRELATIONS COMPNET- EUROSTAT, LEVELS AND GROWTH RATES

	correlation with Eurostat (levels) [1]				correlation with Eurostat (growth rates) [1]				Overlapping years with Eurostat
	Turnover	VA	L	LC	Turnover	VA	L	LC	
BELGIUM [2]	0.77*	0.97*	0.94*	0.98*	0.27*	0.30*	0.16*	0.15*	2008-2010
CZECH REP.	0.96*	0.92*	0.81*	0.97*	0.90*	0.40*	0.92*	0.69*	2005-2010
ESTONIA	0.83*	0.74*	0.84*	0.80*	0.83*	0.52*	0.84*	0.64*	2005-2010
GERMANY [2]	0.96*	0.83*	0.66*	0.83*	0.93*	0.92*	0.87*	0.90*	2008-2010
HUNGARY	0.96*	0.93*	0.52*	0.85*	0.80*	0.71*	0.74*	0.78*	2005-2010
ITALY [2]	0.87*	0.68*	0.69*	0.90*	n/a	n/a	n/a	n/a	2008
POLAND [2]	0.93*	0.77*	0.72*	0.85*	0.83*	0.44*	0.81*	0.76*	2005-2010
SLOVAKIA [2]	0.99*	0.95*	0.67*	0.96*	0.60*	0.62*	0.24*	0.57*	2008-2010
SLOVENIA	0.98*	0.95*	0.90*	0.94*	0.72*	0.58*	0.35*	0.37*	2005-2010
SPAIN [2]	0.77*	0.81*	0.87*	0.92*	0.13*	-0.01	-0.01	0.10*	2008-2010

(*) indicates significance at 5%

[1] Using only overlapping years (shown in the last column) and sectors in Eurostat and Compnet.

[2] Indicates that Eurostat has data for the given country only for 2008-2010. Note that Italy overlaps with Eurostat only in 2008 (hence, no growth rate correlation could be computed) whereas France has no overlapping at all.

[18] Comparisons have been also undertaken with EU KLEMS. EU KLEMS is a sector database of productivity–related variables constructed from the National Accounts and disaggregated to the sector level using harmonised procedures for all countries (see O'Mahony and Timmer 2009 for more information). The advantage of comparing CompNet data with EU KLEMS sector aggregates is that there are more overlapping years than with Eurostat (from the mid-90s to 2007 for 6 out of the 10 available countries). Besides, EU KLEMS is specifically designed to analyse productivity across countries. However, the sector classification of EU KLEMS is different from the one used in CompNet (sector classification is based on NACE rev. 1). For this reason the exercise can be carried out for 7 sectors only (Scientific research and development; Rental and leasing activities; Telecommunications; Manufacture of machinery and equipment nec; Manufacture of basic metals; Manufacture of chemicals and chemical products). Moreover, EU KLEMS does not provide comparable information on labour costs so the exercise can only be done with employment and turnover, neither it provides data for Poland. Due to these limitations we do not show the estimated correlations in the text, although they are available upon request, and very high, around 0.90 in most countries despite the few sectors available to do the exercise.

Correlations between CompNet and EUROSTAT in levels are in general very high, with most values above 0.8 and many above 0.9. Correlations in terms of growth rates are also in general rather high, considering the short time-span we can compare with; notable exceptions are Spain, and to a lesser extent Belgium. However, the low correlations found in Spain are concentrated in the smallest size class (less than 10 employees) whereas correlations in the rest of the size classes are more in line with the rest of the countries, and significant at the 5% level.[19] One possible reason for this mismatch is that EUROSTAT includes self-employed workers in the case of Spanish firms, while the CompNet database does not; and self-employed account for about half of the total number of firms in the smallest size class. As for Belgium, the small correlations in growth rates are partly related to the short overlapping period and to some sectoral outliers.

All in all, the data validation exercise shows that aggregated variables computed from CompNet firm-level data display dynamics which are consistent with those of other aggregated sources.

2.7 COMPARISON WITH AMADEUS

A firm-level cross-country dataset frequently used in the literature is Amadeus, commercially published by Bureau Van Dijk. Amadeus offers firm-level information extracted from firms' balance-sheets on a set of variables for most of the European countries. For this reason, Amadeus has insofar been the closest proxy to a pan-European firm-level dataset that could be used to analyse the micro-foundations of competitiveness. Reliability and actual use of the database, however, is subject to a number of well-known drawbacks. First, although listed in terms of identifier in the data, not all firms provide information about the required variables needed to analyse issues such as productivity or TFP developments. In particular information on employment (typically a non-mandatory item in balance sheets) is generally poor, especially for certain countries. This reduces significantly the available sample for analysis, leading to potential biases across countries. A second drawback, resulting from the fact that firms not reporting data are typically the smallest ones, i.e. selection into the available sample for analysis is not random, is the lack of representativeness in terms of size and sector distribution of some country samples.

As shown in Table 6, the data collected by CompNet rather noticeably improve upon these issues. In particular, Table 6 shows in its first column the firm coverage of CompNet and Amadeus with respect to the population of firms in each country/year. The second column

[19] In Spain, correlations of all variables in growth rates are negative –although close to zero- for firms with less than 10 employees. On the other hand, for the size class 50-249, growth correlations of value added and turnover are about 0.5 and significative. For the size class of large firms, those with more than 250 employees, turnover growth rate correlation is 0.6 and value added growth rate correlation is 0.5.

shows the firm coverage once we require valid data for both value added and employment, the basics to explore productivity-related issues. Lastly, the third column shows the firm coverage available to explore TFP-related issues, that is, the number of firms with data on value added, employment, capital and intermediate inputs.

Take the case of Belgium, for example. Firm coverage in CompNet and Amadeus is quite similar, almost one-quarter of the population of firms. However, the actual number of firms for which information is available to perform productivity and TFP analysis drops in Amadeus from 22.2% to 2.4% and 2.0% respectively, whereas it is still about one-fifth of the total population of firms in CompNet.

TABLE 6 COMPNET AND AMADEUS SAMPLE COVERAGE[20]

	Converage in terms of firms/Eurostat		Coverage in terms of firms with value added and employment data/Eurostat		Coverage in terms of firms with value added, employment, tangible assets and intermdiate inputs data/Eurostat		Year
	Compnet	Amadeus	Compnet	Amadeus	Compnet	Amadeus	
BELGIUM	26.5%	22.2%	19.5%	2.4%	18.4%	2.0%	2008
ESTONIA	65.9%	59.7%	27.3%	21.7%	23.1%	20.5%	2007
FRANCE	30.6%	19.3%	16.1%	9.6%	14.9%	8.1%	2009
GERMANY	3.1%	16.5%	1.7%	0.4%	1.6%	0.3%	2008
HUNGARY	3.6%	3.2%	2.1%	0.9%	1.9%	0.9%	2007
ITALY	2.2%	4.9%	2.0%	4.3%	1.9%	4.1%	2008
POLAND	1.2%	0.7%	0.4%	0.4%	0.4%	0.4%	2007
SLOVAKIA	12.8%	5.4%	9.1%	3.6%	8.7%	3.6%	2008
SLOVENIA	28.4%	0.3%	19.4%	0.0%	16.8%	0.0%	2007
SPAIN	23.6%	6.9%	10.8%	5.9%	9.7%	5.7%	2008

Note: Coverage of CompNet and Amadeus vs. Eurostat is done only for the year specified in the last column.
The only exception is France where data for 2007 in CompNet was compared with the population of firms in 2009, given the lack of overlapping years.

As for the relative representativeness, Table 7 shows the distribution of firms per size and sector in Amadeus and CompNet, compared to the population. Following with the example of Belgium- but the general result applies to all countries with no exception- according to Eurostat only about 0.2% of firms have more than 250 employees; the corresponding share of large firms in Amadeus exceeds 67.3% (0.8% in CompNet).

20 Please note that data for Czech Republic could not be retrieved from Amadeus.

TABLE 7 SECTOR AND SIZE DISTRIBUTION OF FIRMS, COMPNET AND AMADEUS

	Size Distribution									Sector Distribution						Year
	0 to 19 employees			20 to 249 employees			>250 employees			Manufacturing			Non-Manufacturing			
	Compnet	Amadeus	Eurostat	Compnet	Amadeus	Eurostat	Compnet	Amadeus	Eurostat	Compnet	Amadeus	Eurostat	Compnet	Amadeus	Eurostat	
BELGIUM	88.9%	31.1%	96.3%	10.3%	1.6%	3.5%	0.8%	67.3%	0.2%	20.7%	15.5%	14.8%	79.3%	84.5%	85.2%	2008
ESTONIA	86.1%	24.1%	90.0%	13.2%	1.7%	9.5%	0.6%	74.3%	0.5%	24.5%	21.1%	20.6%	75.5%	78.9%	79.4%	2007
FRANCE	86.3%	34.1%	96.4%	12.8%	1.2%	3.4%	0.9%	64.7%	0.2%	19.2%	14.0%	14.9%	80.8%	86.0%	85.1%	2009
GERMANY	24.8%	30.2%	91.8%	61.6%	1.6%	7.6%	13.6%	68.2%	0.6%	52.3%	16.7%	17.0%	47.7%	83.3%	83.0%	2008
HUNGARY	57.6%	7.5%	96.9%	37.9%	7.7%	2.9%	4.5%	84.8%	0.2%	72.8%	38.8%	16.4%	27.2%	61.2%	83.6%	2007
ITALY	56.0%	16.6%	97.4%	43.0%	3.2%	2.5%	1.0%	80.2%	0.1%	58.2%	37.1%	19.5%	41.8%	62.9%	80.5%	2008
POLAND	0.0%	9.6%	93.7%	95.2%	3.0%	5.7%	4.8%	87.3%	0.6%	65.4%	41.7%	45.0%	34.6%	58.3%	55.0%	2007
SLOVAKIA	11.1%	11.4%	87.5%	78.3%	0.0%	11.2%	10.6%	88.6%	1.3%	53.4%	37.3%	27.3%	46.6%	62.7%	72.7%	2008
SLOVENIA	87.9%	29.6%	95.9%	11.1%	2.8%	3.7%	1.0%	67.6%	0.3%	26.0%	33.8%	24.8%	74.0%	66.2%	75.2%	2007
SPAIN	91.5%	13.8%	96.1%	8.1%	2.8%	3.7%	0.4%	83.4%	0.2%	20.3%	30.8%	19.4%	79.7%	69.2%	80.6%	2008

3 DESCRIPTIVE ANALYSIS OF INDUSTRY INDICATORS ACROSS COUNTRIES

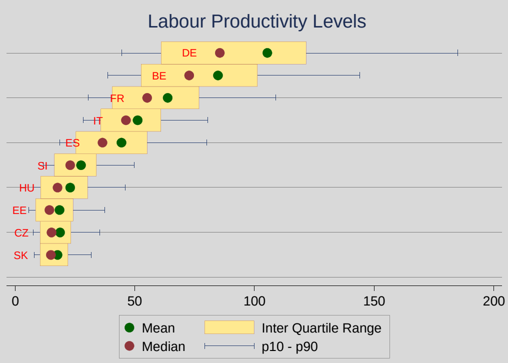
Firm heterogeneity across countries and industries

A compelling reason to use firm-level data as a complement to macro or sector data is that aggregate figures hide the mechanisms underlying firms' response to the policy environment. For example, the response of aggregate unit labour costs (ULC) to collective wage agreements may come through two channels: namely, changes in firm-level productivity and resource reallocation. In these circumstances, as discussed in the Introduction, the extent to which firm productivity will depart from a symmetric distribution will affect the strength of the resource reallocation channel.

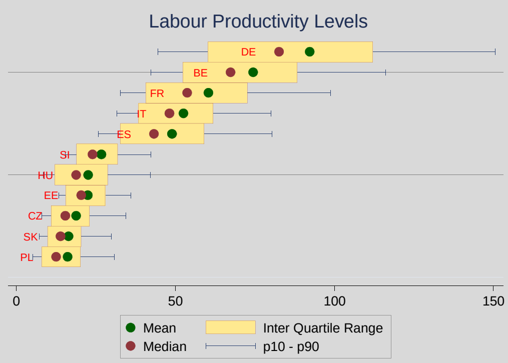
In order to empirically assess the relative importance of the above channels, Figure 1 displays the distribution of labour productivity (calculated as real value added per employee) across countries, averaged over the period 2003-2007 (common to all countries). The left-hand panel of the figure shows distributions using the full sample of firms available in each country, whereas the right-hand panel shows the distributions of productivity across countries using only firms with more than 20 employees, when available (the restricted sample). The latter figure, therefore, uses much more homogeneous samples of firms across countries and should thus be less affected by biases in the distribution potentially affecting our results.

FIGURE 1 LABOUR PRODUCTIVITY DISTRIBUTION ACROSS COUNTRIES, AVERAGE OVER THE PERIOD 2003-2007

a: Full sample



b: Restricted sample of firms



Note: The productivity level for each percentile is computed at a sector-year level within each country. Country-year averages are then computed weighting the percentiles by the common weighting system in Box 1. Then simple averages are computed across years in order to obtain the average distribution of labour productivity for each country. The graph ranks countries according to the p75 of labour productivity. Units are thousands of euros per employee. Recall that the Polish sample has no firms with less than 20 employees and the Slovak sample has no firms with less than 20 employees or less than 5 million euros of turnover.

Although looking at the country mean labour productivity, CompNet firm-level data replicate in both cases well known rankings calculated at the macro (aggregate) level across countries, cross-country comparisons of labour productivity levels have to be done with lots of caution for several reasons. First, labour productivity measures reported above are not expressed in terms of Purchasing Parity Power (PPP), but in country-specific (thousands of) euros. This can be driving some of the country differences in Figure 1. For example, expressing productivity in comparable PPP units for Spain would result in Spanish firms having a similar productivity distribution as that in France. Hence, differences with Germany and Belgium would still be large but not so pronounced. A second caveat is due to the fact that labour productivity differences can be largely driven by differences in capital intensity, which varies widely, above all between the more mature European countries and the Central and Eastern countries. Third, even if the samples of large firms (above 20 employees) are more comparable, there still remain important sample differences that might be affecting rankings in Figure 1.[21]

For all these reasons, Figure 1 is not intended to compare country productivity levels, but rather to point at the large within-country dispersion of productivity as well as the high skewness of the distribution within each country. As such, all the ensuing policy implications in this paper will be derived by looking at the within-country variation of productivity across firms.

In particular, by looking at the shape of the distribution of firm productivity, it is clear that, in each country, the latter is far from being "normal": rather than having many firms centred around an 'average' performance level, with few very bad- or very good-performing ones symmetrically distributed around the mean in equal numbers, data show a large heterogeneity in performance (larger than generally assumed), with many relatively low productive firms, but also a certain number of particularly high productive ones. As a result, median labour productivity is significantly below the mean in every country, while the resulting distribution is characterized by a relatively long right tail (or skewness).

To confirm whether the departure from a normal, symmetric distribution is statistically significant in our data, we have run a t-test on the difference between means and medians of the country-sector-year level distributions, which amounts to verify whether productivity is normally distributed across firms, or is rather asymmetric, thus with higher potential for a

[21] Differences in the representativeness of country samples are, however, not overly distorting the country rankings shown in Figure 1. Using representative samples from Eurostat (we thank Eric Bartelsman for this information), the average productivity of firms in the top quartile of the distribution in Slovenia, for example, was about the same as the productivity of German firms in the second quartile (that is, between the p25 and p50 of the German productivity distribution). In CompNet samples the productivity of the top firms in Slovenia is comparable to German firms in the first quartile. Hence, although Germany has clearly a sample biased to large firms in CompNet, the rough order of magnitude of the rankings shown in Figure 1 seem to correspond to those obtained with representative samples across countries.

reallocation of economic activities towards the most productive firms. It turns out that the two measures are significantly different.[22]

It then follows that in order to understand differences in the average performance across countries, which is what the standard macroeconomic analysis on competitiveness does, we should look at least at two different features of the underlying distribution of productivity: the extent to which the median firm in a given country is more productive than its counterpart in another country; and the extent to which the right tail of the distribution is 'thick' enough to pull away the mean performance from the median, thus triggering resource reallocation potential.

To clarify this point, consider for example the performance of Hungary vs. Slovak Republic in Figure 1a (full sample). The mean labour productivity of Hungarian firms is larger than the one of Slovak ones (24200 euros per employee vs. 23600), but the median firm in Hungary is actually less productive than in Slovak Republic (18600 euros per employee vs. 20500). What explains the higher ranking of Hungary is thus the fact that the density of very productive firms in the country is relatively larger than in Slovak Republic, leading to a compositional effect resulting in a higher average labour productivity.

