USING FIRM-LEVEL DATA TO COMPARE PRODUCTIVITIES ACROSS COUNTRIES AND SECTORS: POSSIBILITIES AND CHALLENGES

Saad Ahmad

Sarah Oliver

Caroline Peters

ECONOMICS WORKING PAPER SERIES

Working Paper 2018-07-A
U.S. INTERNATIONAL TRADE COMMISSION
500 E Street SW
Washington, DC 20436
July 2018

Using firm-level data to compare productivities across countries and sectors: possibilities and challenges

Saad Ahmad, Sarah Oliver, and Caroline Peters

Office of Economics Working Paper 2018-07-A

July 2018

Abstract

A five-year panel of cross-country data for 2012-2016 drawn from the Orbis database is used to evaluate the advantages and shortcomings of this data source in calculating firm level productivity. We find that conditional on the productivity measure employed, country and sector coverage can vary widely in the Orbis database due to different national reporting requirements across countries. This paper also compares the average productivity of the same sector across countries and the average productivity of domestic and foreign owned firms in the same sector. In every type of productivity calculation employed in this analysis, foreign firms are significantly more productive than their domestic counterparts.

Saad Ahmad
Office of Economics
saad.ahmad@usitc.gov

Sarah Oliver
Office of Industries
sarah.oliver@usitc.gov

Caroline Peters
Office of Economics
caroline.peters@usitc.gov

1. Introduction

Theoretical and empirical work have shown that the productivity of a country's firms is an important factor in determining its place in the global economy, as a country's most productive firms are more likely to become exporters (Melitz 2003). This stylized fact has led to a renewed focus in the trade literature on measuring productivity, at the sector- or firm-level, in an accurate and consistent manner. One significant hurdle impeding this research endeavor, however, is finding readily accessible databases that cover a large set of countries, industries, and firms, allowing for meaningful analysis of firm productivity dynamics across countries and sectors.

Our purpose in this paper is to demonstrate the usefulness of firm-level data from the Orbis database by constructing and analyzing measures of firm productivity at the country- and sector-level for both manufacturing and services sectors. In order to evaluate the advantages and shortcomings of Orbis for productivity analysis, we document our data coverage by country and two-digit NACE sector[1], and discuss the variety of productivity measures that can be applied to our dataset. Since different measures of productivity require different financial variables, such as assets or depreciation, country coverage varies by the method choosen to calculate or estimate productivity. Of the methodologies considered, labor productivity provides the best coverage: from 2012 to 2016, 49 countries in our data report total revenue and employment data for at least 30 firms per two-digit sector, while the Levinsohn-Petrin method for estimating total factor productivity (TFP), which requires data on intermediate inputs, had the lowest country coverage at only 27 countries in our sample.

The productivity measures computed from our dataset are also useful in demonstrating firm heterogeneity within a country and sector. Specifically, the Orbis database provides information on firm ownership, which can be used to distinguish between domestic firms and foreign-owned affiliates in a particular market. The second part of this paper finds that on average, foreign firms are significantly more productive than domestic firms in the markets where they operate. This result is consistent across all of our methodologies for calculating productivity.

The remainder of this paper is divided into four sections. The first section provides an overview of the previous literature that relies on firm-level data, including the Orbis database, to analyze productivity. The next section describes the data gathered from Orbis and compares it to other sources of firm-level data. The third section provides an overview of the methodologies used to calculate productivity.

The fourth section, which presents the results of our analysis, is divided into three parts. Part A looks at which countries are most productive and how productivity in particular sectors varies across countries. Part B examines the dispersion of productivity within a country and sector and across different estimation methods. Part C exploits Orbis' firm ownership information to examine the productivity differences between domestic and foreign firms.

[1] We use "sector" here to refer to the 2-digit Nomenclature of Economic Activities (NACE) code.

2. Literature Review

There is a growing consensus in the productivity literature around the advantages of using firm-level data in conducting productivity analysis (Bartelsman and Doms 2000; Syverson 2004; Bartelsman et al. 2009). As discussed in Bartelsman and Doms (2009), firm-level data can be used to establish stylized facts about the dispersion of productivity across firms, the uniformity of changes in productivity, the persistence of productivity differentials, the consequences of entry and exit, and the importance of changes in resource reallocation across firms to aggregate productivity growth. Firm-level data also avoids the issues that accompany productivity analysis when using sector-level data, such as adjustments made for missing establishments by applying productivity assumptions to employment statistics from labor force surveys. Input extrapolation is a frequent occurrence in the services sector where data are thinner overall (OECD 2001). As such, the firm-level nature of the database make Orbis a valuable resource for estimating the productivity of services firms in particular.

There have been several studies that have drawn upon the Orbis database to create firm-level datasets for the purposes of estimating productivity. Some recent works that rely on Orbis as their main data source include Gal (2013), which looks at OECD countries from 2000-2008; Kalemli-Ozcan et al. (2015), which looks at European firms from 1999-2012; and Gopinath et al. (2017), which looks at manufacturing firms in Spain from 1999-2012. Our paper follows in the direction of Gal (2013) who examines Orbis in the context of firm productivity analysis and proposes several imputation strategies to account for coverage issues in Orbis when measuring TFP along with other methods such as re-sampling and PPP-conversion adjustments to make these productivity measures internationally comparable.

While our dataset from Orbis is more current than the datasets used in many of these papers, the total number of years of firm data that we have pulled is less extensive, and we have only queried the database once rather than creating a panel of draws from different database vintages as in Kalemli-Ozcan et al. (2015) and Gopinath et al. (2017). Rather than attempting to build a panel of similar length to conduct more detailed time-series analysis, our primary purpose in this paper is to provide a flavor of the cross-sectional analysis of firm-level productivity that can be performed using the Orbis database.

2. Data

Bureau van Dijk's Orbis dataset reports firm-level financial data that varies in coverage based on the reporting requirements of particular countries (Bureau van Dijk 2017). For our sample, we only include firms that had non-missing revenue and employment data for 2013, 2014, and 2015, and our overall coverage of these firms includes the 5-year span from 2012-2016. We use the EU's Statistical classification of economic activities (NACE) codes to classify firms by industry at the two digit-level, and we include manufacturing sectors corresponding to codes 10-33 and service sectors codes under NACE 41-93, excluding public service (84) and banking and insurance activities (64-66) where revenue is not a good predictor of productivity. After excluding country-sector pairs where there are fewer than 30 firms in a given year, the sample consists of 4.3 million firms, 49 countries and 1,898 country-sector pairs.

To adjust for differences in prices of goods and services across countries, we convert financial variables to purchasing power parity (PPP)-adjusted figures, using the World Bank PPP conversion rate for each year. In order to calculate TFP using the index method, we take the share of labor in total output as reported in either the World KLEMS or OECD Structural Analysis (STAN) databases. A more detailed explanation of these data sources is available in appendix A3.

Among firm-level datasets, Orbis is unique in its coverage of the corporate ownership structure of firms. Collecting data from a variety of reporting sources, Orbis provides fine-grain information on a company's financials. The coverage of firms within a country depends upon reporting requirements and the difficulty of accessing information. In the subset of countries analyzed in this paper (table A.2), Bureau van Dijk reports Orbis as having at least 75 percent coverage of all firms in each country as of January 2018, with the exception of Iceland, Poland, and Luxembourg where the database only captures 50-74 percent of firms, and Greece where less than 25% percentof all firms are captured.[2] Independent verification of this coverage for past versions of the Orbis database has varied, however; in their comparison of the 2008 Orbis database to the OECD's Structural and Demographic Business Statistics (SDBS) database, Ribeiro et al. (2010) found much poorer coverage for certain countries in Orbis, while for other countries, like the United States, there were more business records in Orbis than were reported in official figures from the SDBS.

The countries explored in our analysis are almost exclusively developed countries, with a European bias. Still, compared to other firm-level datasets, Orbis's country coverage is more exhaustive and more current. The World Bank Enterprise Surveys, for example, have data for over 139 countries, but focus on firms in emerging economies, and updated their last full panel wave only as recently as 2002-2006. The European Central Bank's CompNet collects data directly from central banks and national statistical agencies, but its database is limited to 17 European countries with data from 1995-2012. Like Orbis, CompNet does not provide total firm coverage for each country in each sector.[3] The OECD created and distributed a micro-level dataset of firms in 10 countries, developed by extracting raw country-level data from a combination of business registers, enterprise surveys, social security databases, corporate tax rolls, annual industry surveys and manufacturing censuses. The World Bank supplemented this dataset with information from 14 additional countries (Bartelsman et al. 2009).

3. Methodology

This paper considers three methodologies for measuring industry-level productivity using firm-level data: labor productivity, TFP computed using the index method, and TFP estimated from firm-level data (using three different empirical frameworks). These three methodologies are used to illustrate the tradeoffs between methodological rigor and country-sector coverage when using the Orbis database to measure productivity. We briefly discuss these methodologies below, with more technical details available in Gal (2013) and Biesebroeck (2007).

[2] These figures are provided by Bureau van Dijk in their data documentation of Orbis, current as of June 2018. The full list of firm coverage by country is available here (following database login): https://help.bvdinfo.com/LearningZone/Products/orbis/Content/I_Data/Coverage/CompanyDataOverview.pdf .

[3] Some of this is simply due to the threshold for firm reporting requirements, which impact the data sources that the Orbis database aggregates as well.

a) Labor productivity

As shown in equation 1, labor productivity, or output per worker, is simply measured by dividing the operating revenue of a firm *i* by its number of employees in each year *t*. Since the dataset includes the same sample of firms in all five years, labor productivity is calculated for 2014, the midpoint in the data.

$$LaborProductivity_{it} = \frac{OperatingRevenue_{it}}{NumberofEmployees_{it}} \tag{1}$$

b) Index method

Although the simplicity of output per worker as a measure of firm productivity is appealing, it does not account for differences in other inputs across firms. As such, we need to consider measures of productivity like TFP, which accounts for the relative contribution of capital and labor to a firm's output. Following Gal (2013), this paper uses a Cobb-Douglas production function with labor and capital inputs captured by the number of employees and the value of tangible fixed assets, respectively.[4] Such a specification assumes perfect competition and constant returns to scale (Bernard and Jones 1996), and assumes that firms make input choices optimally (Biesebroeck 2007). Equation 2 shows the log linearization of this equation that we used to compute TFP for firm i at time t.[5]

$$TFP_{it} = \log(OperatingRevenue_{it}) - \alpha \log(NumberofEmployees_{it}) - (1 - \alpha) \log(TangibleFixedAssets_{it}) \tag{2}$$

The coefficient **α** shows the share of labor as an input to total operating revenue, and is based on two-digit NACE sector-level estimates of the share of labor input taken from the World KLEMS or OECD Structural Analysis Database (OECD STAN).[6] In these computations, we assume that the rest of the value of total output comes from capital[7], and also that composition of capital is similar across firms.[8] Following Gal (2013), tangible fixed assets are used to approximate capital goods in the productivity equation for this method as well as the subsequent estimation methods. As with labor productivity, we only present results for the TFP index calculated for 2014.