More in general, the role of productivity dispersion as a significant component of aggregate productivity can be shown theoretically within a standard model of firm heterogeneity and international trade (Melitz and Ottaviano, 2008) encompassing linear demand systems and endogenous markups (see Annex 9 for a formal proof). The latter is also statistically confirmed in our data since productivity dispersion across firms turns out to be highly correlated with the average level of productivity at the industry level (see Box 3). Given that sector-specific characteristics could be driving differences across industries, we have regressed (log) productivity levels in each industry against the within-industry (log) productivity dispersion (standard deviation and coefficient of variation), controlling for specific sector and year effects, finding that they are indeed positively and significantly correlated. Section 5 of the paper, on the other hand, shows an application of the CompNet data which reinforces this line of argument by showing how the industry-level skewness and dispersion of the labour productivity and TFP distribution is correlated to the trade performance of the same sector.

[22] In the economic literature on firm heterogeneity, a convenient and often used parameterization of this asymmetric distribution of productivity is the 'Pareto' distribution. Melitz and Redding (2013) show how the latter allows for elegant closed-form solutions in models of international trade with firm heterogeneity. Altomonte et al. (2011) discuss in details the features of a Pareto distribution vs. normal ones in the data.

BOX 3 THE CORRELATION BETWEEN PRODUCTIVITY DISPERSION AND LEVEL

This is a very simple exercise to check the correlation between within-sector dispersion and sector productivity level, once you control for sector and year dummies. The table below shows the OLS pooled regression of industry-specific (log) labor productivity and TFP, relative to the country average, against the within-sector standard deviation (specifications 1 and 3) and within-sector coefficient of variation (specifications 2 and 4) of the same productivity measures. Sector and year dummies are included to control for possible scale effects and year effects.

VARIABLES	(1) sector LP/country LP	(2) sector LP/country LP	(3) sector TFP/country TFP	(4) sector TFP/country TFP
Relative LP sd	0.503***			
Relative LP Coeff of Variation		0.116***		
Relative TFP sd			0.884***	
Relative TFP Coeff of Variation				0.411***
Constant	0.449***	-0.160***	0.432***	-1.090***
Sector dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Observations	4,776	4,776	4,351	4,351
Adjusted R-squared	0.789	0.621	0.907	0.556

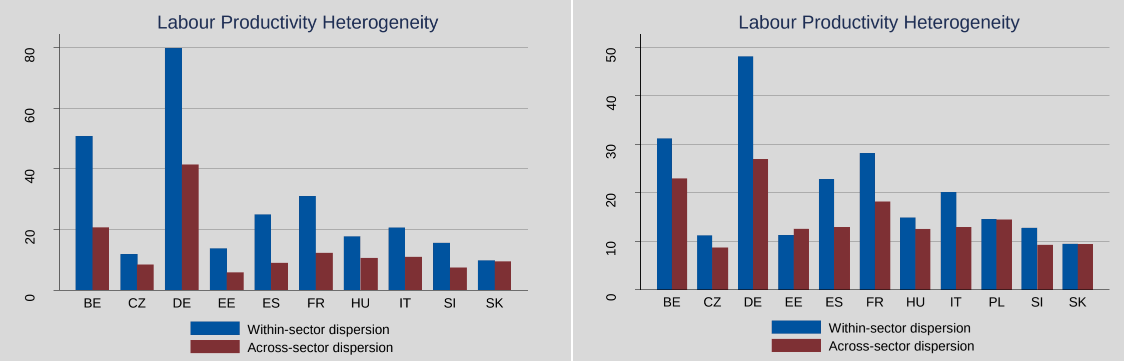
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Consistently with our theoretical priors (see Annex 9), we find that industries with relative high dispersion in terms of productivity, compared with the country average (average of within-sector dispersion in the country), also enjoy relatively higher levels of productivity. This is robust to the use of different measures of dispersion, such as standard deviation or the coefficient of variation, and also to the inclusion of sector and year dummies to control for possible scale effects. This correlation is not driven entirely by differences in capital intensity given that results are very similar if instead of labour productivity the regressions were done with Total Factor Productivity, which takes into account the existing capital stock.

Another (non-parametric) way of considering the importance of dispersion in affecting industry performance can be gauged by looking at Figure 2. It shows how, for most countries, the average difference (here summarized by dispersion of industry mean productivity) in performance across industries is smaller than the average difference in performance across firms within the same industry. The latter implies that the potential gains derived from a reallocation of economic activity from low to high productive firms within any given industry might be at least comparable in terms of magnitude to a change of a country's specialization across industries. Section 4 of the paper will enter in detail in this discussion, capitalising on an already large strand of literature devoted to exploring the role of continuous reallocation for aggregate productivity growth.[23]

FIGURE 2 HETEROGENEITY IN FIRM PERFORMANCE WITHIN AND ACROSS SECTORS. AVERAGE OVER PERIOD 2003-2007

a: Full sample **b: Restricted sample**



Note: Within-sector dispersion is computed as the unweighted average standard deviations of firm-level productivity within each sector, at the country level. Across-sector dispersion is computed as the standard deviation of average industry-specific productivity.

Top vs. bottom firms across countries

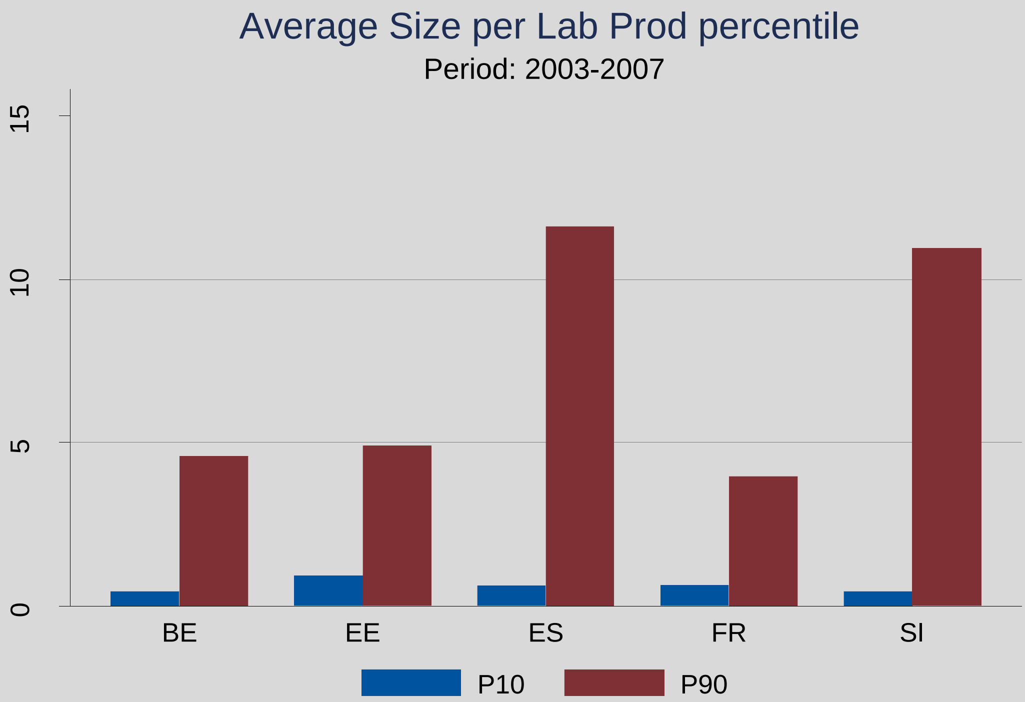
Having shown some of the implications deriving from the high degree of heterogeneity existing across firms within countries and within industries in these countries, it is then interesting to have a closer look at these differences, analysing some of the features of firms residing in different parts (bottom vs. top) of the productivity distribution. In particular, we will be comparing, across countries and over time, the behaviour of firms that are in the first (p10) vs. last (p90) decile of the productivity distribution in their own country and industry.

First of all, Figure 3 shows how, for the five countries with representative samples, the most productive firms are larger (up to ten times larger) in terms of employees than the median (in terms of productivity) firm in the same sector, while the less productive firms are much smaller. Clearly, there are differences across countries both in terms of relative firm size and in terms of distance between the less and the most productive firms, but the finding is remarkably constant across countries. The result is in accordance with some previous empirical analysis (e.g. Mayer and Ottaviano 2007) showing that the most productive firms (those that drive the average productivity of countries, as we have seen) are definitely larger than the average firms. The latter evidence shows the importance for competitiveness of an economic environment in which obstacles to reallocation of resources are removed via the setup of a level playing field in which

[23] See for example Olley and Pakes (1996), Foster et al. 2006, Bartelsman et al. (2004), Bartelsman et al. (2013)

firms might start small (in terms of size / turnover) but are then able to grow exploiting their superior productivity.[24]

FIGURE 3 RELATIVE (TO THE MEDIAN) SIZE OF FIRMS IN EACH PRODUCTIVITY PERCENTILE, 2003-2007. COUNTRIES WITH REPRESENTATIVE SAMPLES



Note: The average firm size in each percentile of the labour productivity distribution within each county/industry/year has been divided by the size of the median firm, in terms of productivity, in that same county/industry/year. The industry-specific ratios have been then aggregated (using the common set of weights) to construct a country-specific relative average size per productivity percentile, each year. An unweighted average across the corresponding years was finally computed.

Figure 4 and 5 below plot instead the evolution over time of labour productivity and unit labour cost, respectively, for different groups of firms over time, namely, the less (p10) and most (p90) productive firms within the sector.[25] We show the dynamics for different groups of countries: (1) Belgium and Germany (BE-DE), (2) Spain and Italy (ES-IT), and (3) Central and Eastern countries (CEE: Estonia, Hungary, Poland, Slovenia and Slovakia). The choice of countries within each group has been driven by data availability for the whole period.[26] We also plot the

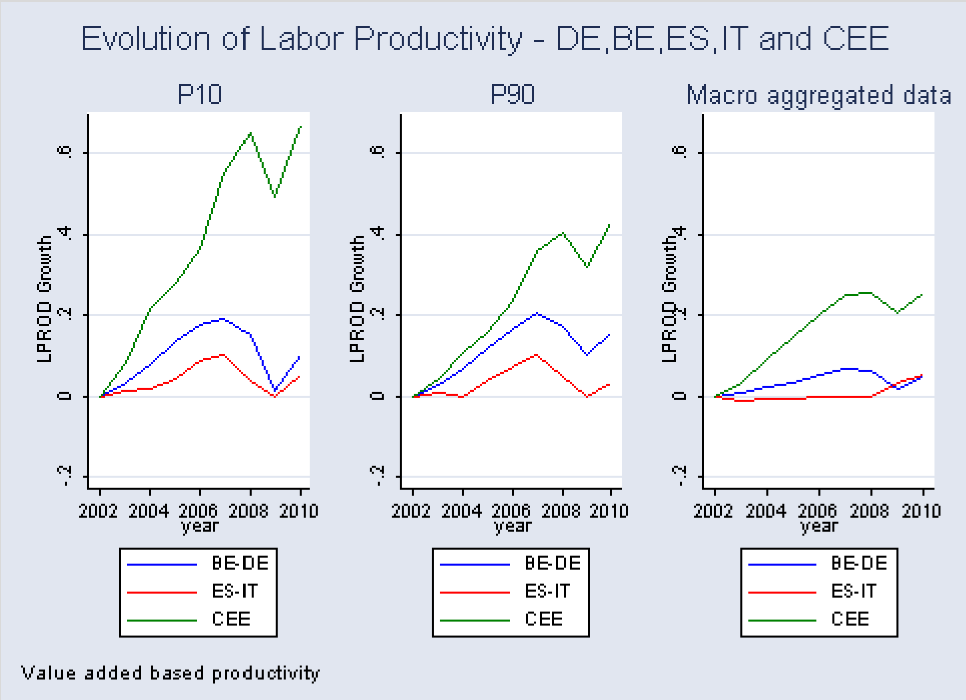
[24] Altomonte et al. (2012) discuss the extent to which firms' growth is conducive to country competitiveness, as well as the firm-specific characteristics that are able to trigger a productivity-enhancing behaviour.

[25] As a robustness check, the dynamics for firms in different percentiles (p25 and p75) have also been analyzed. Results are similar, although, as expected, differences across percentiles are reduced.

[26] Group aggregates are computed as unweighted averages of labour productivity (figure 4) and unit labour cost (figure 5) of the relevant countries.

aggregate figures for labour productivity provided by AMECO (ESA 95) in order to compare the “aggregated message” with that coming from firm-level data.[27] Once again, Figures 4-6 are done for the sample of firms with more than 20 employees to ensure cross-country comparability.[28]

FIGURE 4 LABOUR PRODUCTIVITY EVOLUTION IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, 2002-2010. FIRMS WITH MORE THAN 20 EMPLOYEES.



Note: Growth rates are computed with respect to the base year 2002. The labour productivity corresponding to each percentile is computed at the country/sector/year level. The country labour productivity is obtained using the set of common weights of Box 1. The labour productivity of the country groups are computed as unweighted averages of the relevant countries.

When looking at the graph across country groupings, it is clear that economic convergence and catching-up was taking place in the labour productivity dynamics of Central and Eastern European countries. Also, productivity performance of Germany and Belgium has been generally superior to the one experienced by Spain and Italy.

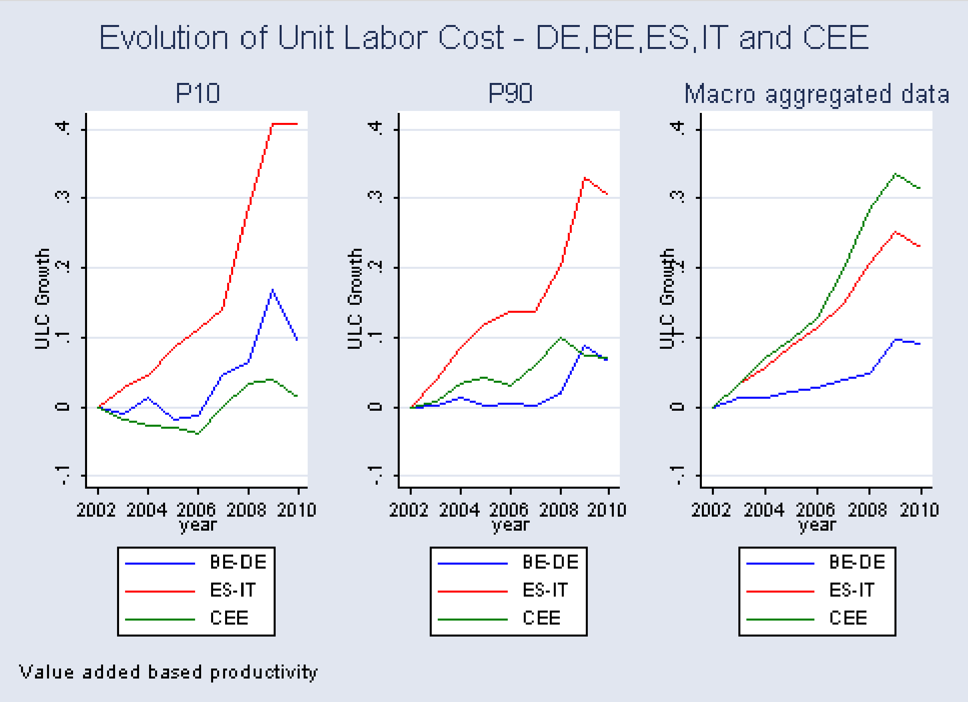
27 Please note that aggregate data from AMECO is provided for the full population of firms whereas figures 4, 5 and 6 show the dynamics of the sample of firms with more than 20 employees in the p10 and p90 of the productivity distribution. This divergence in firm coverage could explain some of the observed differences between aggregate and CompNet figures.

28 In some countries, like Spain, there seems to be a certain decoupling between the Gross Value Added aggregated from available micro data and the one coming from the National Accounts from 2008 onwards, which could be affecting the results. Therefore, as a robustness check, we have replicated Figures 4, 5 and 6 using turnover per employee as our measure of productivity. Results, shown in Annex 5, are virtually the same.

However, when looking at the same graph across different segments of the productivity distribution of firms (p10 vs. p90), more nuanced messages start to appear. In particular, while convergence in productivity seems to be a process valid for all firms in Central and Eastern European countries, that is, irrespective of their relative position in the productivity distribution, in the rest of countries it is the most productive (p90) firms that display the largest dynamics, above all during the crisis period.

In Figure 5 we repeat the same exercise considering unit labour costs (ULC; defined as nominal labour cost per unit of real value added) for firms in both tails of the productivity distribution (p10 and p90), as well as the aggregated developments provided by the European Commission.

FIGURE 5 UNIT LABOUR COST EVOLUTION IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, 2002-2010. SAMPLE OF FIRMS WITH MORE THAN 20 EMPLOYEES.



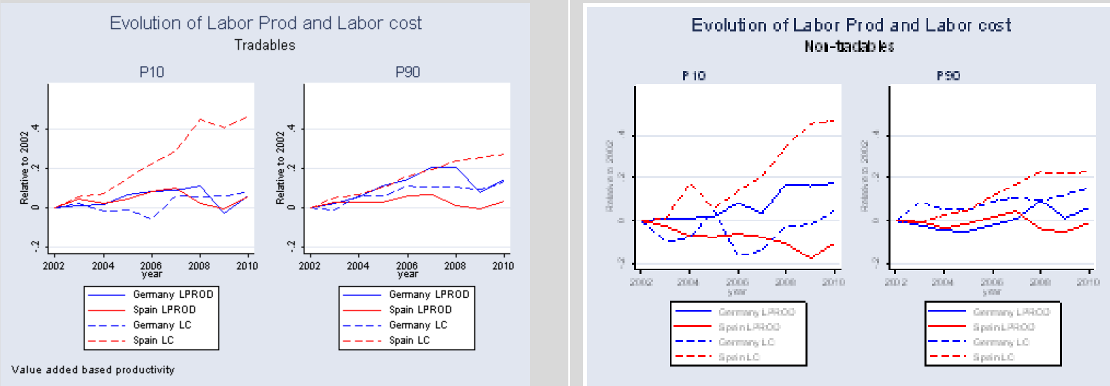
Note: Growth rates are computed with respect to the base year 2002. The ULC corresponding to each percentile is computed at the country/sector/year level. The country ULC is obtained using the set of common weights of Box 1. The ULC of the country groups are computed as unweighted averages of the relevant countries.

Figure 5 shows how firms in different parts of the productivity distribution behave differently. Low productive firms in Spain and Italy have seen their ULC rising during the period of analysis, whereas top productive firms have been somehow better at containing them. Even so, they were not as effective as the firms in Germany and Belgium in controlling costs per unit of output.

Figure 6 provides further insights on Germany and Spain looking at the evolution of the two components of ULC (average labour costs and productivity) in tradable and non-tradable sectors.[29]

FIGURE 6 AVERAGE LABOUR COST AND PRODUCTIVITY IN GERMANY AND SPAIN IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, 2002-2010. RESTRICTED SAMPLE. TRADABLES AND NON-TRADABLES

a: Tradables **b: Non-tradables**



Note: Growth rate computed with respect to the base year 2002. The average cost and labour productivity corresponding to each percentile is computed at the country/sector/year level. An aggregate for each percentile is obtained using the set of common weights of Box 1.

Figure 6 shows clearly, for both tradables and non-tradables, that average cost per employee in Spain differed greatly between low and high productive firms. Until the start of the crisis, low productive Spanish firms lost the ability to compete because of the large and continuous increase in average cost per employee, coupled with a flat productivity performance. Top productive Spanish firms, on the other hand, i.e. those that actually compete in international markets and account for the bulk of Spanish exports, were able to compete in terms of costs with their German counterparts, at least until 2008. Hence, as Antras et al. (2010) show, the so called "Spanish paradox" is a matter of composition bias of the aggregate figures. When one focuses the analysis on those firms that actually export and compete abroad, there is much less of a paradox. Figure 6 shows as well that Spanish firms, across all percentiles and sectors, delivered a flat productivity performance. Although the latter finding might be partly compensated from the fact that some of the positive (but not excessive) cost dynamics of the

[29] The focus on these two countries responds to the existing public debate on the so-called "Spanish paradox", which refers to the fact that during the last decade prices and costs in Spain grew more quickly than in the main developed economies while, at the same time, Spain's export shares did not fall as much. See Antras et al.(2010) and Crespo-Rodriguez and Segura-Cayuela (2013)

most productive firms reflect quality improvements (as also pointed out in Antras et al., 2010), the latter points to a notable weak feature of Spanish firms.

4 ALLOCATIVE EFFICIENCY DIFFERENCES ACROSS COUNTRIES, SECTORS AND PERIODS

Static allocative efficiency

A recent strand of literature, analysing cross-country indicators built up from firm-level data (see Bartelsman et al 2009), shows that cross-country productivity differences partially can be accounted for by differences in allocative efficiency. Aggregate productivity in a country may be lagging partly because available inputs are not allocated efficiently across firms within an industry. This finding provides a potentially new channel for boosting aggregate productivity, namely through reallocation of resources away from poorly performing firms towards the most productive firms.

Although we are still far from fully understanding why allocative efficiency varies across sectors, and countries, there is a growing number of papers relating those observed differences to sector specific regulations of the labour, the product or even the credit market (see for example Restuccia and Rogerson 2008, Arnold et al. 2011, Andrews and Cingano 2012, Aghion et al. 2007 and Martin and Scarpetta 2011).[30] Although identifying the impact of particular regulations or institutions is not an easy task given that many might have more than one effect and others might cancel each other, the fact is that in well-functioning markets resources should flow to more productive plants or firms. Or in other words, there should be a positive correlation between productivity and size at the firm level.

Accordingly, the literature has developed a measure of allocative efficiency by means of the industry-level covariance between productivity and size, a very simple-to-compute and robust indicator[31] first introduced by Olley and Pakes (1996). Olley and Pakes decomposed industry-specific aggregate productivity into the unweighted average labour productivity and a measure of allocative efficiency, as shown in equation 1 (more details in Annex 7):

$$y_{st} = \sum_{i\in s}\theta_{it}\omega_{it} = \overline{\omega_{st}} + \sum_{i\in s}(\theta_{it} - \overline{\theta_{st}})(\omega_{it} - \overline{\omega_{st}}), \qquad (1)$$

where y_{st} is the weighted average productivity of sector s at time t, S is the set of firms belonging to sector s, θ_{it} and $\omega_{i,t}$ represent size and productivity of firm i at time t, respectively, and $\overline{\theta_{st}}$ and $\overline{\omega_{st}}$ represent the unweighted mean size and productivity of industry s at time t, respectively. Hence allocative efficiency is proxied by the covariance

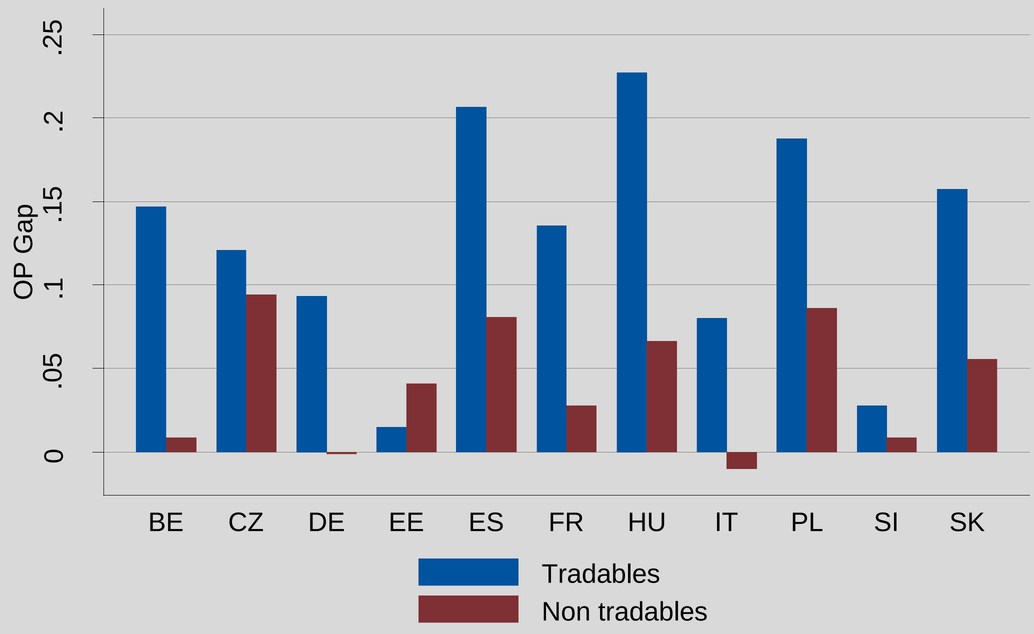
30 Competition can drive productivity through several mechanisms. One of the most important ones is the Darwinian selection among producers with heterogeneous productivity, which moves market share towards more efficient producers, shrinking or forcing the exit of relatively low producers and opening room for more efficient ones. It can also affect entry barriers and incentives of incumbents to undertake within plants improvements aimed at increasing firm-productivity.

31 See Bartelsman et al (2009) for a thorough analysis of different measures of allocative efficiency.

between the relative size of a firm and its relative productivity. That is, if resources were allocated randomly across firms in the industry the covariance measure in the right-hand side of equation (1) would be zero, and aggregate and average productivity would coincide. The larger the covariance, the more efficiently are resources allocated within the sector and the higher the contribution of the (efficient) allocation of resources to the sector productivity, vis-à-vis the unweighted average productivity of the firms operating in the sector.

Figure 7 below presents this indicator of allocative efficiency in each of the CompNet countries, aggregated over the period 2003-2007, distinguishing between tradable and non-tradable sectors (see Box 2 for details on this aggregation). The figure is done with the restricted samples, that is, using only firms with more than 20 employees, in order to ensure as much cross-country comparability as possible.

FIGURE 7 COVARIANCE BETWEEN SIZE AND PRODUCTIVITY, AVERAGE 2003-2007. RESTRICTED SAMPLE



Source: CompNet Dataset

Note: The figure shows the last term on the right-hand side of equation (1). That covariance is first computed at the 2-digit industry level and then aggregated over all tradable and non-tradable sectors in each country. Sector weights are those described in Box 1, but computed separately for tradables and non-tradables so they sum up to one for each of the aggregates. Then, an unweighted average over the years 2003-2007 is computed.

Notice first that numbers are generally quite low, although consistent in terms of ordering of countries, with respect to those shown in Bartelsman et al. (2009). The covariance goes up to 0.2 in Hungary or Spain, which means that sector labour productivity is up to 20 log points larger --20% more-- than it would be with randomly allocated labour. Hence, from an accounting perspective, it is clear that the contribution of the unweighted average productivity of the firms operating in the sector to aggregate productivity is larger than that of the covariance term. Or in other words, the contribution of the allocation of resources to overall productivity is lower than the contribution of the average productivity of the existing firms (10% vs. 90% approximately in our data). However, recent theoretical (Acemoglu et al. 2013) and empirical (Andrews and Cingano 2012) contributions, show that the incentive for firms to increase their own productivity goes up when resources flow easily to the best firms. This means that the direct contribution from better allocation is boosted through its indirect effect on within-firm productivity growth.

Secondly, it is striking the large within-country differences in terms of allocative efficiency in tradable vs. non-tradable sectors, which roughly correspond to manufacturing and services (the covariance in tradables is about 3 times as large, in average, as the non-tradables one). With the exception of Estonia, allocative efficiency is larger in tradables in all countries. Allocative efficiency is actually zero or even negative in the non-tradable sectors of Germany and Italy, which means that the allocation of resources is in fact not contributing at all to the productivity level of non-tradable sectors in those two countries.[32] This large difference in terms of allocative efficiency between tradables and non-tradables is not surprising, and it has also been found in other works like Arnold et al (2008) or European Commission (2013). They might reflect the fact that regulatory reforms in non-tradables have been more hesitant, above all in mature European countries, and that these sectors are sheltered from competition.

Cross-country differences in allocative efficiency are also remarkable. Hungary, Poland and Slovakia, among the Central Eastern countries, and Spain within the mature continental countries, feature the largest covariance between size and productivity, that is, the highest allocative efficiency both in tradables and non-tradables. The high covariance found in some Central and Eastern countries could be reflecting the large foreign direct investments undertaken by western countries which have created some sort of duality between very large and productive foreign-owned firms and small, low productive local firms. However, this pattern could suggest that in some Central and Eastern countries there no longer is much scope for boosting further aggregate productivity through reallocation, which had been a main driver

[32] The non-tradable results for Germany have to be taken with caution given the relatively small sample of non-trading firms.

during the transition period (see Bartelsman et al 2009). Instead, much may be gained by having firms catch up to the global frontier of that industry. In any case, this is a topic that would deserve further research.

As for Spain, the high covariance term shows that the most efficient firms may be large but recall that they are not as productive as the best firms in the world (as shown in Figure 1). The figure also shows that non-tradable sectors' productivity in Germany, Belgium or Italy could be further improved by reallocating better existing resources within sectors. It is beyond the scope of this paper to discuss what policies could be put in place to achieve best this objective, although an increasing number of papers on the topic are pointing to labour and product market regulations as well as financial development or a fair and predictable legal system (see Arnold et al 2008 and Haltiwanger 2011 and the references therein).

Box 4 shows that improving allocative efficiency can indeed have a large impact on sector productivity growth. The estimated elasticity, controlling for sector and year-country dummies, reflects that 1% increase in the covariance term, at the sector level, could increase sector productivity by 0.69%. Box 4 explores this issue further by allowing the productivity benefit from better resource allocation within a sector to vary with the distance to the productivity frontier of the sector, finding that the further to the productivity frontier, the more the sector can benefit from a better allocation of resources.

BOX 4 ESTIMATING THE ELASTICITY OF PRODUCTIVITY TO ALLOCATIVE EFFICIENCY

We perform a very simple exercise to estimate the elasticity of the sector productivity to the covariance term in equation (1). That is, we want to estimate what is the impact on sector productivity growth of a change in the covariance or allocative efficiency of the sector. We also test whether the elasticity is different depending on the distance of the sector to the productivity frontier. We run a pooled OLS of sector productivity growth against the contemporaneous growth in the sector covariance term, in specification (1). Specification (2) adds the interaction between the covariance growth and the (lagged) distance to the productivity frontier of the sector. The distance to the frontier is measured as the ratio of the average productivity of the two most productive countries, in that sector that year, to the sector productivity. More specifically: $\Delta y_{st} = \alpha + \beta \Delta cov(y,\theta)_{st} + \gamma Distance_{s,t-1} + \delta \Delta cov(y,\theta)_{st} \times Distance_{s,t-1} + d_t \times d_c + d_s + u_{st}$ where the change in log productivity at the sector level is related to the contemporaneous change in the sector covariance term (results are robust to including lags), the distance of the sector to the productivity frontier, lagged one period, and the interaction between the change in covariance and the distance to the frontier. Country*year dummies are included to control for country-year specific events as well as sector dummies.

Results confirm that improving resource allocation within the sector can have an important impact on productivity growth, given that the estimated elasticity is 0.69 (that is, a 1% increase in the covariance increases productivity of the sector by 0.69%). This is so even after controlling for the distance to the productivity frontier of the country (although the covariance-productivity elasticity decreases), which as expected is positively related to the sector productivity growth. Moreover, the positive coefficient of the interaction between the covariance and the distance to the frontier confirms that the productivity gains from better resource allocation are larger for sectors further away from the productivity frontier.

VARIABLES	(1) Growth of sector Labour productivity	(2) Growth of sector Labour productivity
Covariance growth	0.6943*** (0.0495)	0.3882*** (0.1280)
Distance to the frontier$_{t-1}$		0.1915*** (0.0095)
Distance to the frontier$_{t-1}$ ^{X} Covariance growth		0.0503* (0.0281)
Constant	-0.0763 (0.0693)	-0.9452*** (0.0788)
Year*country dummies	Yes	Yes
Sector dummies	Yes	Yes
Observations	4,149	4,149
Number of clusters	594	594

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sector productivity growth rate measured as the difference in logs of the sector labour productivity in two periods. Distance to the frontier measured as the log of the ratio of the unweighted average of the sector productivity in the two most productive countries that year, which are considered the frontier, to the sector productivity in a given country and year.

Dynamic allocative efficiency

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. Before looking at this dynamic indicator of allocative efficiency, please recall that we do not have information on the age of firms, nor we have well measured data on entry and exit of firms. That implies that we cannot distinguish between mature and young businesses, nor can we measure the effect of reallocation through the firm entry and exit margin.