Index methods of calculating TFP have been found to be among the best measures for estimating productivity levels, particularly in cases when measurement error is small or there is a great deal of

---

[4]The methodology for collecting and reporting data on labor costs such as total compensation to workers varies widely between different countries and industries in Orbis. Due to these data constraints, we use the reported number of employees in Orbis as our proxy for labor input as in Gal (2013), while mindful of the fact that this may be a biased measure of firm's labor costs as it does not account for the share of part-time vs. full-time employees or different labor skill types.

[5] A non-linear version of this equation more closely resembles the labor productivity equation:

$$TFP_{it} = \frac{Operating\ Revenue_{it}}{(Number\ of\ Employees_{it})^{\alpha} * (Tangible\ Fixed\ Assets_{it})^{(1-\alpha)}}$$

[6] See appendix A3 for a complete description of how these labor shares were determined.

[7] Other contributions to total output can include the value of land and energy inputs, but we do not include these variables due to inadequate coverage of these variables in the Orbis data.

[8] By composition of capital, we mean the type, quality, and depreciation rate of equipment used in production technology, for example. Research has shown that marked differences in the composition of capital across countries can sometimes explain international variation in firm productivity levels (Caselli & Wilson 2004).

variation in the production technology across firms within a sector. In order to calculate a measure of productivity from observables, perfect competition in both input and output markets is generally assumed.[9] If factor shares within a sector vary widely across countries, however, then comparisons of the TFP index are problematic, as these factor shares may be a function of technological constraints across countries.[10] Without knowing or controlling for the level of technology, there is no way of knowing how to attribute differences in productivity to firm performance versus the capital-labor ratio of a sector, and thus there is no way of directly comparing firm-level productivities within the same sector across countries (Bernard and Jones 1996).

c) Estimation Approaches

While index methods provide a useful snapshot of the relationship between a firm's current input and the efficiency of its operations, determining TFP from estimation methods allows researchers to incorporate the dynamic nature of firm decisions that are undertaken to maximize profits. Under the estimation approaches, firm profits are a function of the inputs in preceding periods, and productivity is an unobservable, firm-level characteristic, expressed as a component of the error term of a Cobb-Douglas production function.

Individual productivity can be estimated from standard OLS residuals alone as shown in equation 3a, hereafter referred to as the pooled OLS method.[11]

$$log\ (Revenue_{it}) = \beta_l \log(NumberofEmployees_{it}) + \beta_k \log(TangibleFixedAssets_{it}) + \ \mu_{it} \tag{3a}$$

The downside to using standard OLS is that the coefficients are biased upwards since productivity levels are known by the firm, but unobserved by the researcher (Biesebroeck 2007). As firm-level input choices are likely informed by firm-level productivity, the econometric relationship that results is one where the independent variables are likely correlated with the error term (Del Gatto et al. 2011).

One way to handle this issue of simultaneity is to treat productivity as a fixed, time-invariant firm characteristic θ_i. TFP estimates can then be obtained from a fixed-effect OLS estimation of equation 3b using either least-square dummies or first-differencing methods:

$$log\ (Revenue_{it}) = \beta_l \log(NumberofEmployees_{it}) + \beta_k \log(TangibleFixedAssets_{it}) + \theta_{it} + \mu_{it}$$
$$\text{where} \quad \mu_{it} = TFP_i + \ \varepsilon_{it} \tag{3b}$$

[9] Ideally, some adjustment would be made to account for the nature of scale economies within the industry as well, but it is common practice to leave this unaddressed (Biesenbrock 2007).
[10] The variations in this share across countries could be a function of methodological differences in how these shares are calculated (though it should be mostly consistent, as explained in appendix A3), or cross-country differences in factor allocations. See figure B.1 in appendix B for labor shares across countries for select sectors.
[11] TFP here is the exponentiated difference between fitted and observed values of the dependent variable.

However, the assumption that firm productivity does not change over time has been proven false in the extensive literature documenting the effects of technical improvements and technological innovation on firm-level productivity (see Cardarelli and Lusinyan (2015) and Heshmati and Rashidghalam (2016) for recent examples). By holding firm TFP fixed over time, the fixed effect estimation framework is not able to account for productivity shocks, like changes to regulations at the industry-level or the breakdown of machinery at the firm-level. As such, researchers who employ an estimation strategy to model firm-level productivity as time invariant will have to accept that the accuracy of their TFP estimates may not hold over longer time periods, which in turn limits the usefulness of these estimates in empirical applications.

Olley and Pakes (1996) provide a semi-parametric framework to address the simultaneity concerns in TFP estimations arising from the fact that variations in productivity are known by the firm, but unobservable in the data.[12] They account for the unobserved productivity by treating the firm's investment behavior as a state variable in the firm's dynamic optimization problem that depends on the firm's level of capital and productivity. Thus, the firm's investment decisions can be used as a proxy for unobserved time-varying shocks to productivity. Following Gal (2013), we calculate firm investment by the perpetual inventory method so that capital in the current period is the sum of investment in the current period and the capital in the previous period less depreciation δ_{it-1}:

$$Investment_{it} = TangibleFixedAssets_{it} - TangibleFixedAssets_{it-1} * (1 - \delta_{it-1}) \quad (4)$$

The Olley-Pakes method then uses a two-step procedure for estimating TFP. In the first stage, a function of investment and capital is used to control for unobserved productivity:

$$\log(Revenue_{it}) = \beta_l \log(NumberofEmployees_{it}) + \boldsymbol{f}(\log(Investment_{it}), \log(TangibleFixedAssets_{it})) + \varepsilon_{it} \quad (5a)$$

Here ε_{it} is the idiosyncratic error term for unexpected shocks to firm's revenue. Since the function f() is unknown, the Olley-Pakes method uses a third or fourth order polynomial in investment and capital as an approximation during the estimation to get consistent estimates of the labor elasticity β_l.

In the second stage, the estimated values of β_l and residuals e_{it} from the first stage are used to get consistent estimates of the capital elasticity β_k. The Olley-Pakes method makes use of the fact that the fitted value $\hat{f}$ of the function f() is just the actual revenue minus the residuals and β_l times the number of employees. We can then estimate β_k from equation (5b) using non-linear squares:

$$\log(Revenue_{it}) = \beta_l \log(NumberofEmployees_{it}) + \beta_k \log(TangibleFixedAssets_{it}) + \boldsymbol{g}(\hat{f}_{it-1} - \beta_k \log(TangibleFixedAssets_{it-1})) + \zeta_{it} + \boldsymbol{\varepsilon}_{it} \quad (5b)$$

[12] When applied to a longer panels with information on firm exit, the Olley-Pakes method is also effective at controlling for selection bias (i.e. the fact that firms with higher capital stocks are more likely to remain in the sample after negative productivity shocks) (Gal 2013).

Here ζ_{it} is an unexpected innovation that is uncorrelated with productivity and capital in period t. As in the first stage, the unknown function g() in equation (5b) is treated as a nonparametric term and approximated by a third or fourth order polynomial. The two stages gives us consistent estimates of both β_l and β_k that can be used to compute the firm's TFP:

$$\log(TFP_{it}) = \log(Revenue_{it}) - \beta_l \log(NumberofEmployees_{it}) - \beta_k \log(TangibleFixedAssets_{it}) \quad (6)$$

Levinsohn and Petrin (2003) extend the Olley-Pakes framework by incorporating intermediate inputs, such as electricity or materials, instead of investment as proxies for unobserved time-varying shocks. Including intermediate inputs in the estimation may be preferable to investment due to the comparative smoothness with which they respond to production shocks.[13] Further, unlike Olley-Pakes, this approach can account for periods with zero reported investment, or if the costs to capital adjustment are non-convex.[14] Accounting for intermediate inputs also becomes important when using gross output (as we do in this paper) rather than value-added measures in the TFP estimations. The coverage of intermediate inputs in Orbis is uneven across countries, however, with firms in some countries like the United States not recording material costs as a separate account. In those instances, the Olley-Pakes method is the better option for estimating TFP using data from the Orbis database.

Table 1 summarizes the data requirements for each of the five productivity methods, and gives the number of countries for which the necessary data are available, as well as the number of country-sectors with at least 30 firms with the necessary data.[15] Not surprisingly, labor productivity has the widest coverage across countries and sectors followed by the TFP estimations using simple OLS using pooled estimates or fixed effects. Coverage drops for estimations that require more information: labor share for the index method, investment for Olley-Pakes, and material costs for Levinson-Petrin.[16] As discussed, the Levinson-Petrin method will not be a feasible alternative to Olley-Pakes for a number of countries in our sample and thus we only present results for Olley-Pakes in the sections that follow.

[13] This preference holds under the assumption that intermediate inputs are less costly to adjust than capital.

[14] "Non-convex" adjustment costs are the result of lumpy, mostly non-zero investments. Such marked variations in a firm's level of capital investment are not uncommon. For instance, Doms and Dunne (1998) found in a study of 13,700 manufacturing plants over 16 years that nearly a quarter of a firm's investment over that period is explained by one large investment.

[15] See table B.1 in appendix B for a breakdown of four-digit NACE subsector coverage by country across methodologies.

[16] Following Gal (2013), we consider material costs in our Orbis dataset to be a proxy for the cost of intermediate inputs.