Given this limitation of our database, we start from a traditional aggregate productivity decomposition à la Foster et al. (2006) and identify three sources of productivity growth. The first is the contribution of the within-firm productivity growth, the second one is the contribution of the reallocation of resources between established firms –it is positive if firms with higher productivity than the average are gaining market share- and the last one is the covariance term which captures simultaneous, at the firm-level, growth in productivity and market share (See Annex 7 for details):

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

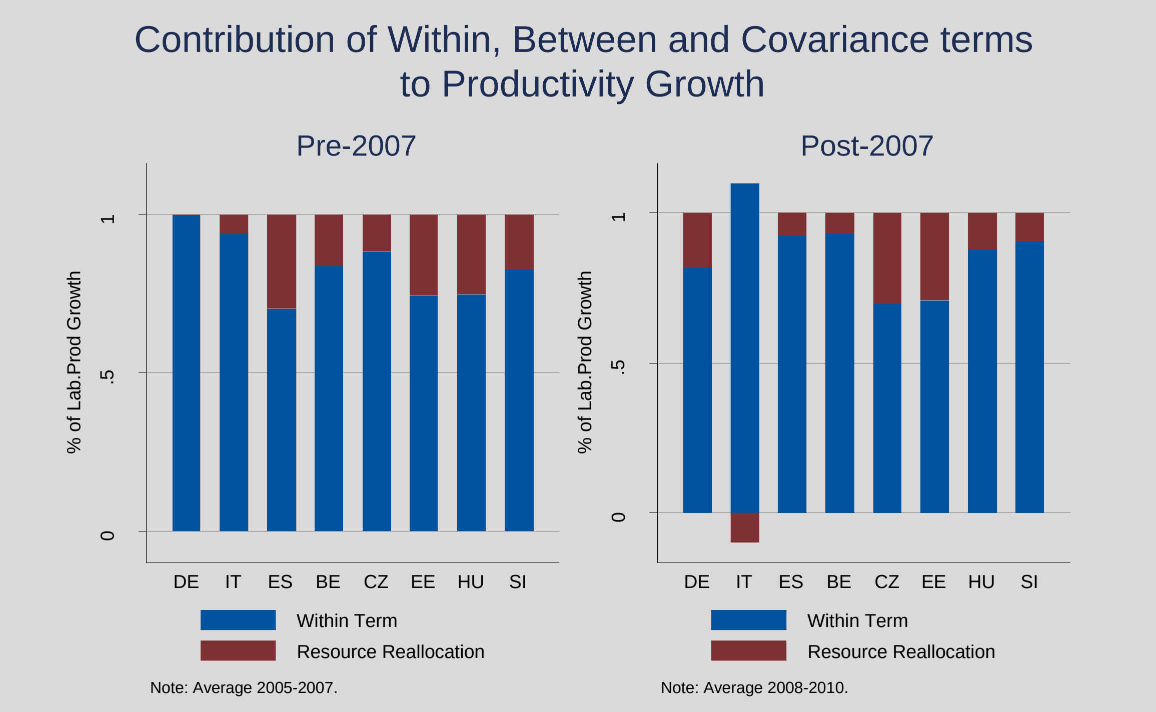
where Δ is the differential operator between t-k and t and y_{st} is the sector productivity as resulting from the sum of the three terms above mentioned, in logs[33]; C denotes continuing firms; θ_{it} and $\omega_{i,t}$ represent size and productivity of firm i at time t respectively and $\widehat{\theta_{st}}$ and $\widehat{\omega_{st}}$ represent the weighted mean size and productivity of industry s at time t respectively. The first term of the decomposition is the contribution of productivity changes for continuing firms with initial weights ("within component"); the second term is the effect of reallocating resources among continuing firms given their initial productivity ("between component") and the third term is the cross-effect of reallocation and productivity changes for continuing firms ("covariance or cross component").

Figure 8 below shows the contribution to sector productivity growth of the "within term" on the one hand, and the sum of the "between" and the "covariance term" on the other hand, which captures the overall contribution of reallocation of resources to productivity growth. Sector-specific contributions of each of the terms are aggregated (using the common set of sector weights explained in Box 1) to compute the country averages. The period 2005-2007, that is,

[33] Note how the within-sample measure of sector productivity used in the decomposition is different from aggregate sector productivity due to the lack of the entry and exit margin.

before the EU debt crisis, and 2008-2010 are shown separately in order to shed some light on the possible impact of the crisis in the measures.[34] Note that the full samples are used to perform the accounting decomposition given that in the restricted samples (firms with more than 20 employees) firm dynamics is quite restricted.

FIGURE 8 CONTRIBUTION OF THE WITHIN PRODUCTIVITY GROWTH AND THE REALLOCATION OF RESOURCES TO PRODUCTIVITY GROWTH, BEFORE AND DURING THE CRISIS. FULL SAMPLE, CONTINUING FIRMS.



Source: CompNet Dataset

Notes: The blue bar is the first term in the right-hand size of equation 2, divided by the sector productivity growth or left-hand term (to show the contribution of the term to the overall sector productivity growth). This ratio is then aggregated using common sector weights to show the country number. The red bar is the sum of the two last terms in the right-hand side of equation 2, again divided by the sector productivity growth or left-hand term, and then aggregated to the country level. Finally, the unweighted average across the corresponding years is taken. Only countries with full sample (that is, Poland and Slovakia could not be included). France did not have data from 2007 on so it was also excluded from the graph.

The most interesting dynamics take place in Germany, Italy and Spain. That is why they have been placed first. The other countries are then shown by alphabetical order. In Germany, the contribution of reallocation to productivity growth before the crisis was approximately zero, turning to be positive during the period post-2007. In Spain and Italy, in contrast, reallocation

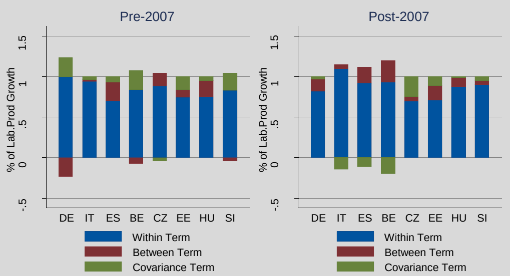
[34] Only countries with data required to do the decomposition on both periods are shown, which excludes France and Slovakia.

had a positive contribution to productivity growth before the crisis, especially in Spain, but it dropped to zero, and even to negative numbers in Italy, during the crisis period. This phenomenon, the reduction of the contribution of the reallocation of resources to productivity growth during the crisis period, has actually taken place in five out of the eight countries shown.

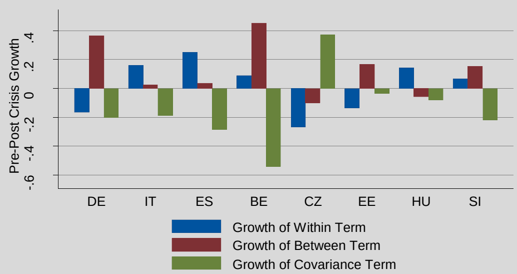
In order to further explore this issue, we show in the next graph the contribution of the covariance and the between term to productivity growth separately, as well as the *change* in the contribution to productivity growth of each of the three terms –within, between and covariance- between the period pre-2007 and the period post-2007.

FIGURE 9 CONTRIBUTION TO PRODUCTIVITY GROWTH OF THE WITHIN, THE BETWEEN AND THE COVARIANCE TERM: PRE-2007 VERSUS POST-2007. FULL SAMPLE, CONTINUING FIRMS.

Contribution of Within and Reallocation Terms to Productivity Growth



Change in the Contribution of Within, Between and Covariance terms to Productivity Growth



Note: Average 2005-2007. Note: Average 2008-2010. Note: Average 2005-2010.

Source: CompNet Dataset

In all countries but the Czech Republic the covariance term drops significantly during the crisis, turning negative in Italy, Spain and Belgium, and about zero in Germany, Slovenia and Hungary. Recall that the covariance term is capturing the simultaneous change in productivity and size at the firm level; a negative productivity covariance could reflect the gain in productivity due to downsizing. On the other hand, the increase in the contribution to productivity growth of the reallocation of resources across firms, the between term, is remarkable in Belgium and Germany, above all if compared to the two southern European countries in this analysis, Spain and Italy. In those two countries the contribution to productivity growth of the transfer of labour from least to the most productive firms drops during the crisis (and becomes virtually zero in tradables[35]). This latter development could be driven by the labour market institutions in both countries, which has led to a decoupling between job destruction and creation during the crisis (see Caballero and Hammour 1996) and/or to the fact that the credit crunch of the last years has hit harder those countries. It is obviously not within the scope of this paper to provide well-grounded explanations to this phenomenon; still, the framework of analysis provided yields a good basis for further investigation.

[35] In non-tradables, on the other hand, the story is the same for Italy although a bit different for Spain where reallocation is still productivity enhancing during the crisis, which could be due to the cleansing effect of the recession in the construction sector.

5 AN APPLICATION OF COMPNET DATA: RELATING AGGREGATE EXPORTS TO FIRM LEVEL EFFICIENCY

This section describes an example of how information on firm heterogeneity can help explain critical outcomes, such as trade performance across countries. This exercise is deeply inspired by the the most recent stream of the international trade literature showing that trade performance depends on firm characteristics, in particular on the characteristics of top performing firms (Melitz, 2003; Mayer and Ottaviano, 2007; Altomonte et al., 2011).

To this purpose, we provide here a preliminary regression analysis where we relate internationalization performance to dispersion's indicators of (lagged) labour productivity and percentiles of the labour productivity distribution. As shown in Annex 8 the results presented here hold also when using TFP instead of labour productivity.

More precisely, we estimate at the country-sector-year level the following equation:

$$Trade_{jct} = a_0 + a_1 D[Prod(m)]_{jct-2} + a_2 SectVA_{jct-2} + a_2 AverEmpl_{jct-2} + S_j + C_c + T_t + ST_{jt} + CT_{jt} + \varepsilon_{jct}\ , \qquad (3)$$

where Trade_{jct} is a measure of internationalization of sector j, in country c, at year t. The explanatory variables of interest are included in D(Prod(m)) and will be measures of productivity dispersion like the standard deviation, the interquartile range and the skewness and the single percentiles of the productivity distribution. We add some controls. The log of value added (SectVa) takes into account scale effects at the sectoral level; average firm size within each sector/country/year is proxied by Aver.Empl. These controls are two-period lagged to minimize simultaneity biases. We also consider a large set of dummy variables to minimize concerns related to omitted variable bias. Sector, country, and year fixed effects (S_j, C_c, and Tt, respectively) controls for systematic differences in country and sector characteristics (i.e., institutional environment), and business cycles, respectively. Then, sector*year (ST_{jt}) and country*year dummies (CT_{ct}) aim at controlling for unobserved shocks, which are country-year, or sector-year specific (i.e., time variant). Finally, ε is the usual *i.i.d.* error term.

The estimation sample includes 11 CompNet countries[36] from 1996 to 2011. We restrict to manufacturing sectors at two digit level (NACE rev.2) and exclude Tobacco (12), Printing and

[36] Belgium, Czech Republic, Estonia, France, Germany, Hungary, Italy, Romania, Slovenia, Slovakia, Spain. Poland is excluded because its sample has no firms under 20 employees.

publishing (18) and Coke and Petroleum (19). Our working sample is an unbalanced panel with 2,382 observations.

The CompNet database is then merged with Eurostat ComExt data, which provides information on trade flows of each CompNet country at sector-year level. In particular, we focus on two trade variables: the real value of exports (nominal exports are deflated with the relative consumer price index; reference year 2005) and a Balassa index measured as the share of total exports accounted by sector s in country c at time t over the same sectoral share at the same year t in the aggregated 11 CompNet countries. A country is considered to have a comparative advantage in a specific sector j if the index is larger than 1. Additional controls are collected from Eurostat or from the CompNet data base. All the variables are transformed in logs.

Results

Table 8 shows the estimation results when using first and higher moments of the productivity distributions. In columns (1) to (4), the dependent variable is the log of exports, in columns (5) to (8) the dependent variable is the Balassa index.

In the first column we show that trade performance is, as expected, positively related to the average size of a country-sector (in terms of value added) and average (unweighted) productivity. When we add higher moments of the productivity distribution the results interestingly change. In line with theoretical predictions, we find that the level of exports is always positively correlated with all the measures of two year-lagged productivity dispersion[37]: standard deviation (column 2), skewness (column 3) and intequartile range (column 3). Since these results hold also when dispersion measures are normalized by average productivity, we conclude that trade outcomes are positively related to the distance of individual observations from the mean, independently from the level of average productivity. Since skewness captures how far the distribution is biased towards higher productivity levels and how fat is the upper tail of the distribution, we confirm that trade outcomes are better when high productivity firms are more productive and have a higher frequency in the distribution.

As shown in columns 5-8, all this evidence extends to the use of a Balassa index as a dependent variable.

Table 9 directly addresses the issue of top performing firms by looking at specific percentiles of the distribution. Consistently with the results of Table 8, we find that the level of productivity at the top of the distribution, in particular the 99^{th,} the 90th and the 75th percentile are strongly

[37] The choice of lag does not affect the results. Also contemporaneous values of productivity dispersions are positively correlated with trade.

positively correlated with export flows and the Balassa index for each country-sector-year triple. Again the same holds with TFP in place of labour productivity.

Though preliminary, the analysis presented in this section shows how relevant can be the knowledge of the shape of productivity distributions (beyond its average) for understanding trade patterns. The main implication of this result is that two countries with an identical average TFP level may have very different trade performances depending on the right tail of the distribution.

TABLE 8 TRADE AND LABOUR PRODUCTIVITY DISPERSION

	1 Exp.	2 Exp.	3 Exp.	4 Exp.	5 Bal.Index	6 Bal.Index	7 Bal.Index	8 Bal.Index
$Lprod_{t-2}$	.1507**	-.4166***	.1984***	-.1086	.1365*	-.3488***	.1655**	-.2121
	(.0692)	(.1177)	(.0689)	(.1295)	(.0766)	(.1289)	(.077)	(.1493)
$VASect_{t-2}$	.7904***	.776***	.7853***	.7846***	.8702***	.8579***	.8671***	.8623***
	(.03)	(.0301)	(.0299)	(.0303)	(.0414)	(.0416)	(.0415)	(.0417)
$Aver.Empl_{t-2}$	.017	.0264	.012	.0241	-.1432**	-.1351**	-.1462***	-.1335**
	(.0465)	(.0459)	(.0463)	(.0465)	(.0559)	(.0552)	(.0556)	(.0561)
$Lprod(s.d.)_{t-2}$		.4794***				.41***		
		(.0744)				(.0801)		
$Lprod(skew)_{t-2}$			.1631***				.0995*	
			(.0417)				(.0525)	
$Lprod(iqr)_{t-2}$				.2078***				.2792***
				(.077)				(.0901)
Obs	1778	1778	1778	1778	1778	1778	1778	1778
R2	.9573	.9587	.9577	.9575	.588	.5946	.589	.5904

All variables are in logs and lagged of two years. Each columns represent a different equation. Exp: export value. Bal.Index: Balassa Index. Estimation unit is defined by the triple country sector year. Robust standard errors are in parenthesis. We include the following groups of dummies: year, country, sector, country X year, and sector X year. Significance level: * significant at 10%; ** significant at 5%; ***significant at 1%.

TABLE 9 TRADE AND PERCENTILES OF LABOUR PRODUCTIVITY DISTRIBUTION

	1 Exp.	2 Exp.	3 Exp.	4 Exp.	5 Bal.Index	6 Bal.Index	7 Bal.Index	8 Bal.Index
$LProd(pc50)_{t-2}$	.0512	-.1865	-.2176	-.0745	.0872	.0543	-.0254	.0102
	(.0748)	(.2413)	(.1417)	(.0901)	(.0803)	(.2615)	(.1557)	(.0976)
$VASect_{t-2}$	.791***	.7861***	.7792***	.7785***	.8714***	.8671***	.8633***	.8626***
	(.03)	(.0303)	(.03)	(.0303)	(.0416)	(.0419)	(.0419)	(.0421)
$Aver.Empl_{t-2}$	.0103	.0228	.0413	.018	-.1482***	-.1371**	-.1256**	-.1427**
	(.0467)	(.0467)	(.0465)	(.046)	(.056)	(.056)	(.0562)	(.0558)
$LProd(pc25)_{t-2}$		-.0691				-.1745		
		(.1658)				(.1806)		
$LProd(pc75)_{t-2}$		.2869**				.1729		
		(.1462)				(.1613)		
$LProd(pc10)_{t-2}$			-.0858				-.1289	
			(.0998)				(.112)	
$LProd(pc90)_{t-2}$			.3179***				.1958**	
			(.0813)				(.0914)	
$LProd(pc1)_{t-2}$				-.052				-.0478
				(.0576)				(.0691)
$LProd(pc99)_{t-2}$				.193***				.127***
				(.0434)				(.0477)
Obs	1778	1778	1778	1778	1778	1778	1778	1778
R2	.9572	.9573	.9577	.9579	.5875	.5882	.5894	.5897

All variables are in logs and lagged of two years. Each columns represent a different equation. Exp: export value. Bal.Index: Balassa Index. Pc:percentile. Estimation unit is defined by the triple country sector year. Robust standard errors are in parenthesis. We include the following groups of dummies: year, country, sector, country X year, and sector X year. Significance level: * significant at 10%; ** significant at 5%; ***significant at 1%

6 CONCLUSIONS

This paper documents the construction of the CompNet database, an ambitious project of collection of harmonised firm-level based indicators for 11 EU countries and 58 sectors over the period 1995-2011, aimed at analysing the micro-foundations of competitiveness. The paper presents some preliminary findings, which illustrates the importance of firm-level information in assessing country competitiveness. Moreover, it provides an application of the newly developed dataset to explain cross-country differences in export performance.

With respect to existing databases at the sector level (e.g. Eurostat), CompNet has very rich information not only for what concerns sector and country averages, but also with respect to firm level performance distribution across several interrelated dimensions – productivity, size, sectors, employment, labour costs. Knowing more about the underlying firm level distribution helps to ascertain the scope for increasing aggregate productivity via resources reallocation across and within sectors, and to eventually design policies, which foster such reallocation.

We show that there is large heterogeneity in terms of firm performance within narrowly defined sectors, even more than across sector. And this is a fact in most countries. Moreover, firm performance is not distributed following a “Normal probability distribution”, but rather highly skewed, with a large number of underperforming firms and few high-performance firms within each sector. In this type of distributions, the average firm is not representative of the underlying population of firms. Hence, acknowledging this heterogeneity and skewness is fundamental from a policy point of view given that firms at both extremes of the distribution follow different dynamics, and might be differently affected by aggregate shocks or policies. One clear example is given in our analysis of the unit labour cost developments in Spain and Germany. We show that, in Spain, while top productive – exporting – firms were able to contain their average labour costs, the opposite occurred for the least productive, smaller firms in the pre-crisis period. It is likely that such different behaviour across productivity segments lies behind the so-called “Spanish paradox” whereby developments of cost-competitiveness when measured with economy-wide aggregated indicators appear to be unrelated to trade performance.

The importance of considering the dispersion rather than merely the average of the productivity distribution is confirmed when we use our dataset to explain sector export performance across Europe. In sectors where the labor productivity distribution is more highly skewed (fatter and longer right tail), it appears more likely to observe higher volume of exports. This would support the claim that the performance of the top percentiles drives aggregate trade outcomes, independently of the average productivity level.

Although the link between firm heterogeneity and aggregate productivity is complex, it is likely to be associated to the extent to which resources (i.e. workers and capital) can be reallocated within the sector from relatively less to relatively more productive firms, what it is called "allocative efficiency". We show that one of the standard measures of allocative efficiency, the covariance between relative productivity and relative size of firms in a given sector, varies widely across sectors (grouped in tradable and non-tradable) as well as across countries. Although resource allocation contributes relatively less than within-firm productivity performance to sector aggregate productivity, we find that it offers great potential from a policy point of view, especially in the non-tradable sectors.

REFERENCES

Acemoglu, D., Akcigit, U., Bloom, N. and W. R. Kerr (2013): “Innovation, Reallocation and Growth,” NBER Working Paper No. 18993.

Ackerberg, D., Caves, K. and F. Garth (2006): "Structural identification of production functions," MPRA Paper 38349, University Library of Munich, Germany.

Aghion, P., Fally, T. and S. Scarpetta (2007): “Credit constraints as a barrier to the entry and post-entry growth of firms,” Economic Policy (October): 731-779

Altomonte, C., Barba Navaretti, G., di Mauro, F., and G. Ottaviano (2011): “Assessing Competitiveness: how firm level data can help”, Bruegel Policy Contribution, Issue 2011/16, November.

Altomonte, C., Aquilante, T. and G. Ottaviano (2012): “The triggers of competitiveness: The EFIGE cross-country report,” Bruegel blueprint

Altomonte, C. and T. Aquilante (2012): “The EU-EFIGE/BRUEGEL-UNICREDIT dataset” Bruegel Working Paper October 2012.

Andrews, D. and F. Cingano (2012): "Public Policy and Resource Allocation: Evidence from Firms in OECD Countries," OECD Economics Department Working Papers 996, OECD Publishing.

Antras, P., Segura-Cayuela, R. and D. Rodriguez-Rodriguez (2010): “Firms in international trade, with an applicaiton to Spain,” In *SERIEs* invited lecture at the XXXV Simposio de la Asociación Espanola de Economía.

Arnold, J.M., Nicoletti, G. and S.Scarpetta (2011): “Does anti-competitive regulation matter for productivity? Evidence from European firms,” IZA DP N. 5511

Bartelsman, E.J. and M. Doms (2000): “Understanding Productivity: Lessons from Longitudinal Microdata,” *Journal of Economic Literature*, vol. 38: 569-594.

Bartelsman, E, Haltiwanger, J. and S. Scarpetta (2004): "Microeconomic evidence of creative destruction in industrial and developing countries." The World Bank, Policy Research Working Paper No. 3464, December.

Bartelsman, E., Haltiwanger, J. and S. Scarpetta (2009): “Measuring and analyzing cross-country differences in firm dynamics,” in Producer dynamics: New evidence from micro data, eds(Dunne, Bradford and Roberts), University of Chicago Press.

Bartelsman, E., Haltiwanger, J. and S. Scarpetta (2013): “Corss-country differences in productivity: the role of allocation and selection”, American Econoic Review 103(1): 305-334

Bugamelli, M., Barba Navaretti, G., Castellani, D. and E. Forlani (2013): Mimeo

Caballero, R.J. and M.L. Hammour (1996): “"On the Timing and Efficiency of Creative Destruction," The Quarterly Journal of Economics, vol. 111(3): 805-52

Caves, R.E. (1998): “Industrial Organization and New Findings on the Turnover and Mobility of Firms,” Journal of Economic Literature, vol. 36: 1947-1982.

Crespo-Rodríguez, A. and R. Segura-Rayuela (2013): "Understanding competitiveness," Mimeo.

Draghi, M. (2012). "Competitiveness: the key to balanced growth in monetary union" within the conference "A European strategy for growth and integration with solidarity", organised by the Directorate General of the Treasury, Ministry of Economy and Finance –Ministry for Foreign Trade, Paris, 30 November.

European Commission (2013): "Business churn, sectoral performance, and economic policy", Chapter 1 in Product Market Review, Directorate General Economic and Financial Affairs.

Fernandez, C. and P. Lopez-Garcia (2013): "Why are small firms less productives? An analysis using Spanish, French and Italian firm-level data. Mimeo

Foster, L., Haltiwanger, J and C. J. Krizan, (2006): "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s," The Review of Economics and Statistics 88(4): 748-758

Foster, L, Haltiwanger, J. and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" The American Economic Review 98(1): 394-425

Galuscak, K. and L. Lizal (2011): "The Impact of Capital Measurement Error Correction on Firm-Level Production Function Estimation," Working Papers 2011/09, Czech National Bank, Research Department.

Griliches, Z, and J. Mairesse (1995): "Production Functions: The Search for Identification," NBER Working Paper No. 5067.

Haltiwanger, J. (2011): "Firm dynamics and productivity growth," EIB Papers 5/2011, European Investment Bank, Economics Department.

Levinsohn, J. and a. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," Review of Economic Studies Vol. 70 (2): 317-342

Levinsohn, J. and a. Petrin (2005): "Measuring and Mismeasuring Industry Productivity Growth Using Plant-Level Data," Society for Economic Dynamics.

Martin, J.P. and S. Scarpetta (2011): "Setting it right: Employment protection, labour reallocation and productivity," IZA Policy Paper 27.

Mayer, T. and Ottaviano, G. (2008): "The Happy Few: The Internationalisation of European Firms," *Intereconomics: Review of European Economic Policy*, vol. 43(3): 135-148.

Melitz, M.J. (2012): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, vol. 71(6): 1695-1725.

Melitz, M.J. and G. I. Ottaviano (2008): "Market Size, Trade, and Productivity," Review of Economic Studies, vol. 75(1: 295-316.

Melitz, M.J. and S. Redding (2013): "Heterogeneous Firms and Trade," CEPR Discussion Papers 9317.

Olley, G.S. and A. Pakes (1996): "The dynamics of productivity in the telecommunications equipment industry," Econometrica 64(6): 1263-1297

O'Mahony, M. and M.P. Timmer (2009): "Output, input and productivity measures at the industry level: The EU KLEMS database," The Economic Journal 119(June): F374-F403

Petrin, A. and J. Levinsohn (2012): "Measuring Aggregate Productivity Growth Using Plant-Level Data," RAND Journal of Economics 43(4): 705-725.

Restuccia, D. and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," Review of Economic Dynamics 11: 707-720

Wooldridge, J.M (2009): "On estimating firm-level production functions using proxy variables to control for unobservables," Economics Letters, 104(3):112-114

ANNEX 1: COUNTRY-SPECIFIC NOTES

BELGIUM

Coverage:	Universe of firms that need to provide annual accounts.
Source:	National Central Bank data. Information for small firms is completed using the VAT declaration which is compulsory for all firms.
Unit of analysis:	Legal business entities, including affiliates of foreign businesses.
Threshold:	None
Deflators:	NACE 2 digit value added deflators provided by the National Bank of Belgium (Belgostat) and Eurostat.
Additional filters:	In the dataset that will be released, all cells for which less than 4 firms were used to make the computation will be removed.
Other info:	Firms were classified into different sectors according to the NACE code they specify in their VAT declaration. The last NACE code provided by the firm was used; for firms for which the last NACE code was in NACE rev1.1, the code to was converted to NACE v.2 according to a conversion table provided by the Belgian Statistical Institute. Large firms provide all their financial information, while smaller firms have to provide a limited number of variables (value added, employment, wage).

CZECH REPUBLIC

Coverage:	Universe of firms with >20 employees in 2002-2007; firms >50 employees since 2008. Sample of smaller firms.
Source: Czech	Statistical Office data collected through regular reporting.
Unit of analysis:	Legal and physical persons registered in the business register and selected physical persons not registered in the business register are included.
Threshold:	None
Deflators:	Deflators in CZK (as provided by the Czech Statistical Office) are value added in current prices divided by value added in constant (2005)

	prices. Deflators in CZK (defl_CZK) converted to EUR as: defl_CZK / (CZK_EUR/CZK_EUR2005), where CZK_EUR is the mean exchange rate in a given year, CZK_EUR2005 is the mean exchange rate in 2005.
Additional filters:	No additional filters were applied.

GERMANY

Coverage:	Sample.
Source:	Financial Statements Statistics (secondary statistics). The data are compiled from: • -commercial data suppliers • -financial statements data pool: firm-level information mainly sourced from financial service providers that dispose of extensive balance sheet information of their customers • -Bundesbank's own data resources: information from firms that are creditors by banks which use these commercial credits as collateral for their refinancing.
Unit of analysis:	Firms as a legal entity are included.
Threshold:	None
Deflators:	Value-added deflators at the NACE rev. 2 disaggregation level provided by Eurostat.
Additional filters:	Due to confidentiality constraints, less than 20 observations per cell at the sector level were dropped.
Other info:	Data comprises reports by firms as a legal entity as opposed to the Eurostat classification, which includes reports by organizational units (e.g. the leading entity of a corporate group).This may result in a shift across sectors with respect to the Eurostat classification in the case legal entity and the organizational unit belong to different sectors and in differences in the aggregate turnover since reports from organizational units are consolidated. Sample is biased towards larger firms. Employment information may be missing especially for small and medium-sized firms.

ESTONIA

Coverage:	Universe.
Source:	Business Register.
Unit of analysis:	All legal forms of business entities, including affiliates of foreign.
Threshold:	None
Deflators:	Value-added deflators at the NACE rev. 2 disaggregation level provided by Eurostat.
Additional filters:	No additional filters were applied.

Other info: Eurostat statistics on Estonian businesses are based on a survey of firms conducted by Statistics Estonia, EKOMAR. The main differences with respect to Business Register data are the following:

- -EKOMAR covers the universe of larger firms (>20 employees) and random sample of smaller firms (1-19 employees), while Business Register covers all the firms.
- -EKOMAR does not include firms that offer services only for public sector and firms that are not value added tax payers (yearly turnover <16,000 EUR). All these groups are part of Business Register.
- Until 2010, employment was collected in the Business Register by a supplementary non-compulsory report which was still submitted by most firms. Since 2010 it is not possible to submit a report without employment data however employment has becomes much noisier.

FRANCE

Coverage: Sample.

Source: INSEE data. Calculations involving individual data from the BRN (Bénéfices Réels Normaux) were performed by the researchers of the Banque de France having a formal agreement from the Comité National du Secret Statistique (CNIS), in respect of all confidentiality rules implied by the agreement.

Unit of analysis: Fiscal regime “normal” (small firms under-represented). “Simplified” fiscal regime is adopted by a large number of very small firms. The version of the BRN data provided by INSEE to the researchers under

	agreement at Banque de France do not include these very small firms or the self-employed.
Threshold:	None
Deflators:	Deflators provided by EUklems (identical to deflators from INSEE)
Additional filters:	Due to confidentiality constraints, a sector was dropped when at least one of the following was fulfilled: • One firm has > 85% of the sector's turnover • 5 firms exist

The following sectors have been deleted:

Data_20E	
year	sectors deleted
1995	5 6 9 12 39 61 75
1996	5 6 9 12 39 75
1997	5 9 12 75
1998	5 9 12 75
1999	5 9 12 75
2000	5 9 12
2001	5 9 12 75
2002	5 9 12 75
2003	5 7 9 12 75
2004	5 7 9 12
2005	5 9 12
2006	5 7 9 12
2007	7 9 12

Data_ALL	
year	sectors deleted
1995	5 12 61
1996	6 12
1997	12
1998	12
1999	12
2000	12
2001	5 12
2002	12
2003	5 12
2004	5 12
2005	5 12
2006	5 12
2007	5 12

HUNGARY

Coverage:	Universe of exporters and importers (above threshold) firms. Sample (representative) of domestic producers and of non-trading firms.
Source:	Corporate income tax return data augmented by the statistical office by NACE codes; corporate registry data (minor source).
Unit of analysis:	Limited liability companies, limited partnerships, corporations are included. Self-employed persons are excluded (as they are not subject to corporate taxation).
Threshold:	None
Deflators:	Value-added deflators at the NACE rev. 2 disaggregation level provided by Eurostat.

Additional filters:	Due to confidentiality constraints, the cell size used by the Statistical Office is 4 firms.

ITALY

Coverage:	Sample taken from the universe of corporate firms with >10 employees. Sample of smaller firms.
Source:	Firms' Balance Sheets data; Business Register data (only for employment).
Unit of analysis:	Corporation (in Italian "Società di capitali"). Self-employed persons are not included.
Threshold:	None
Deflators:	Value-added deflators at the NACE rev. 2 disaggregation level provided by Eurostat.
Additional filters:	Due to confidentiality constraints, NACE sector 12 (Tobacco) was deleted.
Other info:	All firms that in the period considered have undergone m&a or other legal changes are excluded.

POLAND

Coverage:	Universe of non-financial firms.
Source:	Central Statistical Office data.
Unit of analysis:	forms of ownership with >10 employees. Self-employed persons are excluded.
Threshold:	> 20 employees
Deflators:	Value-added deflators at the NACE rev. 2 disaggregation level provided by Eurostat.
Additional filters:	No additional filters were applied.
Other info:	All non-financial enterprises with >10 employees are requested to fill the financial statement questionnaire (every 6 months) and the balance sheet questionnaire (once a year). Firms with >50 employees need to fill the financial statement questionnaire every quarter.

The requirement applies to NACE sections A (excluding individual farms), B, C, D, E, F, G, H, I, J, K (excluding banks, cooperative banking, insurance enterprises, currency and stock dealing enterprises, investment funds and mutual funds), L, M, N, P (excluding higher education), Q (excluding public health care providers), S.

SLOVENIA

Coverage:	Sample.
Source:	Agency of the Republic of Slovenia for Public Legal Records and Related Services (AJPES).
Unit of analysis:	Companies, namely limited liability companies, joint stock companies, and partnerships, that have submitted annual reports.
Threshold:	None
Deflators:	Deflators provided by the Statistical Office of the Republic of Slovenia (http://pxweb.stat.si/pxweb/Dialog/varval.asp?ma=0301910E&ti=&path=../Database/Economy/03_national_accounts/05_03019_GDP_annual/&lang=1).
Additional filters:	Due to confidentiality constraints, results exposing the performance of a single firm were dropped.

SLOVAKIA

Coverage:	Sample of all firms above the threshold. The survey is designed as exhaustive for the given group of companies and covers >80% of firms.
Source:	Statistical Office of the Slovak Republic data (Annual reports on production industries gathered through annual questionnaire of productive branches).
Unit of analysis:	Limited liability companies, joint-stock companies, general partnerships, limited partnerships or co-operatives are included. Self-employed persons are excluded.
Threshold:	>20 employees or alternatively >5 million EUR turnover
Deflators:	Official value added deflators provided by the Statistical Office of the Slovak Republic (Chained volume method used for constant prices).

Additional filters:	No additional filters were applied.
Other info:	The Annual questionnaire of productive branches is the main source for Structural Business Statistics published by Eurostat. There are three types of the questionnaire of productive branches differing by the number of surveyed indicators with regard to the size category of enterprise and coverage. Structural Business Statistics take into account results from all three types of the questionnaire. The sample contains the results of the questionnaire for large enterprises (see the official threshold), since the other two questionnaires are not exhaustive and small enterprises/legal units cannot be tracked in time.

SPAIN

Coverage:	Sample of mercantile firms.
Source:	Spanish Central Balance Sheet Database (CBSD) for most of large firms in Spain; Firm Registries for small and medium firms.
Unit of analysis:	Mercantile firms (Sociedades Mercantiles) are included. Self-employed persons are excluded.
Threshold:	None
Deflators:	Deflators based on National Accounts data provided by National Statistical Office for 2000-2011; Banco de España estimates for 1995-1999.
Additional filters:	Due to confidentiality constraints, an observation was dropped if at least one of the following was fulfilled: • <= 3 firms were included • or one firm accounted for >80% VA.
Other info:	Detailed info regarding the databases used (only in Spanish) available at: http://www.bde.es/bde/es/secciones/informes/Publicaciones_an/Central_de_Balan/anoactual/. The data are still preliminary for the year 2011, especially in the case of the Firm Registries.