Table 1. Data Requirements and Coverage for Productivity Measures

Methods	Variables Required (and source)	Number of years of data required	Countries with necessary data	Country-sectors with necessary data
Labor Productivity	Total firm revenue (Orbis) Number of employees (Orbis)	1	49	1898
Index Method	Total firms revenue (Orbis) Number of employees (Orbis) Tangible fixed assets (Orbis) Labor Share (KLEMS/OECD STAN)	1	34	1413
Estimation Methods				
(a) OLS/FE	Total firms revenue (Orbis) Number of employees (Orbis) Tangible fixed assets (Orbis)	3 years	47	1722
(b) Olley-Pakes	Total firms revenue (Orbis) Number of employees (Orbis) Tangible fixed assets (Orbis) Depreciation rate (Orbis)	3 years	43	1445
(c) Levinsohn-Petrin	Total firms revenue (Orbis) Number of employees (Orbis) Tangible fixed assets (Orbis) Materials cost (Orbis)	3 years	27	1156

Source: Authors' calculations using data from Bureau van Dijk's Orbis database

4. Results

a) Productivity at the country and sector levels

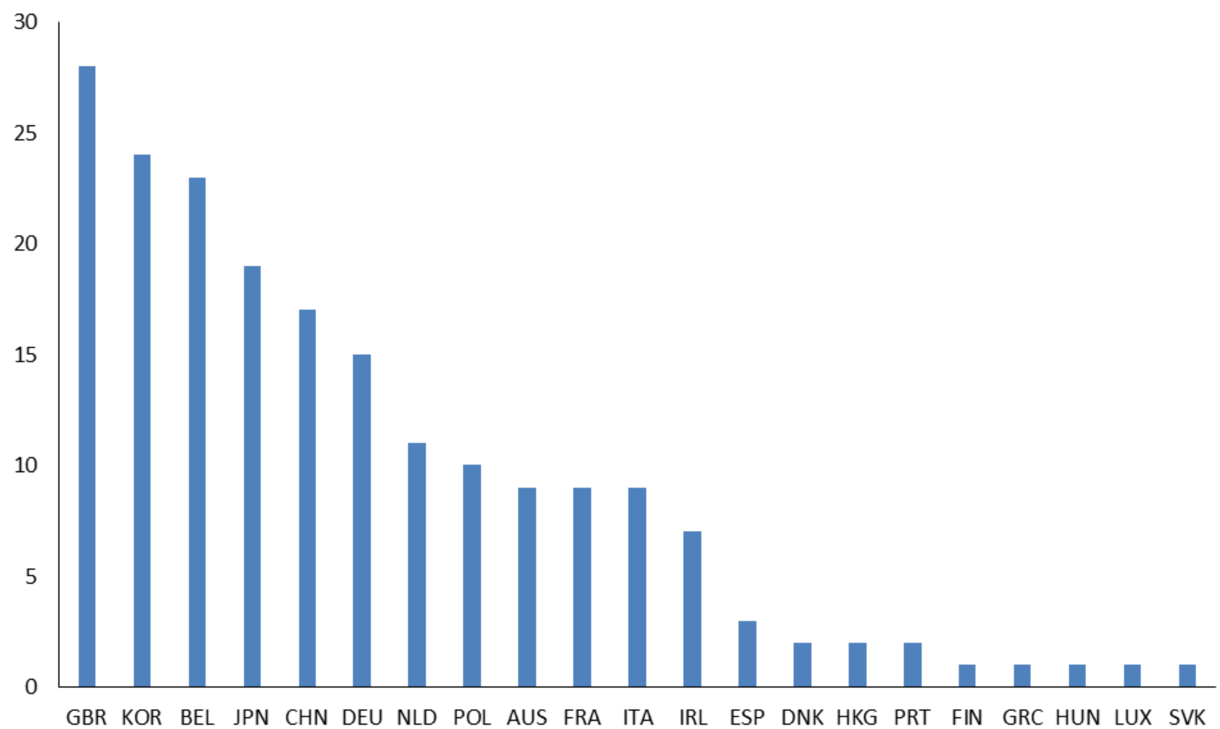
We start our analysis by looking at the differences in the computed productivity measures across countries. Using simple averages, we summarize firm productivity at the NACE 2-digit sector-level for each country in our sample. We focus on labor productivity in this analysis since of the three types of productivity measures discussed in the paper, labor productivity provides the widest country and sector coverage. Further, labor productivity may be more easily compared across countries than more sophisticated measures such as TFP as it requires less information about a firm's capital stock. Although using PPP-adjusted output data does help correct for cross-country differences in prices of goods and services, we note that differences in sectoral allocations within national economies still make labor productivity an imperfect measure of productivity differences across countries.[17]

We start by identifying the most productive countries in each of the 65 NACE 2-digit sectors in our dataset. Figure 1 shows the number of times a country, based on its average labor productivity, is ranked as a top 3 country for a given sector. For brevity, figure 1 only includes countries that have been

[17] Differences in productivity have also been attributed to country-specific institutions and policies that provide firms with a comparative advantage in their sector (Hall and Jones 1999). It is worth keeping this and other country-level sources of variation in mind when considering our results.

ranked as a top 3 country in at least 1 2-digit NACE sector.[18] We find that highly-developed countries such as the United Kingdom (28 times) , Korea (24 times), and Belgium (23 times) dominate these rankings, with nearly half of all possible spots taken by these three countries.[19] It is not surprising that these advanced economies are the most productive for a large number of sectors, although China (17 times) is quickly becoming a strong competitor in a number of sectors. By contrast, we see smaller and less advanced European countries like Hungary and Greece ranking among the most productive countries for only a single sector.

Figure 1: Number of Sectoral Top 3 Rankings by Country (Labor Productivity, 2014)



Source: Authors' calculations using data from Bureau van Dijk's Orbis database

As discussed above, differences in sectoral composition and factor prices may limit cross-country comparisons of productivity. However, we can also use our constructed productivity measures to examine differences across sectors within a country, and thus identify the top and bottom performing sectors in terms of output per worker.

[18] Overall, 20 countries in our sample never appear as a Top 3 country in any 2-digit NACE sector, however some of this is a result of inadequate coverage in Orbis. For instance, we only have labor productivity measures in 17 sectors for American firms out of a possible 65 sectors.

[19] Note that the maximum number of times a country can be in a top 3 list is 65.

Table 2 shows the top and bottom 3 manufacturing 2-digit sectors at the country-level. In the table, we only include the countries that had coverage on labor productivity in more than 8 sectors in the dataset. We see a fair degree of heterogeneity in the top 3 manufacturing sectors among countries in the sample, with food and paper products ranking among the most productive sectors in countries like Australia and Serbia that have a strong agricultural base and natural resource endowments; while basic metals, chemicals, and motor vehicles are found to be the most productive sectors in countries with strong manufacturing bases such as Italy, China, and Korea. For the bottom 3 sectors, we see less heterogeneity with wearing apparel, textiles, and furniture ranking among the least productive sectors for a number of countries.

Table 2: Top and Bottom Two Digit NACE Sectors by Country (Manufacturing)

	Top 3 Manufacturing			Bottom 3 Manufacturing		
AUS	food products	paper	coke/refined petroleum	repair/install of machinery	**furniture**	motor vehicles
AUT	repair/install of machinery	rubber/ plastics	machinery/ equipment	furniture	food products	fabricated metal products
BEL	computer, electronic	chemicals	food products	recorded media	machinery/ equipment	other non-metallic
BGR	electrical equipment	chemicals	motor vehicles	other mnf	wearing apparel	furniture
CHN	basic metals	motor vehicles	coke/refined petroleum	furniture	rubber and plastics	wearing apparel
CZE	computer, electronic	paper	motor vehicles	wearing apparel	other non-metallic	beverages
DEU	paper	chemicals	coke/refined petroleum	other mnf	furniture	wood products
ESP	pharmaceutical	chemicals	motor vehicles	furniture	recorded media	fabricated metal
EST	other mnf	fabricated metal	repair/install of machinery	wearing apparel	textiles	wood
FIN	rubber and plastics	electrical equipment	chemicals	wearing apparel	textiles	leather
FRA	pharmaceutical	beverages	other non-metallic	recorded media	other mnf	repair/install of machinery
GBR	wood products	beverages	chemicals	recorded media	furniture	fabricated metal products
HRV	electrical equipment	computer, electronic	paper	wearing apparel	furniture	leather
HUN	pharmaceutical	chemicals	computer, electronic	wearing apparel	other mnf	furniture
ITA	pharmaceutical	basic metals	coke/refined petroleum	recorded media	fabricated metal	furniture
JPN	beverages	wood	coke/refined petroleum	recorded media	repair/install of machinery	other transport equipment
KOR	basic metals	chemicals	coke/refined petroleum	other transport equipment	other mnf	recorded media
LTU	fabricated metal	wood products	rubber and plastics	wearing apparel	other non-metallic	other mnf
LVA	rubber and plastics	food products	paper	wearing apparel	textiles	furniture
MKD	paper	beverages	electrical equipment	leather	furniture	repair/install of machinery
PRT	motor vehicles	chemicals	pharmaceutical	furniture	wearing apparel	other mnf
ROU	motor vehicles	basic metals	coke/refined petroleum	wearing apparel	leather	other mnf
RUS	basic metals	motor vehicles	coke/refined petroleum	leather	recorded media	wood products
SRB	paper	food products	chemicals	wearing apparel	textiles	other mnf
SVK	other non-metallic	computer, electronic	motor vehicles	wearing apparel	food products	furniture
SVN	rubber and plastics	machinery and equipment	computer, electronic	wearing apparel	other non-metallic	furniture
SWE	machinery and equipment	other non-metallic	paper	wearing apparel	recorded media	repair/install of machinery
UKR	basic metals	electrical equipment	coke/refined petroleum	wearing apparel	other mnf	furniture

b) Productivity distribution across countries and sectors

We next examine the difference in the distribution of productivities across manufacturing and service sectors within each country. A number of studies have documented large amounts of heterogeneity across firms in terms of their productivity and have explored the key factors behind this heterogeneity within the framework of firm behavior (Bartelsman et al. 2013). Our goal in this section is to establish whether a similar heterogeneity in productivity exists within our dataset.

Figure 2 shows the distribution of labor productivities across manufacturing (NACE codes 10-33) and services sectors (NACE codes 41-93, excluding codes 64-66 and 84).[20] In manufacturing (top panel) countries with the largest interquartile range (the middle 50 percent of firms from that country in the sample) are the Netherlands and Ireland, while in services (bottom panel), Luxembourg and China have the largest interquartile ranges. This suggest that within manufacturing, there is more room for aggregate productivity increases in the Netherlands and Ireland as resources get reallocated from less productive to more productive firms, while Luxembourg and China have the most room for aggregate productivity increases in services. Other countries, such as Russia, have tighter productivity distributions within manufacturing and services, indicating fewer possible productivity gains from reallocation.

We next turn to TFP measures of productivity in order to control for capital inputs across countries and sectors. Figure 3 shows the dispersion in TFP computed by the index method across countries. Here we find a similar pattern as in figure 2. Ireland, the Netherlands, and Belgium have the biggest inter-quartile range for manufacturing, while Lithuania and Romania have the smallest levels of dispersion in computed TFP.[21] In services, China and Luxembourg stand out again with very dispersed productivities.

Using estimation methods for obtaining TFP values, we observe that the levels of dispersion for manufacturing and services firms change, and are no longer consistent across methodologies.[22] One explanation for this result could be the variance in the sample of firms across estimation methods, which may be causing changes in estimated TFP values and the level of dispersion across countries.

While figures 2 and 3 show productivity distribution at the country level, we can also compare productivity distribution across sectors for our sample countries. Figure 4 compares the distribution of labor productivity for manufacturing sectors in Germany (top panel) and Russia (bottom panel), two countries with very different levels of economic development that also are well represented in the data. In manufacturing, the distribution of productivities across two-digit NACE sectors are similar: in both countries, coke and refined petroleum products have high average output per worker and a wide distribution of firm productivities. Other sectors that see high levels of dispersion include chemicals, pharmaceuticals, and food products for Germany, and basic metals and motor vehicles for Russia. Leather and wood products have less dispersion in both countries.

[20] Codes 64-66 represent banking and insurance services, and 84 represents public administration and defense services. See appendix A1 for more information on the composition of industries in the dataset.