ANNEX 2: INDUSTRY CLASSIFICATION

Code	NACE rev.2-2 digits	NACE Section	Selected NACE Industries
5	Mining of coal and lignite	B	MINING AND QUARRING
6	Extraction of crude petroleum and natural gas	B	
7	Mining of metal ores	B	
8	Other mining and quarrying	B	
9	Mining support service activities	B	
10	Manufacture of food products	C	MANUFACTURING
11	Manufacture of beverages	C	
12	Manufacture of tobacco products	C	
13	Manufacture of textiles	C	
14	Manufacture of wearing apparel	C	
15	Manufacture of leather and related products	C	
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	C	
17	Manufacture of paper and paper products	C	
18	Printing and reproduction of recorded media	C	
19	Manufacture of coke and refined petroleum products	C	
20	Manufacture of chemicals and chemical products	C	
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	C	
22	Manufacture of rubber and plastic products	C	
23	Manufacture of other non-metallic mineral products	C	
24	Manufacture of basic metals	C	
25	Manufacture of fabricated metal products, except machinery and equipment	C	
26	Manufacture of computer, electronic and optical products	C	
27	Manufacture of electrical equipment	C	
28	Manufacture of machinery and equipment n.e.c.	C	
29	Manufacture of motor vehicles, trailers and semi-trailers	C	
30	Manufacture of other transport equipment	C	
31	Manufacture of furniture	C	
32	Other manufacturing	C	
33	Repair and installation of machinery and equipment	C	
35	Electricity, gas, steam and air conditioning supply	D	ELECTRICITY[…]
36	Water collection, treatment and supply	E	WATER […] AND REMEDIATION ACTIVITIES
37	Sewerage	E	
38	Waste collection, treatment and disposal activities; materials recovery	E	

39	Remediation activities and other waste management services	E	
41	Construction of buildings	F	CONSTRUCTION
42	Civil engineering	F	
43	Specialised construction activities	F	
55	Accommodation	I	ACCOMMODATION[…]
56	Food and beverage service activities	I	
58	Publishing activities	J	INFORMATION AND COMMUNICATION
59	Motion picture, video and television programme production, sound recording and music publishing activities	J	
60	Programming and broadcasting activities	J	
61	Telecommunications	J	
62	Computer programming, consultancy and related activities	J	
63	Information service activities	J	
69	Legal and accounting activities	M	PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES
70	Activities of head offices; management consultancy activities	M	
71	Architectural and engineering activities; technical testing and analysis	M	
72	Scientific research and development	M	
73	Advertising and market research	M	
74	Other professional, scientific and technical activities	M	
75	Veterinary activities	M	
77	Rental and leasing activities	N	ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES
78	Employment activities	N	
79	Travel agency, tour operator reservation service and related activities	N	
80	Security and investigation activities	N	
81	Services to buildings and landscape activities	N	
82	Office administrative, office support and other business support activities	N	

ANNEX 3: SENSITIVITY ANALYSIS TO THE USE OF DIFFERENT SECTOR WEIGHTS

In this Annex we redo figures 1 and 5 using 4 different (apart from the one used in the paper) sector weight schemes, defined as follows:

1) The one used in the paper: Time variant and common weights across countries

Let s denote the sector, i the country and t the year. The sector weight is defined as:

$$\theta_{st} = \sum_i VA_{sit} / \sum_s \sum_i VA_{sit}$$

In this case the weight does not depend on the specific country but is allowed to differ across time. Hence the across sector weighted average for each country-year is:

$$y_{it} = \sum_s \theta_{st} \varpi_{sit}$$

where ϖ_{sit} denotes labour productivity (or unit labour cost) in sector s, country i and time t

2) Time Variant - Country Specific

Let s denote the sector, i the country and t the year. The sector weight is defined as:

$$\theta_{sit} = VA_{sit} / \sum_s VA_{sit}$$

In this case the weight does depend on the specific country. Hence the across sector weighted average for each country-year is:

$$y_{it} = \sum_s \theta_{sit} \varpi_{sit}$$

where ϖ_{sit} denotes labour productivity (or ulc) in sector s, country I and time t

3) Time invariant, country common

Let s denote the sector, i the country and t the year. The sector weight is defined as:

$$\theta_s = \sum_i \sum_t VA_{sit} / \sum_s \sum_i \sum_t VA_{sit}$$

In this case the weight does not depend on the specific country nor on the specific year. We computed the across sector weighted average for each country-year as follows:

$$y_{it} = \sum_s \theta_s \varpi_{sit},$$

where ϖ_{sit} denotes labour productivity (or unit labour cost) in sector s, country i and time t

4) Time invariant, country specific

Let s denote the sector, i the country and t the year. The sector weight is defined as:

$$\theta_{si} = \sum_t VA_{sit} / \sum_s \sum_t VA_{sit}$$

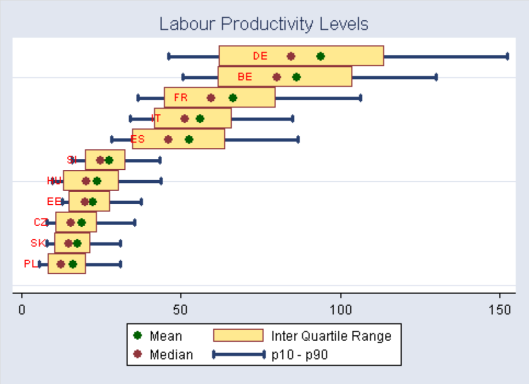
In this case the weight does not depend on year but is different across countries. Hence the across sector weighted average for each country-year is:

$$y_{it} = \sum_s \theta_{si} \varpi_{sit},$$

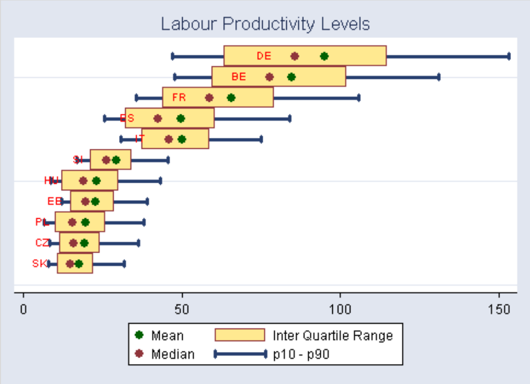
where ϖ_{sit} denotes labour productivity (or ulc) in secotor s, country i and time t.

FIGURE 1 LABOUR PRODUCTIVITY DISTRIBUTION, RESTRICTED SAMPLE

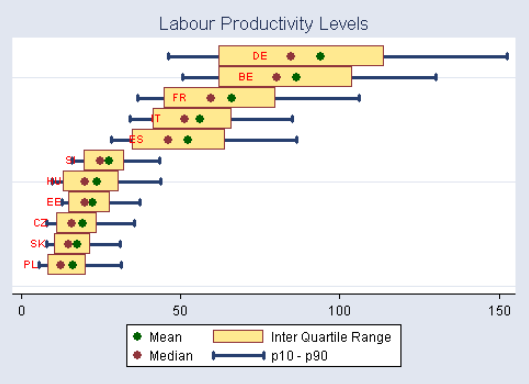
Sector weight 1: Common, time variant



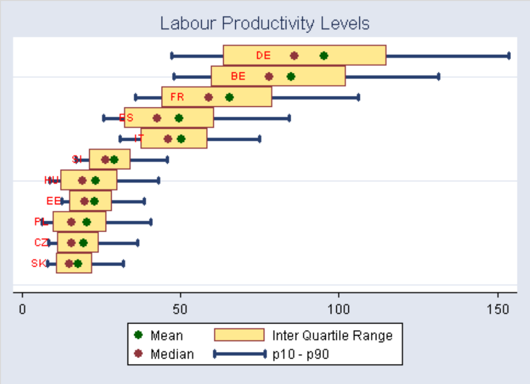
Sector weight 2: Country-specific, time variant



Sector weight 3: common, time invariant



Sector weight 4: country-specific, time invariant



Sector weight 5: Common, time variant (turnover share)

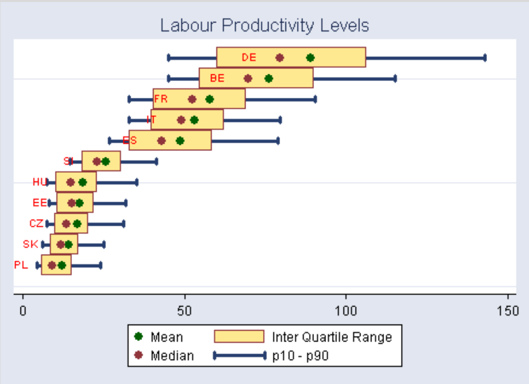
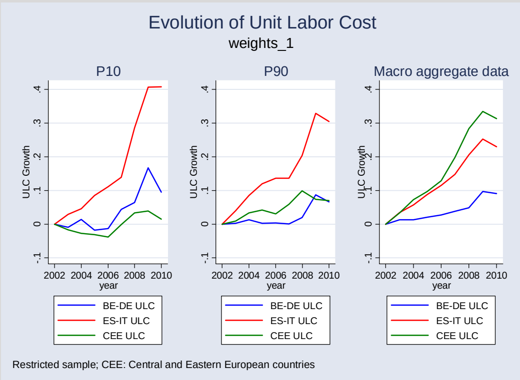
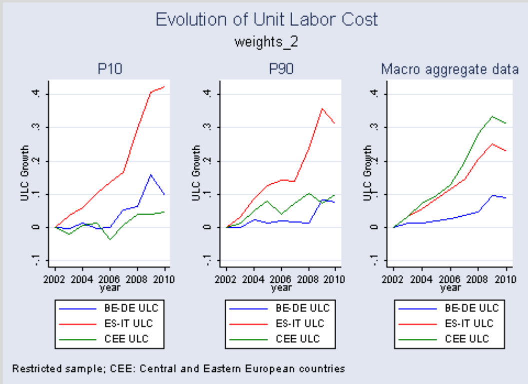


FIGURE 2 ULC EVOLUTION. 2002-2010. RESTRICTED SAMPLE

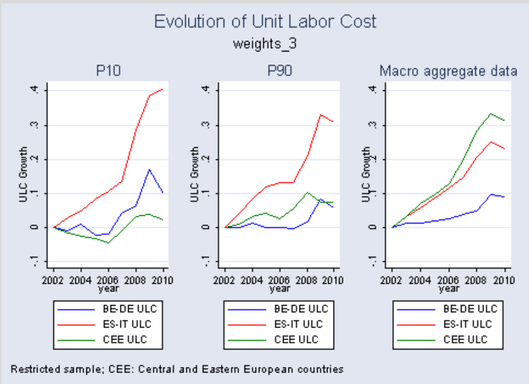
Sector weight 1: Common, time variant



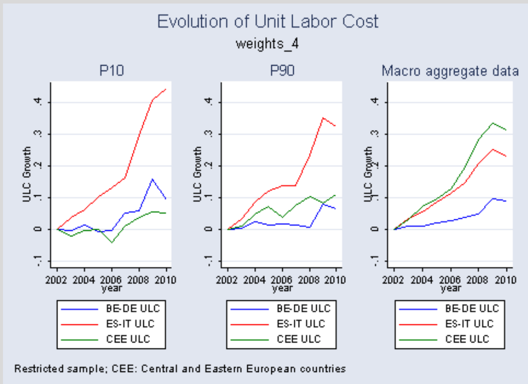
Sector weight 2: Country-specific, time variant



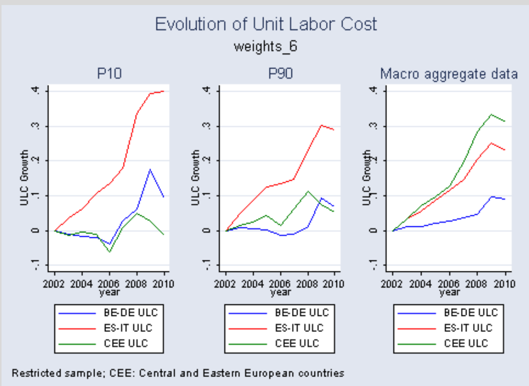
Sector weight 3: common, time invariant



Sector weight 4: country-specific, time invariant



Sector weight 5: Common, time variant (turnover share)



ANNEX 4: COMPLETE LIST OF VARIABLES

Indicators	Description	Var_list		
		Level	38Log[1]	First diff.[2]
Nr. of Employees		l	lnl	d_lnl
Value-added		va	lnva	d_lnva
Capital		k		
Material Costs	m/defl_va_35		lnm	d_lnlc
Labour Cost		lc	lnlc	
Turnover		turnover	lnturnover	d_lnturnover
Real VA	va/defl_va	rva	lnrva	d_lnrva
Real Capital	k/defl_va_35	rk	lnrk	
Real Turnover	turnover/ defl_va_99	rturnover	lnrturnover	
Capital/Labour	k/l	k_l	lnk_l	d_lnk_l
Turnover/Labour	rturnover_l	turnover_l	lnturnover_l	
Real Turnover/Labour	rturnover_l	rturnover_l	lnrturnover_l	
Cost per employee	lc/l	lc_l	lnlc_l	
Wage Share	lc/va	wageshare		d_wageshare
Labour Productivity	rva/l	lprod	lnlprod	d_lnlprod
Revenue based Labour Productivity	turnover/l	lprod_rev	lnlprod_rev	d_lnlprod_rev
Capital Productivity	rva/rk	kprod	kprod	d_kprod
Unit Labour Cost (ULC)	lc/rva	ulc	lnulc	d_lnulc
Total Factor Productivity (TFP)	See methodological session		tfp	d_tfp
Productivity Decomposition				
Olley-Pakes Decomposition of:				
Labour Productivity	total(lnlprod*l)	totwlprod		
	total(l)	totl_weight		
	OPgap= (totwlprod/totl_weight)-avlprod	lopgap		
Revenue Based Labour Productivity	total(lnlprod_rev*l)	totwlprod_rev		
	OPgap= (totwlprod_rev/totl_weight)-avlprod_rev	lopgap_rev		
Capital Productivity	total(lnkprod*k)	totwkprod		
	total(k)	totk_weight		
	OPgap= (totwkprod/totk_weight)-avkprod	kopgap		

38 [1] General Notation of variables in logarithm is: lnvar_list.

38 [2] General Notation of the first difference of logarithmic variables is: d_lnvar_list. First difference is calculated as follows: d_lny=(lny-lny[_n-1]).

TFP, Input weighted	total(lntfp*inp)	totw_inp_tfp
	total(inp)	tottfp_inp_weight
	OPgap= (totw_inp_tfp/tottfp_inp_weight)-avtfp	tfp_inp_opgap
TFP, Output weighted	total(lntfp*avrva)	totw_out_tfp
	total(avrva)	tottfp_out_weight
	OPgap= (totw_out_tfp/tottfp_out_weight)-avtfp	tfp_out_opgap
Foster Decomposition of TFP (without taking into account entry-exit factors)		
2 Year Lag	Labour Productivity	labTOT_within, labTOT_between, labTOT_covariance, labLP_1, labTOT_diff
	Revenue Based Labour Productivity	lab_revTOT_within, lab_revTOT_between, lab_revTOT_covariance, lab_revLP_1, lab_revTOT_diff
	TFP	TOT_within, TOT_between, TOT_covariance, LP_1, TOT_diff
	ULC	ulcTOT_within, ulcTOT_between, ulcTOT_covariance, ulcLP_1, ulcTOT_diff
5 Year Lag	Labour Productivity	tlabTOT_within, tlabTOT_between, tlabTOT_covariance, tlabLP_1, tlabTOT_diff
	Revenue Based Labour Productivity	tlab_revTOT_within, tlab_revTOT_between, tlab_revTOT_covariance, tlab_revLP_1, tlab_revTOT_diff
	TFP	tTOT_within, tTOT_between, tTOT_covariance, tLP_1, tTOT_diff
	ULC	tulcTOT_within, tulcTOT_between, tulcTOT_covariance, tulcLP_1, tulcTOT_diff

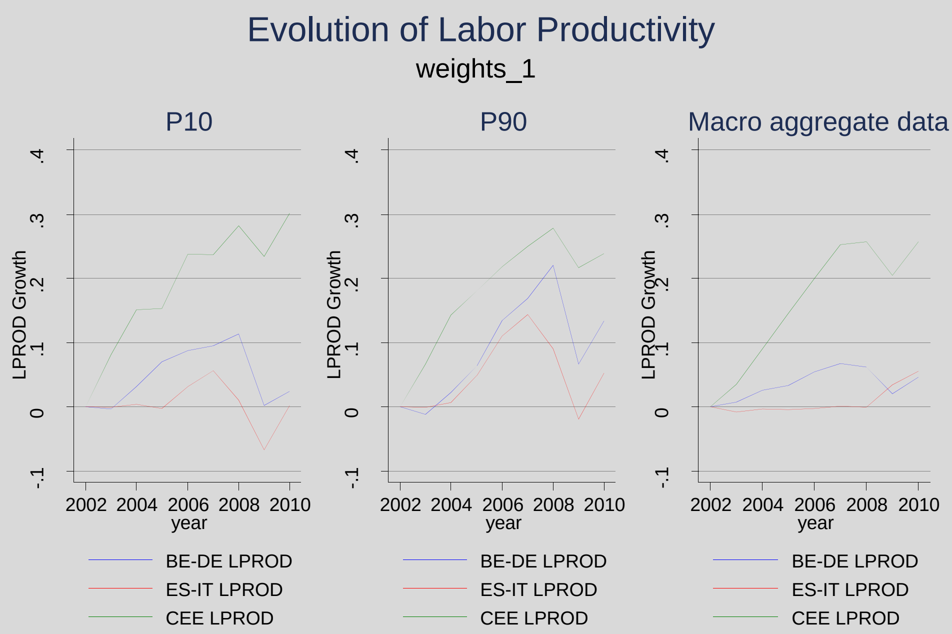
Additional average variables, matching the moments of productivity distributions	Labour Productivity	lprod_p1_`z'
Where z= l, va, rva, k, rk, lc, lc_l, wageshare, turnover		lprod_p10_`z'
		lprod_p25_`z'
		lprod_p50_`z'
		lprod_p75_`z'
		lprod_p90_`z'
		lprod_p99_`z'
	Revenue Based Labour Productivity	lprod_rev_p1_`z'
		lprod_rev_p10_`z'
		lprod_rev_p25_`z'
		lprod_rev_p50_`z'
		lprod_rev_p75_`z'
		lprod_rev_p90_`z'
		lprod_rev_p99_`z'
	Capital Productivity	kprod_p1_`z'

		kprod_p10_`z'
		kprod_p25_`z'
		kprod_p50_`z'
		kprod_p75_`z'
		kprod_p90_`z'
		kprod_p99_`z'
	TFP	tfp_p1_`z'
		tfp_p10_`z'
		tfp_p25_`z'
		tfp_p50_`z'
		tfp_p75_`z'
		tfp_p90_`z'
		tfp_p99_`z'
	ULC	ulc _p1_`z'
		ulc _p10_`z'
		ulc _p25_`z'
		ulc _p50_`z'
		ulc _p75_`z'
		ulc _p90_`z'
		ulc _p99_`z'

ANNEX 5: TURNOVER PRODUCTIVITY

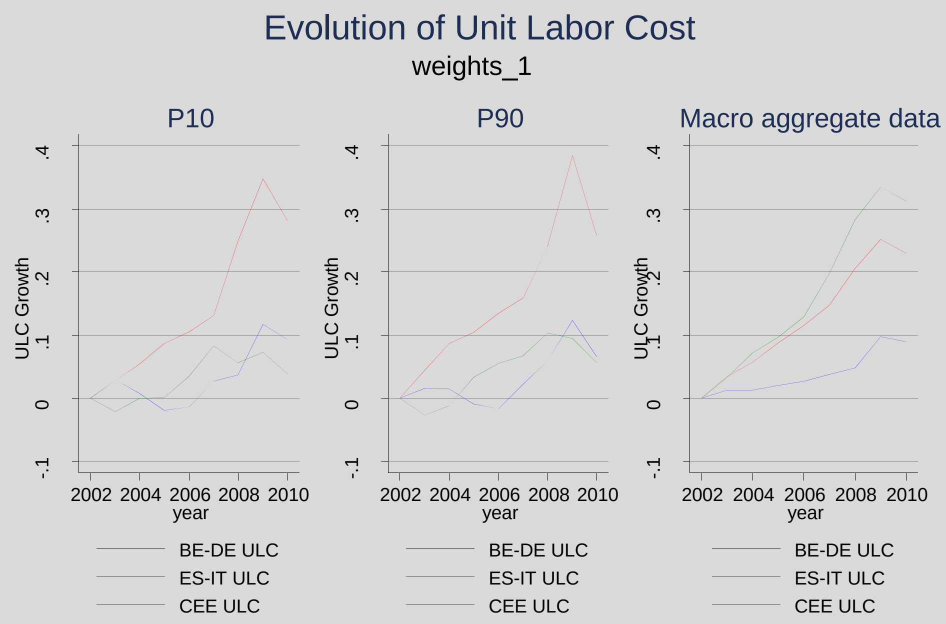
This Annex shows figures 4, 5 and 6 of the main text redrawn using productivity defined as turnover over employees, instead of real value added over employees.

FIGURE A4 EVOLUTION OF LABOUR PRODUCTIVITY IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, RESTRICTED SAMPLE. PRODUCTIVITY DEFINED IN TERMS OF TURNOVER OVER EMPLOYEES.



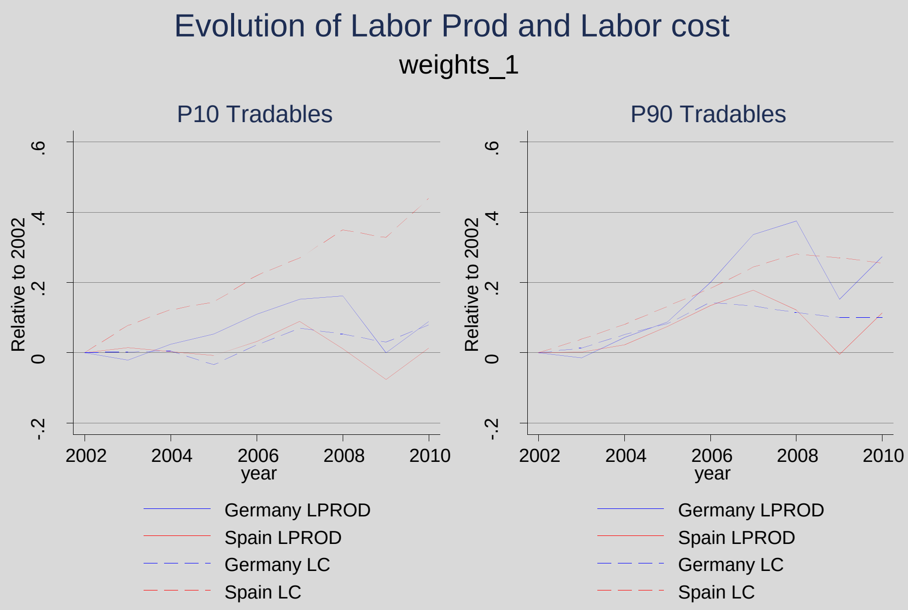
Labor prod: revenue/Labor

FIGURE A5 EVOLUTION OF UNIT LABOUR COST (AVERAGE LABOUR COST OVER PRODUCTIVITY) IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, RESTRICTED SAMPLE. PRODUCTIVITY DEFINED IN TERMS OF TURNOVER OVER EMPLOYEES



Labor prod: revenue/Labor

FIGURE A6 EVOLUTION OF AVERAGE LABOUR COST AND PRODUCTIVITY IN DIFFERENT TAILS OF THE PRODUCTIVITY DISTRIBUTION, RESTRICTED SAMPLE. PRODUCTIVITY DEFINED IN TERMS OF TURNOVER OVER EMPLOYEES. SPAIN AND GERMANY



Labor prod: revenue/Labor

ANNEX 6: TOTAL FACTOR PRODUCTIVITY ESTIMATION

Assume the following Cobb-Douglas production function:

$$Y_{it} = A_{it} K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M} \qquad (1)$$

Where Y is physical output of firm i at time t, K, L and M are inputs and A is the Hicksian neutral efficiency level of the firm. Y, L, K and M are econometrically observed whereas A is not, although it is known by the firm.

Taking equation (1) in logs:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_L l_{it} + \beta_M m_{it} + \omega_{it} + u_{it} \qquad (2)$$

where $\ln(A_{it}) = \beta_0 + \omega_{it} + u_{it}$ (3)

with β_0 representing the mean-efficient level across firms and over time and $\omega_{it} + u_{it}$ is a firm-specific deviation from that mean. The first component of which refers to an unobserved firm-level time-variant productivity level, known by the firm, and the second component is an i.i.d error term representing unexpected (by the firm) shocks, and therefore independent of the rest of explanatory variables.

Equation (2) could be consistently estimated by OLS only if firm's variable input choices are independent of the unobserved shocks, including firm-level productivity. That is very unlikely to be the case since productivity is observed by the firm. Therefore it will influence the choice of the optimal bundle of inputs:

$$E(\omega_{it}, x_{it}) > 0$$

where x represent firm's variable (or semi-variable) inputs. If it is assumed that the higher the firm-level productivity, the larger the quantities of the inputs chosen by the firm, the technology coefficients of labour and materials will be upward biased. If labour is the only freely available input and capital being quasi-fixed (Levinsohn and Petrin, 2003), the technology coefficient of capital will be downward biased. The simultaneity bias caused by the relationship between unobserved productivity shocks and production inputs (Marschak & Andrews, 1944[39]) is the major problem to be tackled in production function estimation.

One of the solutions provided for solving this problem is represented by semi-parametric estimators or control function approach. Olley & Pakes (1996) proposed a structural approach to the problem, by using observed input choices to instrument for unobserved productivity. Their

39 See also Griliches and Mairesse (1995)

original work proposes to use investment demand to proxy unobserved productivity. More concretely, they assume that investment is a monotonically increasing function of two state variables: capital stock, which depends on past investment decisions, and unobserved productivity:

Assumption 1: $i_{it} = \mathrm{f}(k_{it}, \omega_{it})$ with f being a function which increases monotonically with productivity, and therefore can be inverted to factor out unobserved productivity:

$\omega_{it} = f^{-1}(i_{it}, k_{it})$ (4)

Assumption 2: $K_{it} = I_{it-1} + (1-\delta)K_{it-1}$, that is, K_{it} is a function of past values of investment, not current.

Substituting (4) into (2):

$y_{it} = \beta_L l_{it} + \beta_M m_{it} + \emptyset(i_{it}, k_{it}) + u_{it}$ (5)

With $\emptyset(i_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + f^{-1}(i_{it}, k_{it})$ (6)

And $E(u_{it} \backslash l_{it}, m_{it}, k_{it}) = 0$ (7)

In the first stage of the OP algorithm, (6) is treated non-parametrically and approximated by a third-order polynomial in i_{it} *and* k_{it}. Then (5) can be estimated consistently with OLS. Although the coefficients of labour and materials are identified in this first-stage regression, the coefficient of capital cannot be identified given it is collinear in the non-parametric function.

In order to identify the capital coefficient, one further assumption on firm dynamics is required:

Assumption 3: the unobserved productivity is assumed to follow a first-order Markov process, that is, the level of productivity at t is uncorrelated to the change in productivity between t-1 and t:

$\omega_{it} = E(\omega_{it} \backslash \omega_{it-1}) + \varepsilon_{it}$ (8)

Where ε_{it} is an unexpected shock to the productivity. To proceed to the second stage of the algorithm in which the capital coefficient is identified, it is important to note that by Assumption 2 (capital only depends on past values of investment), the error term in (8) is uncorrelated to capital:

$E(\varepsilon_{it} \backslash k_{it}) = 0$ (9)

Now, given (5) and the estimated coefficients of labour and materials, it is possible to compute $\widehat{\emptyset}(i_{it}, k_{it}) = y_{it} - \widehat{\beta_L} l_{it} - \widehat{\beta_M} m_{it}$. For a given value of $\dot{\beta}_k$, and given (6) and (4), the productivity level for t can be predicted (up to a constant) as follow:

$\widehat{\omega_{it}} = \widehat{\emptyset}(i_{it}, k_{it}) - \dot{\beta}_k k_{it}$ (10)

Approximation to $E(\omega_{it}\backslash\omega_{it-1})$ (called $E(\omega_{\iota t}\widehat{\backslash\omega_{\iota t-1}})$) is given by predicted values from regression:

$$\widehat{\omega_{\iota t}} = \gamma_0 + \gamma_1\widehat{\omega_{\iota t-1}} + \gamma_2\omega_{it-1}\widehat{\omega_{\iota t-1}}^2 + \gamma_3\widehat{\omega_{\iota t-1}}^3 + \varepsilon_{it} \quad (11)$$

Lastly, given $\widehat{\beta_L}$, $\dot{\beta}_k$ and $E(\omega_{\iota t}\widehat{\backslash\omega_{\iota t-1}})$, the estimated capital coefficient will be chosen to minimize the sum of the squared sample residuals:

$$min_{\dot{\beta}_k} \sum_t \left[y_{it} - \widehat{\beta_L}l_{it} - \widehat{\beta_M}m_{it} - \dot{\beta}_k k_{it} - E(\omega_{\iota t}\widehat{\backslash\omega_{\iota t-1}})\right]^2$$

Finally, a bootstrap based on random sampling from observations is used to construct standard errors for the estimates of the technology coefficients.

The first problem with the OP (1996) algorithm was highlighted by Levinsohn & Petrin (2003). They noted that the strict monotonicity of the investment function, with respect to productivity and capital, was broken given the many zeros reported by firms. If all observations with zero or negative investment had to be dropped, there would be an important efficiency loss in the estimation.

Levinsohn and Petrin (2003) propose as a solution to proxy productivity with material inputs demand, instead of investment demand. Hence $m_{it} = \mathrm{h}(k_{it}, \omega_{it})$, which can be claimed to be strictly increasing in productivity and, therefore, can be inverted out to factor productivity. Moreover, there are few missings or zero observations in variables such as energy or some other intermediate input consumption at the firm level. Equation (4) would be rewritten as:

$$\omega_{it} = h^{-1}(mi_{it}, k_{it}) \quad (4a)$$

and

$$y_{it} = \beta_L l_{it} + \breve{\emptyset}(m_{it}, k_{it}) + u_{it} \text{ with} \quad (5a)$$

$$\breve{\emptyset}(m_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + \beta_m m_{it} + f^{-1}(m_{it}, k_{it}) \quad (6a)$$

The only difference with OP (1996) algorithm is that the coefficient of material inputs has to be recovered now in the second stage. For that purpose, the following identification assumption has to be made, apart from (9):

$$E(\varepsilon_{it}\backslash m_{it-1}) = 0 \quad (9a)$$

The main critique to OP (1996) and LP (2003) came from Ackerberg, Caves and Frazer (2006). They claimed that the labour coefficient cannot be identified in the first stage of the algorithm due to collinearity problems. They argue that either labour is chosen before material inputs, hence $m_{it} = h'(k_{it}, \omega_{it}, l_{it})$ in which case the labour coefficient cannot be identified in the first-stage, or material inputs and labour are chosen at the same time, once the productivity has been observed. In that latter case, the labour choice is assumed to depend, as material inputs do, on the state variables at t, productivity and capital:

$$l_{it} = \mathrm{g}(k_{it}, \omega_{it})$$

Hence, substituting ω_{it} by its value in (4a),

$$l_{it} = g\left(k_{it}, h^{-1}(mi_{it}, k_{it})\right) = g'(mi_{it}, k_{it})$$

which means that labour is a function of material input and capital which invalidates the identification of labour coefficient in the first step.

Finally, Wooldridge (2009) shows a method to implement OP/LP in a GMM framework with at least three advantages over Ackerberg et al: (1) the first stage of the algorithm contains identifying information for the parameters on the variable inputs; (2) fully robust standard error are easy to obtain (no need to bootstrap), and (3) estimators are more efficient.

The advantage about the efficiency of the estimators comes about because the algorithm proposed by Wooldridge takes account of the potential contemporaneous correlation across errors of the two stages as well as accounts for heteroskedasticity or serial correlation. Wooldridge strengthens assumption (7) to include independence of past values of input choices:

$$E\left(u_{it} \backslash l_{it}, m_{it}, k_{it}, l_{it-1}, m_{it-1}, k_{it-1,\dots,} l_1, m_1, k_1\right) = 0 \,. \qquad (7a)$$

In this way, Wooldridge allows for serial dependence on the idiosyncratic shock u_{it}.

Wooldridge notes that given the assumption on productivity dynamics in (8), in the second stage lagged values of the variables inputs l and m are used to estimate the capital coefficient. Hence consistency requires reinforcing (9) so the stochastic shock to productivity is independent on the current value of capital as well as on the lagged values of the variable inputs:

$$E(\varepsilon_{it} \backslash k_{it}, k_{it-1}, m_{it-1}, l_{it-1}, \dots) = 0 \qquad (9b)$$

This means that:

$$\omega_{it} = E(\omega_{it} \backslash \omega_{it-1}) + \varepsilon_{it} = \mathrm{j}[h^{-1}(k_{it-1}, m_{it-1}) + \varepsilon_{it} \quad (12)$$

Now, plugging (6a) into (5a) and (12) into (2), we get a system of two equations with the same dependent variable and fixed and variable inputs as explanatory variables. The difference between both is the way we approximate unobserved productivity. These two different ways provide a different set of instruments for identification:

$$y_{it} = \beta_L l_{it} + \beta_M m_{it} + \beta_0 + \beta_k k_{it} + h^{-1}(m_{it}, k_{it}) + u_{it} \qquad (13)$$

$$y_{it} = \beta_L l_{it} + \beta_M m_{it} + \beta_0 + \beta_k k_{it} + j[h^{-1}(k_{it-1}, m_{it-1})] + \varepsilon_{it} + u_{it} \qquad (14)$$

with $E\left(u_{it} \backslash l_{it}, m_{it}, k_{it}, l_{it-1}, m_{it-1}, k_{it-1,\dots,} l_1, m_1, k_1\right) = 0$

and $E(\varepsilon_{it} + u_{it} \backslash k_{it}, k_{it-1}, m_{it-1}, l_{it-1}, \dots) = 0$

Both equations can be estimated simultaneously in a GMM framework using the appropriate set of instruments. Alternatively, equation (14) can be estimated by pooled IV. In particular, if y_{it} is (real) value added and the productivity process is a random walk with a drift $\omega_{it} = \tau + \omega_{it-1} + \varepsilon_{it}$, equation (14) becomes

$$y_{it} = \beta_L l_{it} + (\beta_0 + \tau) + \beta_k k_{it} + h^{-1}(k_{it-1}, m_{it-1}) + \varepsilon_{it} + u_{it} \qquad (15)$$

We approximate h^{-1} with a low-degree polynomial (of order up to three) and estimate equation (15) using pooled IV with instruments for l_t.[40] The table A6 below shows the estimated labour and capital coefficients estimated using this methodology, for all countries, grouped in broad industries (1 digit NACE level).