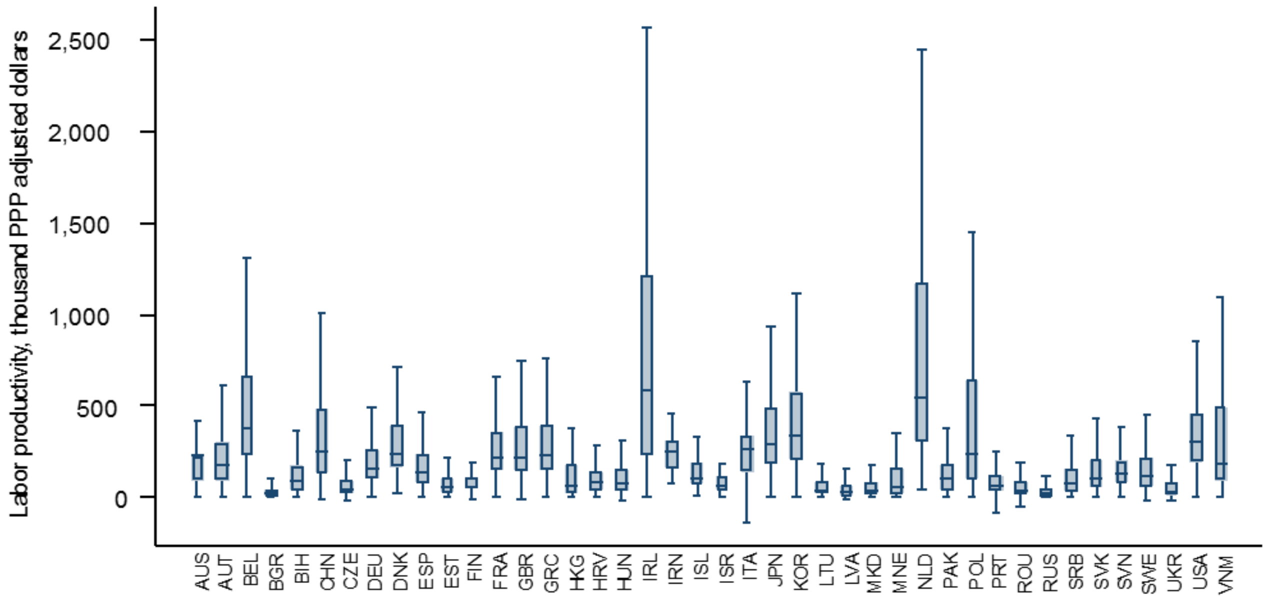
[21] As noted in Bernard and Jones (1996), the labor share used in this method varies across countries, and so it may be misleading to use this specific measure to compare firm-level productivities across countries.

[22] See appendix B for figures showing the dispersion across countries for estimated TFP values.

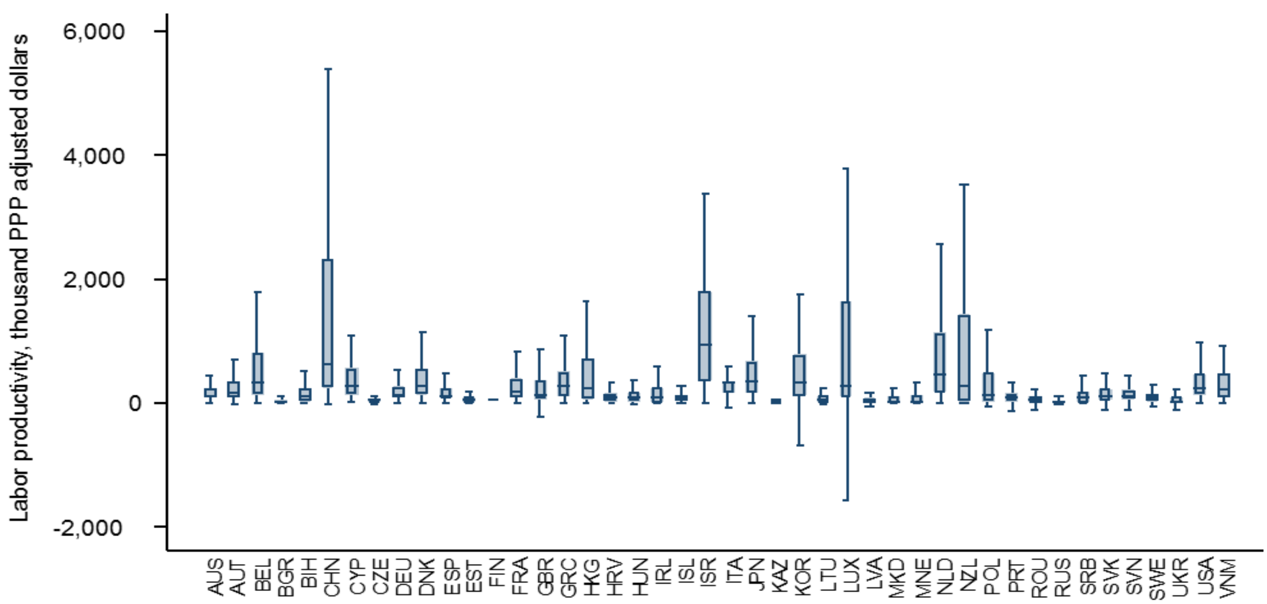
Figure 5 compares the distribution of labor productivity for services sectors in Germany (top panel) and Russia (bottom panel). We see that in services, there is a fairly even distribution of labor productivity in Russia, while in Germany, water transport has by far the highest level of dispersion and average output per worker. Real estate and broadcasting are other German sectors that have relatively high levels of dispersion.

Figure 2: Labor productivity distribution in Manufacturing and Services by Country, 2014.

Manufacturing (NACE 10-33)



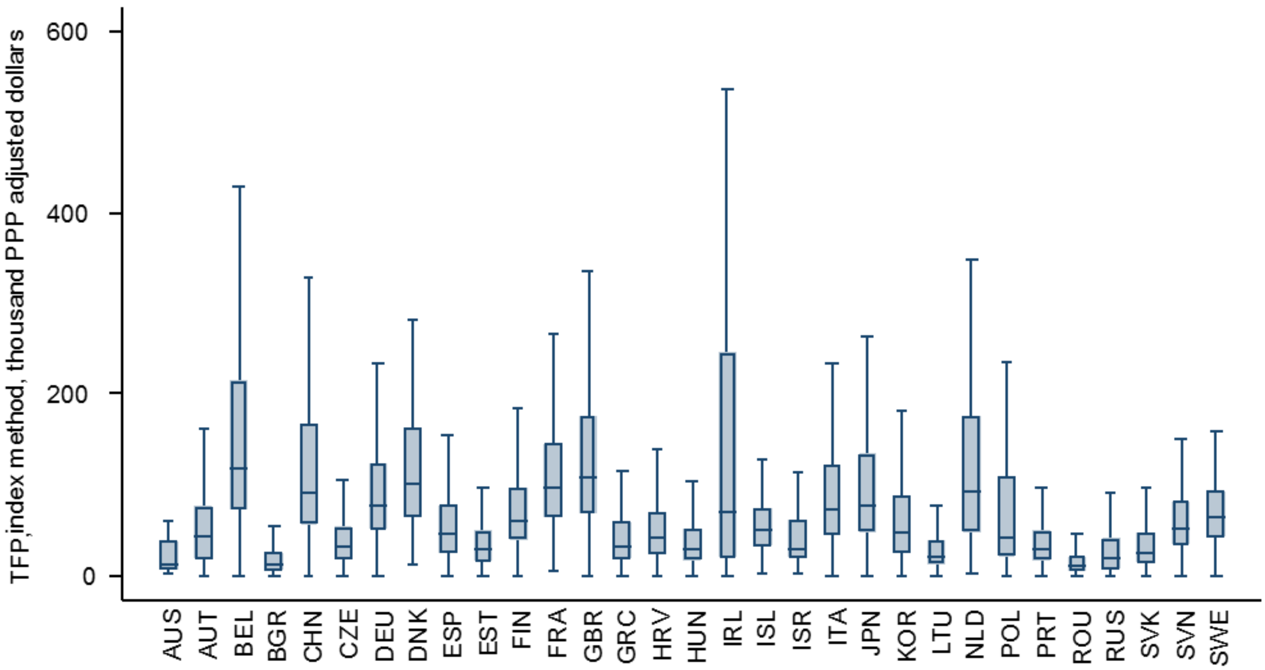
Services (NACE 41-64, 66-83, 85-99)



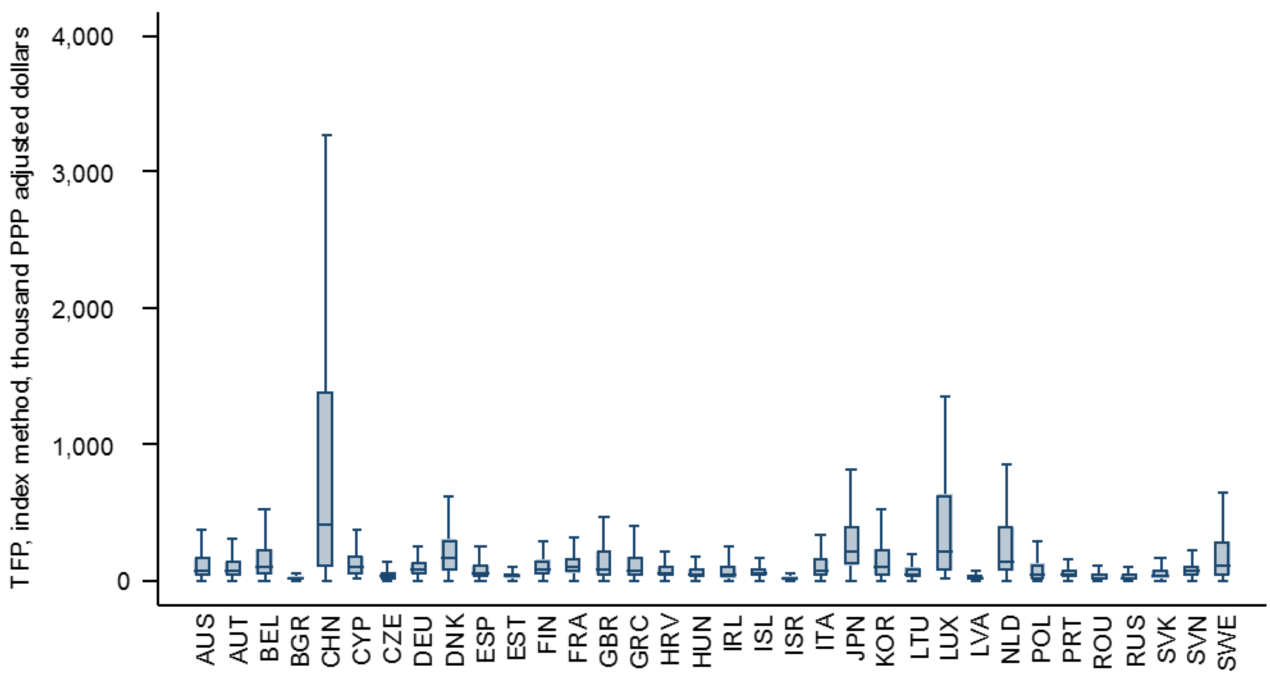
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure 3: TFP distribution (index method) in Manufacturing and Services by Country, 2014.

Manufacturing (NACE 10-33)



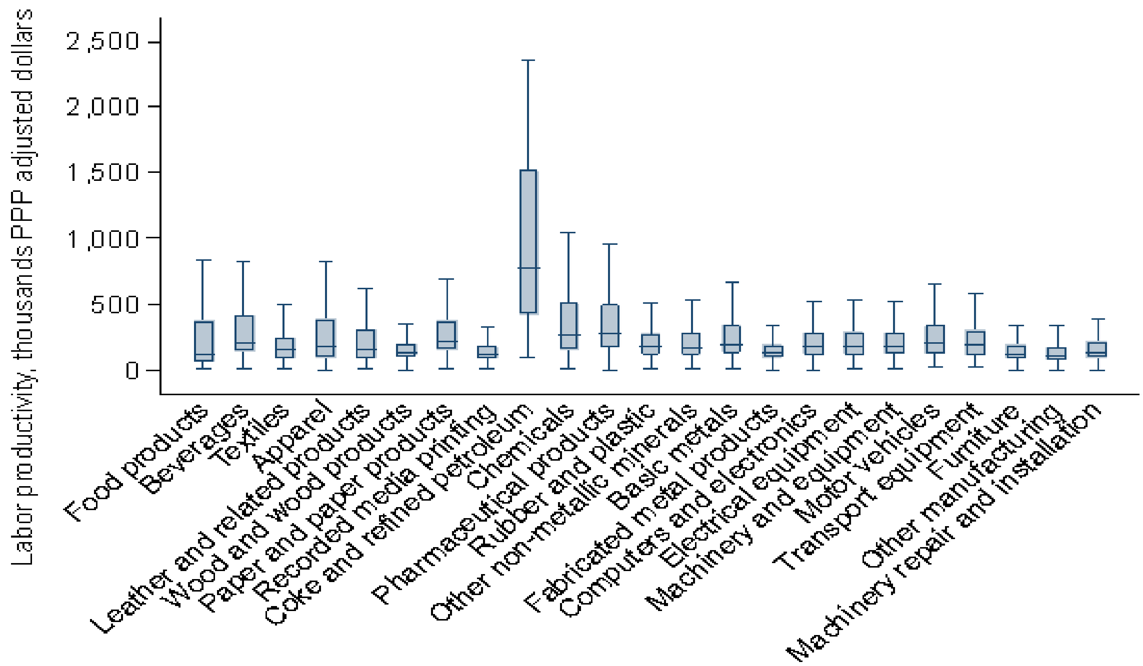
Services (NACE 41-64, 66-83, 85-99)



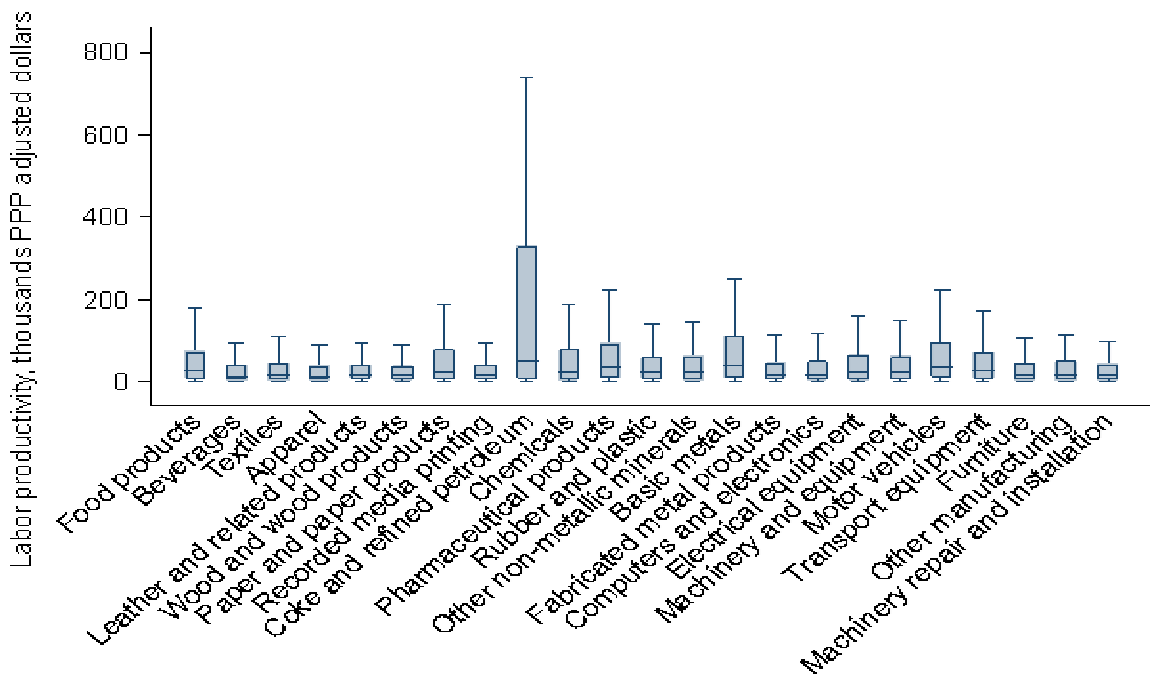
Source: Authors' estimates using data from Bureau van Dijk's Orbis database. Contact authors for information on labor share estimates used for individual country-sectors.

Figure 4: Labor productivity by manufacturing sector, 2014.

Germany



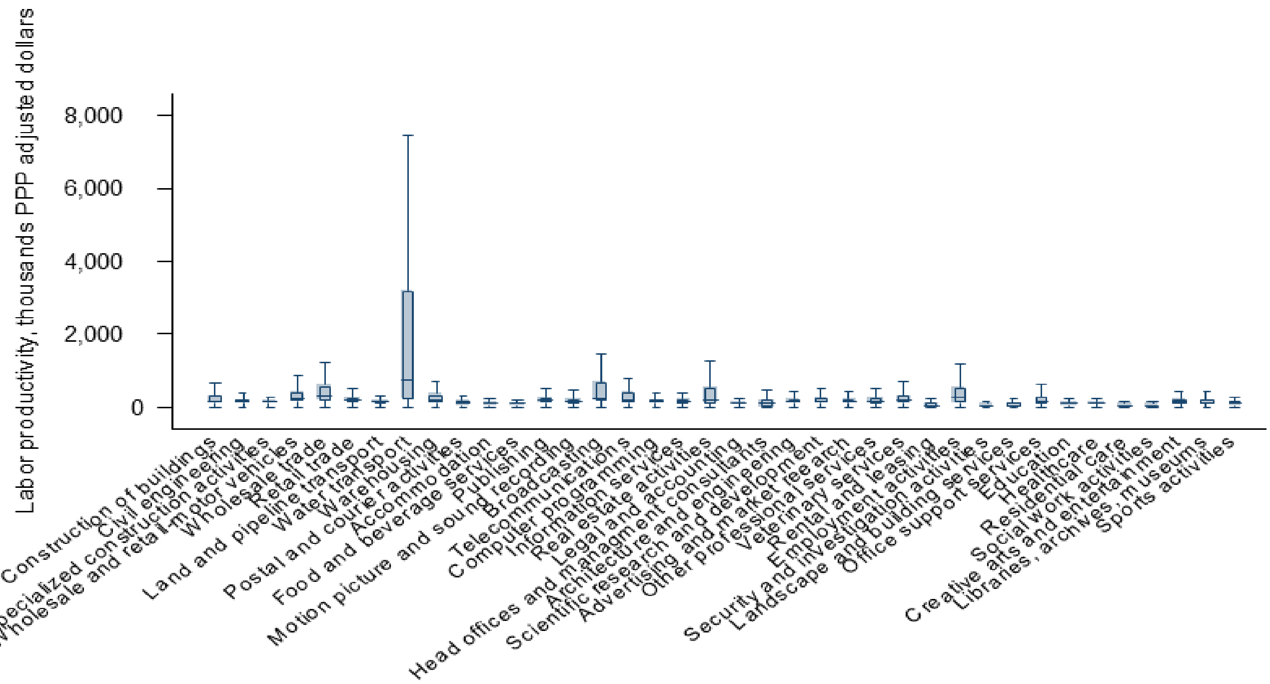
Russia



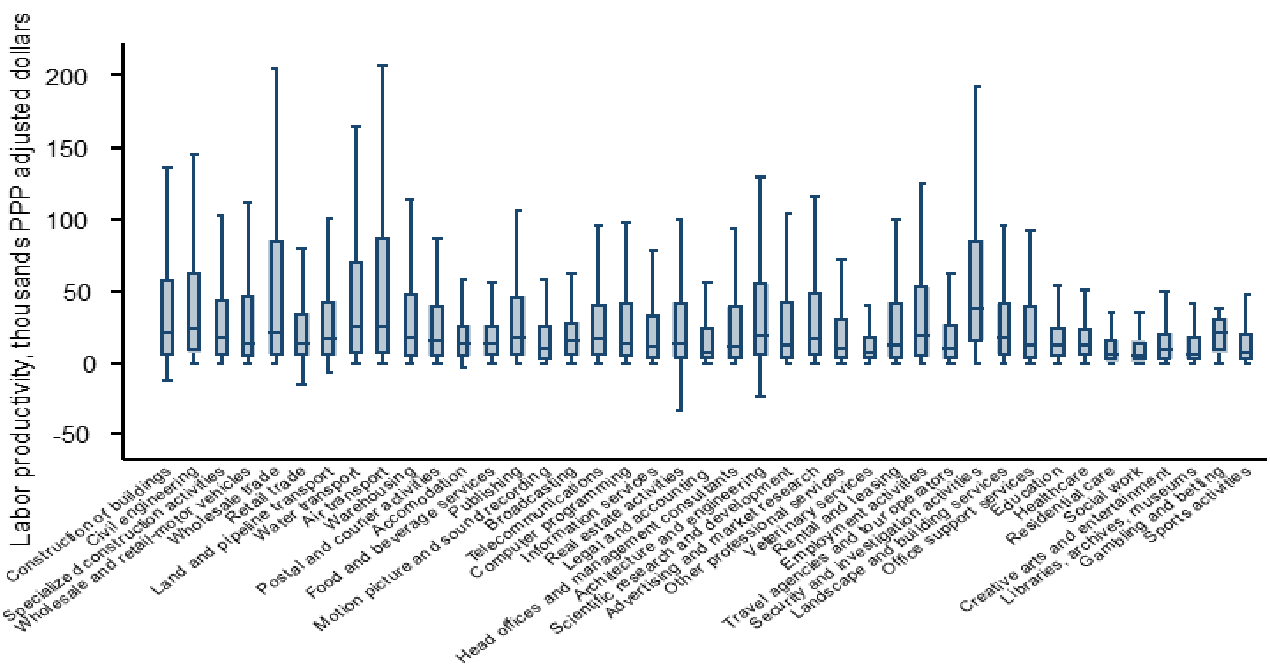
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure 5: Labor productivity by services sector, 2014.