[40] This implementation of Wooldrige (2009) method is used in Petrin and Levinsohn (2012) and Galuscak and Lizal (2011).

TABLE A6 ESTIMATED LABOUR AND CAPITAL COEFFICIENTS, BROAD SECTORS.WOOLDRIDGE (2009) AMENDED LIKE IN PETRIN AND LEVINSOHN (2012) AND GALUSCAK AND LIZAL (2011).

	BE		CZ		DE		EE		ES		FR		HU		IT		SI		SK	
Broad Sector	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**	**k**	**l**
Mining	n.s.	0.4	n.s.	0.6	0.1	0.7	n.s.	n.s.	0.1	0.6	0.4	0.5	0.5	0.5	0.1	0.6	0.3	0.5	0.6	0.6
Manufacturing	0.1	0.6	0.2	0.6	0.1	0.6	0.2	0.6	0.0	0.8	0.3	0.7	0.2	0.5	0.1	0.8	0.2	0.6	0.4	0.4
Electricity	n.s.	0.1	n.s.	0.7	0.2	0.3	0.1	0.5	n.s.	0.3	0.3	0.4	n.s.	0.5	0.1	0.5	0.2	0.4	0.3	0.2
water supply	0.1	0.5	n.s.	0.6	0.1	0.5	0.3	0.6	0.1	0.7	0.3	0.6	0.2	0.1	0.1	0.7	0.2	0.6	0.4	0.5
Construction	0.1	0.6	0.2	0.7	0.0	0.6	0.2	0.7	0.0	0.7	0.2	0.6	0.3	0.5	0.0	0.9	0.2	0.6	0.3	0.7
accomodation and food	0.1	0.5	n.s.	0.5	n.s.	0.5	0.0	0.8	0.1	0.6	0.3	0.6	0.1	0.2	0.0	0.8	0.1	0.6	0.2	0.9
information and communication	0.1	0.7	n.s.	0.9	0.1	0.7	0.1	0.8	0.1	0.8	0.2	0.8	0.2	0.6	0.0	1.0	0.1	0.8	0.3	0.8
professional activities	0.1	0.7	n.s.	0.9	n.s.	0.7	0.1	0.9	0.1	0.8	0.2	0.7	0.1	0.7	0.0	0.9	0.2	0.8	0.7	0.9
administrative activities	0.1	0.7	n.s.	0.7	0.1	0.6	0.1	0.7	0.1	0.8	0.2	0.7	0.1	0.6	0.0	0.8	0.1	0.7	0.3	0.7

Notes: Poland is not included because estimations were run on full samples. We show the weighted average of coefficients for broader sectors of activity. All available years for each country are used for the regression.

ANNEX 7: ALLOCATIVE EFFICIENCY INDICATORS

Static Allocative Efficiency (Olley and Pakes, 1996)

Let 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. The productivity of industry s can be decomposed as:

$$y_{st} = \sum_{i\in s} \theta_{it}\omega_{it} = \overline{\omega_{st}} + \sum_{i\epsilon s}(\theta_{it} - \overline{\theta_{st}})(\omega_{it} - \overline{\omega_{st}}), \qquad (15)$$

where S is the set of firms belonging to industry s, θ_{it} and $\omega_{i,t}$ represent size and productivity of firm i at time t, respectively, $\overline{\theta_{st}}$ and $\overline{\omega_{st}}$ represent the unweighted mean size and productivity of industry s at time t, respectively. The decomposition in equation (15) splits the weighted average of firm productivity in two components: the unweighted industry mean and the covariance between productivity and size. The larger the covariance, the higher the share of size associated to more productive firms, and the higher is industry productivity. This term, labelled "OP Gap", captures the static allocative efficiency. It represents the increase in industry productivity y_{st} due to reallocation of resources from less productive to more productive firms, and not due to an increase in average productivity ($\overline{\omega_{st}}$).

The shares used in the decomposition are calculated according to different variables and depending on the productivity indicators analyzed. In particular:

Labour Productivity: the decomposition is weighted by number of employees, with weights calculated as

$$\theta_{it} = l_{it} / \sum_i l_{it}$$

Revenue based Labour Productivity: the decomposition is weighted by number of employees, with weights calculated as

$$\theta_{it} = l_{it} / \sum_i l_{it}$$

Capital Productivity: the decomposition is weighted by total asset, with weights calculated as

$$\theta_{it} = k_{it} / \sum_i k_{it}$$

TFP: the decomposition is weighted by both input and output weighted, with weights calculated as follows

Inputs weighted:

$$xqm_{it} = 0.5 * ((m_{it}/RVA_{it})(m_{it-1}/RVA_{it-1}))$$

$$xqe_{it} = 0.5 * ((lc_{it}/RVA_{it})(lc/RVA_{it-1}))$$

$$imp_{it} = \left(\left(m_{it}^{xqm_{it}}\right)\left(l_{it}^{xqe_{it}}\right)\left(k_{it}^{(1-\,xqm_{it}-xqe_{it})}\right)\right)$$

$$\theta_{it} = imp_{it}\Big/\sum\nolimits_{i} imp_{it}$$

Output weights:

$$avRVA_{it} = 0.5 * (RVA_{it} + RVA_{it-1})$$

$$\theta_{it} = avRVA_{it}\Big/\sum\nolimits_{i} avRVA_{it}$$

Dynamic Allocative Efficiency (Foster, Haltiwanger, and Krizan, 2006)

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. The change in productivity of industry s from time t-k to time t can be decomposed as:

$$\Delta y_{st} = \sum_{i\in C}\theta_{it-k}\Delta\omega_{it} + \sum_{i\in C}(\omega_{i,t-k} - \widehat{\omega_{st-k}})\Delta\theta_{it} + \sum_{i\in C}\Delta\theta_{it}\,\Delta\omega_{it} + \sum_{i\in N}\theta_{it}(\omega_{it} - \widehat{\omega_{st-k}}) - \sum_{i\in X}\theta_{it-k}(\omega_{it-k} - \widehat{\omega_{st-k}}),$$

where Δ is the differential operator between t-k and t; C denotes continuing firms, N denotes entering firms, and X denotes exiting firms; θ_{it} and $\omega_{i,t}$ represent size and productivity of firm i at time t, respectively, $\widehat{\theta_{st}}$ and $\widehat{\omega_{st}}$ represent the weighted mean size and productivity of industry s at time t, respectively. The first term of the decomposition is the contribution of productivity changes for continuing firms with initial weights ("within component"); the second term is the effect of reallocating resources among continuing firms given their initial productivity ("between component"); the third term is the cross-effect of reallocation and productivity changes for continuing firms ("covariance or cross component"); finally, the fourth and fifth terms are the contribution from entry and from exit, respectively.

The Foster Decomposition implemented in this exercise does not take into account the contribution of exiting and entering firms. The decomposition is carried out for k=2 as well as k=5.

ANNEX 8: RELATING AGGREGATE EXPORTS TO FIRM'S LEVEL TFP

In this Annex Tables 8 and 9 from section 5 are replicated using TFP instead of labour productivity as indicator of the firm's efficiency as a robustness check.

TABLE A8 TRADE AND TFP DISPERSION

	1 Exp	2 Exp	3 Exp	4 Exp	5 Bal.Index	6 Bal.Index	7 Bal.Index	8 Bal.Index
TFP_{t-2}	.0835***	-.1778***	.1152***	-.07	.0346	-.23***	.0656***	-.2323***
	(.0229)	(.0521)	(.0246)	(.0657)	(.0231)	(.0585)	(.0248)	(.0755)
$SectVA_{t-211}$	.7309***	.7243***	.7159***	.7314***	.8006***	.794***	.786***	.8015***
	(.0255)	(.0255)	(.0257)	(.0255)	(.0361)	(.0361)	(.0363)	(.0359)
$Aver.Empl_{t-2}$	.0256	.0398	.0268	.0313	-.1515***	-.1371**	-.1503***	-.1416***
	(.0451)	(.0446)	(.0448)	(.0451)	(.0546)	(.0539)	(.0542)	(.0541)
TFP(s.d.)t-2		.2902***				.2939***		
		(.0517)				(.0604)		
TFP(skew)t-2			.1189***				.1164***	
			(.0274)				(.033)	
TFP(iqr)t-2				.1515***				.2634***
				(.0582)				(.0681)
Obs.	1877	1877	1877	1877	1877	1877	1877	1877
R2	.9606	.9613	.961	.9607	.571	.5766	.5738	.5743

All variables are in logs and lagged of two years. Each columns represent a different equation. *Exp:* export value. *TradeBal:* trade balance. *Imp:* import values. Estimation unit is defined by the triple country sector year. Robust standard errors are in parenthesis. We include the following groups of dummies: year, country, sector, country X year, and sector X year. Significance level: * significant at 10%; ** significant at 5%; ***significant at 1%

TABLE A9 TRADE AND PERCENTILES OF TFP DISTRIBUTION

	1	2	3	4	5	6	7	8
	Exp	Exp	Exp	Exp	Bal.Index	Bal.Index	Bal.Index	Bal.Index
$TFP(pc50)_{t-2}$	.0591**	-.6805***	-.3673***	-.0581	.0171	-.2021	-.1699	-.0645
	(.0237)	(.2208)	(.1289)	(.0562)	(.0236)	(.2605)	(.1472)	(.0609)
$SectVA_{t-2}$	.7329***	.7271***	.721***	.7086***	.8019***	.7969***	.794***	.7812***
	(.0255)	(.0255)	(.0257)	(.0254)	(.0361)	(.0362)	(.0365)	(.0362)
$Aver.Empl_{t-2}$	.0196	.0424	.0523	.0271	-.1551***	-.1382**	-.1331**	-.1492***
	(.0456)	(.0446)	(.0447)	(.0438)	(.0546)	(.0546)	(.0547)	(.0544)
$TFP(pc25)_{t-2}$		-.0526				-.3143*		
		(.1386)				(.1653)		
$TFP(pc75)_{t-2}$		.7565***				.4876***		
		(.1348)				(.1584)		
$TFP(pc10)_{t-2}$			-.0278				-.0925	
			(.0859)				(.104)	
$TFP(pc90)_{t-2}$			.4191***				.245***	
			(.0763)				(.0834)	
$TFP(pc1)_{t-2}$				-.1844***				-.1731***
				(.0539)				(.0609)
$TFP(pc99)_{t-2}$				.2258***				.1828***
				(.0335)				(.0388)
Obs.	1877	1877	1877	1877	1877	1877	1877	1877
R2	.9604	.9614	.9614	.962	.5707	.5753	.5736	.5788

All variables are in logs and lagged of two years. Each columns represent a different equation. *Exp:* export value. *TradeBal:* trade balance. *Imp:* import values. *Pc:*percentile. Estimation unit is defined by the triple country sector year. Robust standard errors are in parenthesis. We include the following groups of dummies: year, country, sector, country X year, and sector X year. Significance level: * significant at 10%; ** significant at 5%; ***significant at 1%.

ANNEX 9: THE LINK BETWEEN THE DISPERSION OF THE UNDERLYING DISTRIBUTION OF FIRMS AND AGGREGATE PERFORMANCE: EXTENSION OF THE MELITZ-OTTAVIANO (2008) FRAMEWORK[41]

Consider an industry that in period s is populated byN_s heterogeneous firms with unit labor requirement (i.e. workers per unit output) c distributed according to a cumulative density function $G_s(c)$ over the support $[0, c_s]$. The demand and inverse demand of a firm with unit labor requirement c are respectively equal to

$$q_s(c) = \frac{L}{\gamma}(c_s - p_s(c))$$

(1)

$$p_s(\omega) = c_s - \frac{\gamma}{L} q_s(c)$$

Profit maximization requires marginal revenue to match the marginal cost c (with wage equal to one). Given inverse demand (1), the FOC for profit maximization in period s by a firm with unit labor requirement c implies output

$$q_s(c) = \frac{L}{2\gamma}(c_s - c)$$

which can be plugged into total inverse demand (1) to obtain the corresponding price, markup, revenue and profit:

$$p_s(c) = \frac{1}{2}(c_s + c) \qquad \mu_s(c) = \frac{1}{2}(c_s - c)$$

(2)

$$r_s(c) = \frac{L}{4\gamma}((c_s)^2 - (c)^2) \qquad \pi_s(c) = \frac{L}{4\gamma}(c_s - c)^2$$

By (2) and unit wage, employment is

$$l_s(c) = r_s(c) - \pi_s(c) = \frac{L}{2\gamma} c(c_s - c)$$

At the industry level, revenues can be calculated as follows

---

41 We thank Gianmarco Ottaviano for the extension to the aggregate case of the Melitz and Ottaviano (2008) framework.

$$R_s = \int_0^{c_s} r_s(c) dG_s^*(c)$$

$$= \frac{L}{4\gamma} N_s \left((c_s)^2 - \int_0^{c_s} (c)^2 dG_s^*(c) \right)$$

$$= \frac{L}{4\gamma} N_s \big((c_s)^2 - (\tilde{c}_s)^2 - \tilde{\sigma}_s{}^2 \big)$$

being the mean and variance of $G_s(c)$. Note that the above result implies

$$\frac{R_s}{N_s} = r(\tilde{c}_s) - \frac{L}{4\gamma} \tilde{\sigma}_s{}^2$$

which shows that average industry revenues equal the revenues of the average firm minus a linear transformation of the variance of marginal costs.

Analogouly, it can be shown that industry profits (here also value added) and employment equal respectively

$$\Pi_s = \pi(\tilde{c}_s) + \frac{L}{4\gamma} \tilde{\sigma}_s{}^2$$

$$L_s = R_s - D_s = N_s \frac{L}{2\gamma} \big(\tilde{c}_s (c_s - \tilde{c}_s) - \tilde{\sigma}_s{}^2 \big)$$

Thus average industry profit (value added) equals the profit (value added) of the average firm plus a linear transformation of the variance of marginal costs

$$\frac{\Pi_s}{N_s} = \pi(\tilde{c}_s) + \frac{L}{4\gamma} \tilde{\sigma}_s{}^2$$

while average industry employment equals the employment of the average firm minus a linear transformation of the variance of marginal costs.

$$\frac{L_s}{N_s} = l(\tilde{c}_s) - \frac{L}{2\gamma} \tilde{\sigma}_s{}^2$$

Finally, average industry productivity equals

$$\frac{R_s}{L_s} = \frac{1}{2} \frac{(c_s)^2 - (\tilde{c}_s)^2 - \tilde{\sigma}_s{}^2}{\tilde{c}_s (c_s - \tilde{c}_s) - \tilde{\sigma}_s{}^2} = \frac{r(\tilde{c}_s) - \frac{L}{4\gamma} \tilde{\sigma}_s{}^2}{l(\tilde{c}_s) - \frac{L}{2\gamma} \tilde{\sigma}_s{}^2}$$

when computed as industry revenues per worker, while it equals

$$\frac{\Pi_s}{L_s} = \frac{1}{2}\frac{(c_s - \tilde{c}_s)^2 + \tilde{\sigma}_s{}^2}{\tilde{c}_s(c_s - \tilde{c}_s) - \tilde{\sigma}_s{}^2} = \frac{\pi(\tilde{c}_s) + \frac{L}{4\gamma}\,\tilde{\sigma}_s{}^2}{l(\tilde{c}_s) - \frac{L}{2\gamma}\,\tilde{\sigma}_s{}^2}$$

when computed as industry value added per worker. Both productivity measures are increasing functions of $\tilde{\sigma}_s{}^2$, which implies that, *for given firm averages*, productivity is higher in more heterogeneous industries. The reason is that with linear demand revenues and profits are skewed towards more productive (lower c) firms away from less productive ones. Larger variance of firm level TFP (inverse of c) implies that there is more scope for reallocating resources towards better firms and thereby increasing observed average industry productivity.