Germany



Russia



Source: Authors' estimates using data from Bureau van Dijk's Orbis database

c) Foreign vs Domestic firms

While previous sections of this paper demonstrated the usefulness of Orbis firm-level data for constructing and comparing sector-level measures of productivity, we now use the firm-specific characteristics to better understand differences in productivity across categories of firms within specific countries and sectors. Modern trade theory predicts that only the most productive firms will be able to make an investment to set up operations in foreign countries (Helpman et al., 2004) and so we can use our computed productivity measures to test if this holds true for the firms in our dataset. The additional benefit of examining foreign and domestic firms within a specific country, is that the foreign firms in a particular market should face the same prices for factor inputs as domestic firms, so even non-PPP adjusted financial variables are comparable. We use both a two-sample t-test of means and a Kolmogorov-Smirnov (K-S) test of distribution to compare the two types of firms across all of our estimated productivity variables. For each of the estimation methods used in this analysis, foreign firms are significantly more productive than domestic firms.

To illustrate the differences in domestic and foreign firm productivity, we compare labor productivity of foreign and domestic firms in two sectors: computer programming, consultancy, and related activities (NACE 62) and manufacture of fabricated metal products, except machinery and equipment (NACE 25).[23] We choose to focus on these two particular sectors because of the relatively high number of countries that meet the data requirements to calculate labor productivity in these sectors (26 and 23 countries, respectively), and because these are two sectors in which Orbis provides relatively good coverage in terms of the overall employment.[24] Overall, there are approximately 29,000 firm observations in the manufacture of fabricated metal products, and 20,000 firm observations in computer programming.

In Orbis, the name and country of origin of the global ultimate owner (GUO) of a firm is provided if that GUO controls at least 50 percent of a company observation. If the GUO is located in a different country than the firm, that firm is considered foreign. Domestic firms are those with a GUO located in the same country. Firm observations for the GUO itself, or firm observations that do not have any ownership information are excluded from this analysis. For more information on the classification of foreign and domestic firms, see appendix A2.

Table 3 provides summary statistics for the sample of firms in NACE code 25 and 62, separated by foreign and domestic firms. In both cases, the sample includes more domestic firms than foreign firms, but there are more than 1,500 foreign firm observations in each sector. In both sectors, foreign firms have higher revenues, more employees and higher-value tangible fixed assets than domestic firms.

---

[23] See table B.2 in appendix B for the list of sectors covered under these two NACE subheadings.

[24] Overall coverage of employment was measured by comparing the total employment by country-sector in 2012 as reported in the KLEMS data, to the total employment of all the firms in our dataset for the same country-sector in 2012. Across countries, computer programming firms in our dataset on covered 33.6 percent of total sector employment on average, while firms from the manufacture of fabricated metal products sector covered 27.4 percent of total sector employment on average.

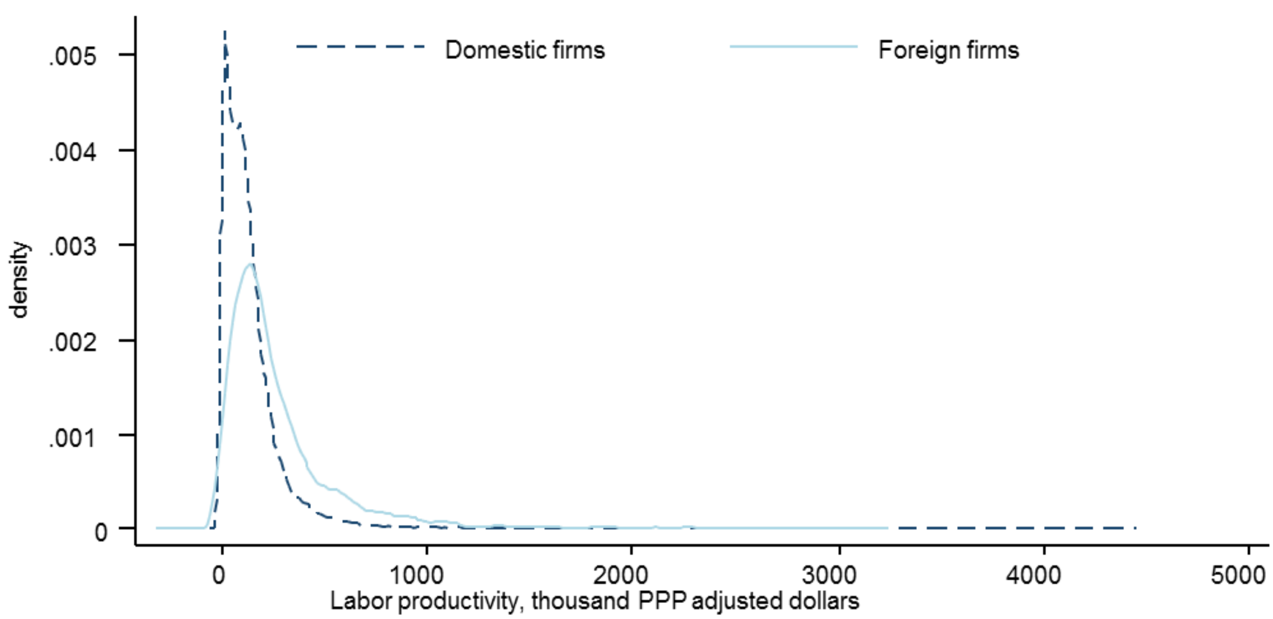
Table 3: Summary statistics for foreign and domestic firms in NACE codes 25 and 62

		Number of employees		Firm revenue (PPP-adjusted)		Tangible Fixed Assets (PPP-adjusted)	
	Number of firms	Mean	SD	Mean	SD	Mean	SD
Manufacturing Sector: NACE 25 - Manufacture of fabricated metal products, not machinery and equipment							
Foreign	1,791	129.8	(569.4)	$41,237	($151,568)	$8,364	($31,751)
Domestic	27,222	31.5	(215.9)	$8,582	($115,443)	$2,026	($22,974)
Services Sector: NACE 62 - Computer programming, consultancy, and related activities							
Foreign	2,982	166.6	(665.9)	$63,109	($503,011)	$3,723	($31,930)
Domestic	16,884	35.8	(441.9)	$8,310	($104,679)	$1,261	($39,886)

Source: Authors' calculations using data from Bureau van Dijk's Orbis database

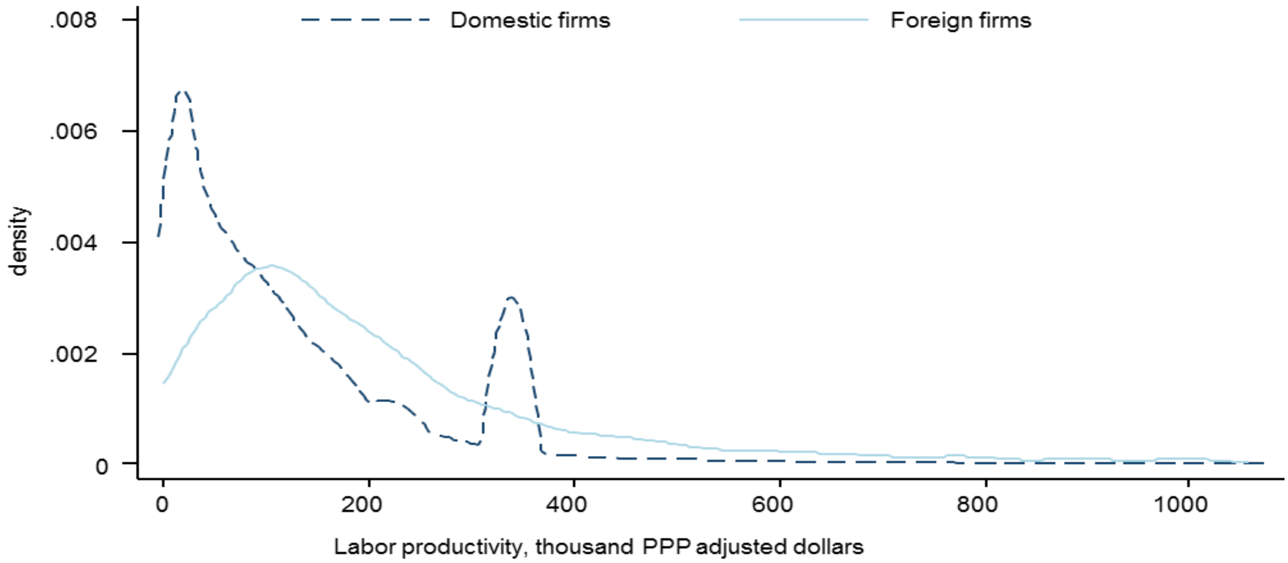
The difference in labor productivity between foreign and domestic firms is also apparent looking at the data at the firm and country level. Figure 6 compares the distribution of domestic and foreign firms' labor productivities, pooled across countries in the fabricated metals sector. In that sector, domestic firms tend to be more tightly concentrated at lower productivity levels than foreign firms. In computer programming, the difference in the distribution of labor productivities between domestic and foreign firms is even more pronounced, as shown in figure 7. Again, domestic firms tend to be more tightly concentrated at lower productivity levels than foreign firms.

Figure 6: Distribution of productivity by firm ownership (Fabricated metals – NACE 25)



Note: For clarity, excludes firms in the top 95th percentile of observations in this sector.
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure 7: Distribution of labor productivity by firm ownership (Computer Programming – NACE 62)

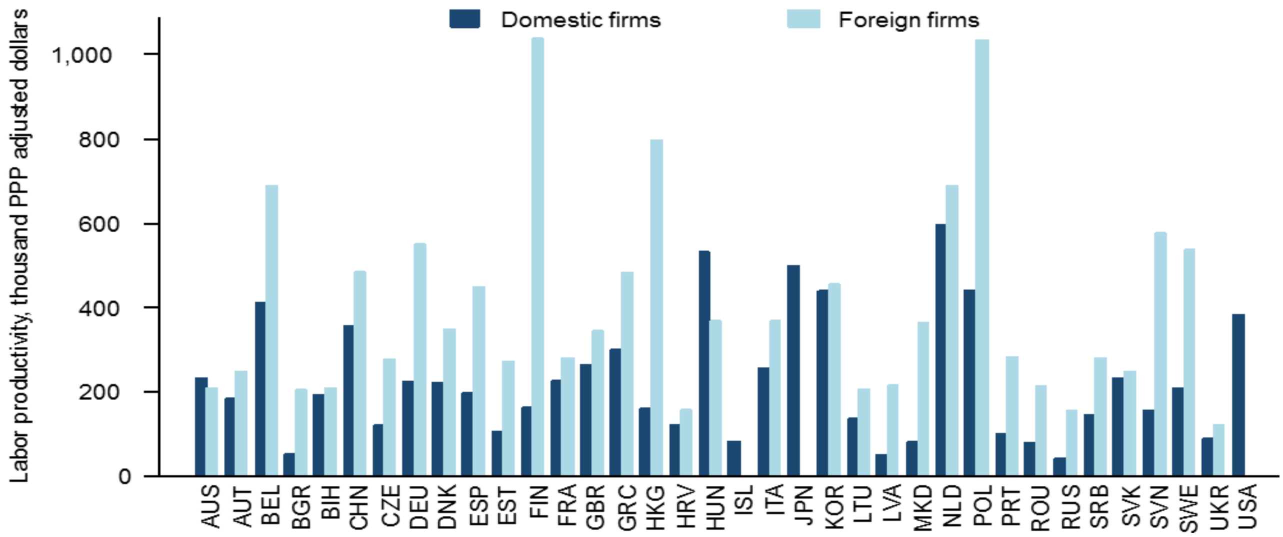


Note: For clarity, excludes firms in the top 95th percentile of observations in this sector.
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

At the country level, average labor productivity of foreign and domestic firms also lend support to the idea that foreign firms tend to be more productive than domestic firms. In the fabricated metals sector, labor productivity is only higher for domestic firms than for foreign firms in 2 of the countries covered in the sample: Australia and Hungary. Figure 9 compares the average labor productivity in computer programming for foreign-owned and domestic companies across countries. Productivity outliers (Ireland and the Netherlands) left-skew the labor productivity distribution of output per worker.[25] In every country but Hungary, Japan, and the United States, however, the average productivity of foreign firms exceeds the average productivity of domestic firms. This result is not surprising given the prominence of U.S. and Japanese companies in computer services.

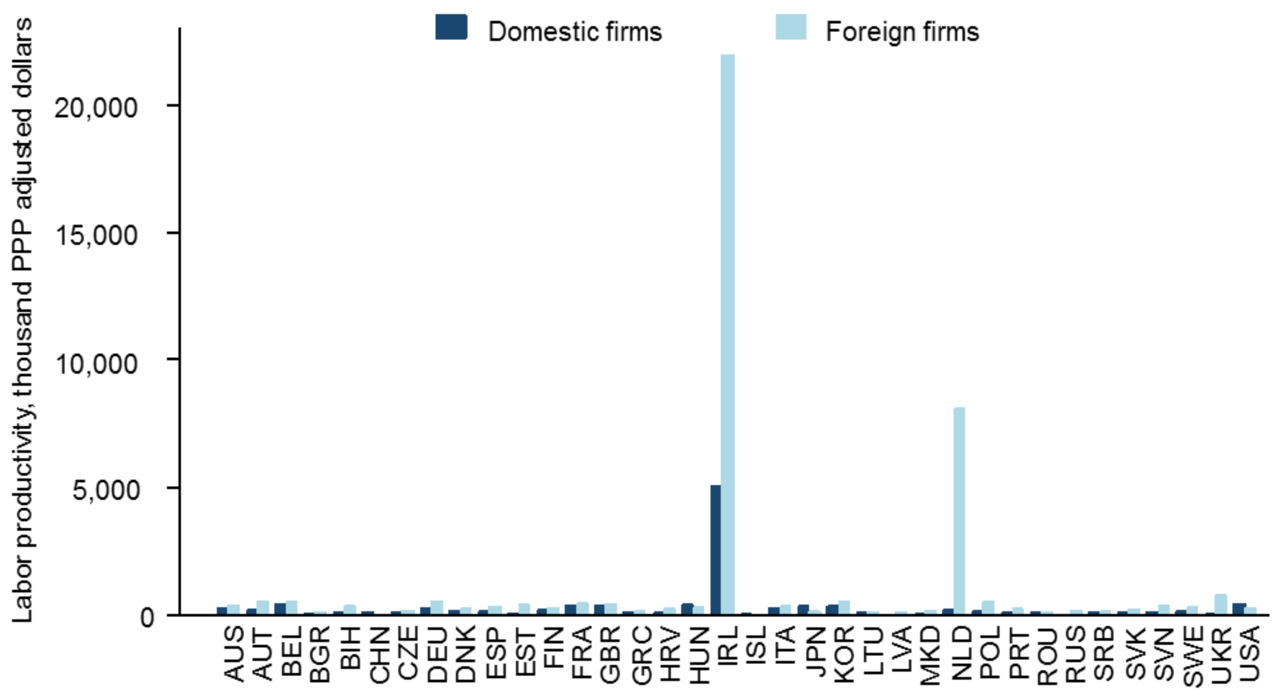
[25] It is likely that large tech firms like Google who have headquarters in Ireland are driving this result.

Figure 8: Average labor productivity by country (Fabricated Metal – NACE 25)



Note: There are no foreign-owned fabricated metal firms in U.S, Iceland, or Japan in our dataset Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure 9: Average labor productivity by country (Computer Programming – NACE 62)



Source: Authors' estimates using data from Bureau van Dijk's Orbis database

To test the relationship between foreign ownership and firm productivity, we use the two-sample t-test and the K-S test, which are sensitive to differences in both the mean and the shape of the distribution of the two samples. The two-sample t-test indicates whether there is a significant difference in the average labor productivity across domestic and foreign firms, while the K-S test indicates whether the distribution of the two samples is significantly different. Table 4 presents results for all of the methods for calculating and estimating firm level productivity used in this paper.

Table 4: Two-sample tests for all reporting countries in 2014

	Foreign			Domestic				
	Countries	**Mean TFP**	**Firms**	**Countries**	**Mean TFP**	**Firms**	**Difference in Means**	**K-S Stat**
Services Sector: NACE 62 - Computer programming, consultancy, and related activities								
Labor Productivity	35	714.69	4,367	33	166.98	39,168	547.00***	0.25***
Index Method	26	4.8	2,982	27	4.0	16,884	0.79***	0.31***
OLS Method	32	29.53	3,344	33	15.09	18,554	14.44**	0.14**
OLS FE Method	32	5.02	3,344	33	1.82	18,554	3.20***	0.36***
Olley-Pakes Method	26	496.06	2,134	28	257.09	7,135	239.00***	0.24***
Manufacturing Sector: NACE 25 - Manufacture of fabricated metal products, not machinery and equipment								
Labor Productivity	29	339.08	2,318	31	196.15	52,893	142.92***	0.19***
Index Method	22	4.1	1,791	24	3.84	27,222	0.25***	0.18***
OLS Method	27	10.07	1,959	29	6.94	29,388	3.13***	0.22***
OLS FE Method	27	4.31	1,959	29	1.59	29,388	2.72***	0.39***
Olley-Pakes Method	22	456.35	1,297	24	251.29	13,867	205.06***	0.14***

Note: Both GUOs and firms with no ownership are excluded from calculations in the table above. For estimation-based productivity methods, we dropped results where the estimated labor or capital coefficient in the regression (i.e. the labor and capital elasticities within a sector) were less than zero.
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

The higher productivity of foreign-owned firms persists, regardless of the method of estimation. The trend is common to both the services and manufacturing sectors we are exploring. Further, under all productivity estimation methods, the difference in the mean productivity between foreign and domestic firms is larger within the computer programming and consultancy sector than within the fabricated metal products sector. Overall, the finding that foreign-owned firms are, on average, more productive than domestic firms, is consistent with modern trade theory and indicates that foreign-owned firms are present in a country because they are able to compete, at least in terms of productivity, with domestic firms.

5. Conclusion

This paper considers the utility of firm-level data for constructing cross-country and sector measures of firm productivity. Variation in productivity estimates across the three methods considered in this analysis show that country and sector coverage in the Orbis database is contingent on choice of estimation strategy. While the Olley-Pakes method may be a more methodologically rigorous way to calculate TFP than labor productivity, TFP index, or simple OLS-based estimation methods, the need for multiple years of data and more detailed financial information decreases the firm sample size, preventing the estimation of productivity for some country-sectors in our dataset entirely.

One of the advantages of using Orbis as a source of firm-level data is the ability to distinguish between domestic and foreign-owned firms operating in particular country markets. We use this distinction to test whether foreign-owned firms have significantly different average productivities and productivity distributions. Using two-sample tests of means we find that foreign firms tend to be more productive than domestic firms on average. We also find that foreign and domestic firms have significantly different productivity distributions.

Because we are calculating productivity from cross-sections rather than focusing on growth rates, the issue of true comparability between country-sectors looms large. Factors precluding balanced comparisons of productivity measures between country-sectors stem from both the nature of the data collected and the limitations of the estimation strategies used to overcome it. Our estimates here should be viewed with these caveats in mind. Our hope for future work in this area is to build on these findings by considering specific sectors and testing the empirical relationship between productivity and international trade.

References

Bartelsman, Eric J., and Mark Doms. “Understanding Productivity: Lessons from Longitudinal Microdata.” *Journal of Economic Literature*. 38, no. 3 (Sept. 2000): 569-594. http://www.jstor.org/stable/2565420

Bartelsman, Eric J., John Haltiwanger, and Stefano Scarpetta. “Measuring and Analyzing Cross-country Differences in Firm Dynamics.” In *Producer Dynamics: New Evidence from Micro Data,* 15-76. Edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. University of Chicago Press: 2009. http://www.nber.org/chapters/c0480

Bartelsman, Eric J., John Haltiwanger, and Stefano Scarpetta. “Cross-Country Differences in Productivity: The Role of Allocation and Selection.” The American Economic Review. 103, no. 1 (Feb. 2013): 305-334. http://www.jstor.org/stable/23469644

Bernard, Andrew and Charles I. Jones. “Comparing Apples to Oranges: Productivity Convergence and Measurement Across Industries and Countries.” *The American Economic Review.* 86, no. 5 (Dec. 1996): 1216-1238. http://www.jstor.org/stable/2118287

Biesebroeck, Johannes Van. “Robustness of Productivity Estimates.” *The Journal of Industrial Economics.* 55, no. 3 (Sept. 2007): 529-569. http://www.jstor.org/stable/4622397

Bureau van Dijk. “Untangling the world of private company information.” White Paper. (2017). https://www.bvdinfo.com/en-gb/contact-us/view-white-paper?White%20Paper%20Title={BF133221-3609-4A43-A375-C5F02E0DDFBE}

Cardarelli, Roberto and Lusine Lusinyan. “U.S. Total Factor Productivity Slowdown: Evidence for the U.S. States.” *IMF Working Paper Series.* WP/15/116. May 2015. https://www.imf.org/external/pubs/ft/wp/2015/wp15116.pdf

Caselli, Francesco and Daniel J. Wilson. “Importing technology.” *Journal of Monetary Economics*. 51, (2004): 1-32.

Del Gatto, Massimo, Adriana Di Liberto, and Carmelo Petraglia. “Measuring Productivity.” *Journal of Economic Surveys*. 25, no. 5 (2011): 952-1008.

Doms, Mark and Timothy Dunne. “Capital Adjustment Patterns in Manufacturing Plants.” *Review of Economic Dynamics*. 1, (1998): 409-429.

Gal, Peter N. “Measuring Total Factor Productivity at the Firm Level using OECD-Orbis.” *OECD Economics Department Working Papers*, No. 1049, OECD Publishing, Paris. (2013). http://dx.doi.org/10.1787/5k46dsb25ls6-en

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. “Captial Allocation and Productivity in South Europe.” *Quarterly Journal of Economics.* 132, no. 4 (June 2017): 1915-1967. https://doi.org/10.1093/qje/qjx024

Hall, Robert E., and Charles I. Jones. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *The Quarterly Journal of Economics*. 114, no. 1 (Feb. 1999): 83-116. http://www.jstor.org/stable/2586948

Helpman, Elhanan, Marc Melitz, and Stephen Yeaple. "Export Versus FDI with Hetrogenous Firms." *American Economic Review*. 94, no. 1 (Mar. 2004): 300-316. https://www.aeaweb.org/articles?id=10.1257/000282804322970814

Heshmati, Almas, and Masoomeh Rashidghlalam. "Estimation of Technical Change and TFP Growth based on Observable Technology Shifters." GLO Discussion Paper, no. 8. Global Labor Organization, Maastricht. (2017). http://hdl.handle.net/10419/152325

Jorgenson, Dale W. "The World KLEMS Initiative: Measuring Productivity at the Industry Level." In *The Oxford Handbook of Productivity Analysis*, 1: Ch. 20. Oxford: The Oxford University Press, Forthcoming.

Kalemli-Ozcan, Sebnem, Bent Sorenson, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas. "How to construct nationally Representative Firm Level Data from the Orbis Global Database." NBER Working Paper No. 21558. (Sept. 2015). http://www.nber.org/papers/w21558.pdf

OECD. *STAN Database for Structural Analysis (ISIC Rev. 4)*. https://stats.oecd.org/Index.aspx?DataSetCode=STANI4 (Accessed April 2018)

OECD. "Measurement of Aggregate and Industry-Level Productivity Growth." In *Measuring Productivity: OECD Manual*. (2001). http://unstats.un.org/unsd/EconStatKB/Attachment291.aspx?AttachmentType=1

Olley, G. Steven, and Ariel Pakes. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*. 64, no. 6 (Nov. 1996): 1263-1297. https://www.jstor.org/stable/2171831

Melitz, Marc J. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." Econometrica. 71, no. 6 (Nov. 2003): 1695-1725. https://www.jstor.org/stable/1555536

Levinsohn, James, and Amil Petrin. "Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies*. 70, no. 2 (Apr. 2003): 317-341. http://www.jstor.org/stable/3648636

Ribeiro, Samuel Pinto, Stefano Menghinello, and Koen De Backer. "The OECD Orbis Database: Responding to the Need for Firm-Level Micro-Data in the OECD." OECD Statistics Working Papers, 2010/01. OECD Publishing, Paris. (2010) http://dx.doi.org/10.1787/5kmhds8mzj8w-en

Syverson, Chad. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*. 112, no. 6 (Dec. 2004): 1181-1222. http://www.jstor.org/stable/10.1086/424743

Appendix A: Data Construction

A1: Creating a productivity dataset from Orbis.

The dataset used in this paper covers financial information from 2012-2016 for all companies in the Orbis database with non-missing operational revenue and employment data for 2013, 2014 and 2015. Data was downloaded from Orbis' online portal (https://Orbis4.bvdinfo.com) in September and October 2017. This portal provides 10 years of firm-level financial data, depending on the last available year for a firm's financial data. For example, a firm that has ten years of observations in Orbis where the last available year is 2016 has data from 2006 to 2016, while a firm that has ten years of observations in Orbis where the last available year is 2017, has data from 2007-2017. The Orbis online portal also updates company data when new data becomes available rather than annually. This makes it more difficult to construct a representative time series from the Orbis online portal than through annual versions of the Orbis database released on CDs, as was the case the compilation of datasets in Kalemli-Ozcan et al (2015) and Gopinath et al (2017). Table A.1 lists the number of firms available by country in our dataset.

Table A.1: Number of firm observations in our dataset pulled from the Orbis database in November 2017

Country	Obs.	Country	Obs.	Country	Obs.	Country	Obs.
Italy	1410750	United States	8896	Sri Lanka	194	Zambia	21
Russia	616588	China	7450	Liechtenstein	192	Zimbabwe	21
Romania	474281	Poland	6949	South Africa	132	Brazil	18
Spain	376121	Bosnia and Herzegovina	6919	Philippines	105	Ghana	18
Australia	211548	Ireland	6308	Malaysia	104	Kenya	18
Finland	207900	Netherlands	4296	Jordan	100	Chile	17
Bulgaria	201768	Greece	4284	Monaco	95	Colombia	15
Ukraine	201069	Denmark	3907	Taiwan	91	Thailand	15
Portugal	194362	Hong Kong	3408	Bangladesh	88	Angola	12
Germany	176969	Kazakhstan	2231	Nigeria	88	Albania	12
Hungary	168401	Iceland	2126	Malta	78	Kuwait	12
Sweden	138237	Belarus	1877	Mexico	74	Marshall Islands	12
Japan	127222	Montenegro	1361	Oman	49	Peru	12
Czech Republic	87660	Vietnam	746	Cuba	44	Uganda	11
Latvia	81332	Luxembourg	733	Moldova	39	Nepal	10
United Kingdom	61778	Cayman Islands	610	Canada	35	Lebanon	9
Slovakia	56043	Israel	479	Iraq	35	United Arab Emirates	8
Serbia	51525	Switzerland	446	Cambodia	35	Ethiopia	8
Croatia	51180	Indonesia	370	Egypt	34	Gibraltar	8
Korea	49991	Turkey	365	British Virgin Islands	30	Panama	8
Macedonia, FYR	39760	Bermuda	334	Palestine	26	Azerbaijan	7
Slovenia	36768	Cyprus	281	Saudi Arabia	26	Botswana	7
Lithuania	35384	New Zealand	262	Tanzania	24	Kosovo	7
France	26055	Iran	258	Bahrain	23	Curacao	6
Estonia	22513	India	235	Costa Rica	22	Mongolia	6
Belgium	16954	Norway	232	Dominican Republic	21		
Austria	12214	Pakistan	224	Singapore	21		

Note: the dataset contains 14 additional observations with the ISO codes of II, which indicates an international organization such as the World Bank.
Source: Authors' calculations using data from Bureau van Dijk's Orbis database

The BvDID number is used as a unique identifier for firms, and firms are classified into sectors by two-digit NACE codes.[26] The subsample of this dataset used for our analysis only includes country-sector pairs with at least 30 firm-level observations. Additionally, we limit the NACE codes covered in our analysis to those falling under NACE two-digit codes 10-33 (manufacturing) and 41-93 (services), excluding codes 64-66, which include banking and insurance activities, and 84, which covers public administration and defense. Banking and insurance are excluded not because of poor coverage, but because return on assets (rather than output per worker) tends to reflect productivity of these sectors.[27]

These restrictions produce a sample of 49 countries (listed below) that can be used for our calculations of county-sector productivity.

Table A.2: Country observations for productivity analysis

Country	Number of sectors	Country	Number of sectors
Italy	543	Lithuania	176
Russia	462	Macedonia, FYR	165
Spain	427	Estonia	133
Germany	425	Belgium	127
Romania	362	Austria	92
Ukraine	361	China	72
Hungary	346	Poland	44
Portugal	324	Bosnia and Herzegovina	38
Bulgaria	301	Greece	34
Finland	289	Ireland	26
Japan	260	Hong Kong	22
United Kingdom	256	Netherlands	20
Latvia	254	Denmark	16
Czech Republic	249	Island	16
Sweden	242	United States	12
Korea	216	Belarus	7
Australia	214	Kazakhstan	7
Croatia	210	Montenegro	6
Slovakia	204	Israel	1
Serbia	200	Luxembourg	1
Slovenia	182	Vietnam	1
France	178		

Source: Authors' calculations using data from Bureau van Dijk's Orbis database

[26] The database also categorizes companies by NAICS code.

[27] Gal (2013) also excludes financial services from the analysis of firm level productivity using Orbis.

A2: Foreign Ownership

We use Orbis' Global Ultimate Owner (GUO) variables to determine foreign ownership. A GUO owns at least 51 percent of a company, either directly or through at least 51 percent ownership of a subsidiary that owns the company. In addition to identifying the GUO, the dataset includes information on the GUO country of origin, which allows us to classify subsidiaries as either domestic or foreign-owned. Firms for which the firm country and the GUO country match are considered domestic firms, while firms for which the firm country and the GUO country do not match are considered foreign firms. One limitation of the Orbis database's prioritization of up-to-date information over historical information is that the GUO variable only reflects the latest ownership information, so we do not know whether firms have changed ownership during our sample timeframe. Additionally, we are unable to distinguish between foreign acquisitions of companies and greenfield investment.

This methodology can be misleading in cases where large multinational companies have GUOs that are holding companies in a separate country for tax purposes. For example, because Baidu's GUO is a holding company in the Cayman Islands, Baidu's main operating arm would be considered foreign in China. To correct for this problem, we considered firms to be domestic if the GUO was a holding company (classified under primary NACE code 6420) located in either the Cayman Islands or Bermuda.

There are additional cases of firms where assigning a classification of foreign or domestic is less straightforward. Some firms in the sample have BvDID numbers that match the GUO ID number, indicating that these firm observations are Global Ultimate Owners. Since many of these GUOs have consolidated accounts that include their global operations, it is difficult to classify them as foreign or domestic firms, since their financial variables may reflect conditions in a market other than their headquarters market. Additionally, there are company observations with no information on the GUO. This could indicate that there are no shareholders and therefore that the firm is a domestic entity. However, this could also indicate that although the firm is majority foreign-owned, it is owned by multiple foreign entities, none of which have a total share above 51 percent. As a result, both GUOs and firms with no ownership are included in our calculations of country and sector level productivity, but excluded when comparing the productivity of foreign and domestic firms.

Finally, for about 270,000 firms, the GUO country variable is marked n.a. This indicates that the GUO is an individual, trust, or investment firm (such as a private equity firm). In these cases, the following rules were used to assign country codes to the GUO:

- When there was only one company associated with the GUO, the company was considered a domestic firm (212,032 observations).
- When there were multiple companies associated with the GUO but all were located in the same country, all companies associated with that GUO were considered domestic firms (57,451 observations).
- When there were multiple companies in different countries associated with the GUO, we conducted internet searches to assign country codes to GUOs based on the headquarter location of the individual's primary company, based on sources such as company websites, Bloomberg's Executive profiles, Forbes Billionaires lists, and news articles. For example, while

the Walton family is listed as the GUO of Walmart's subsidiaries, since we can connect the Walton family to Walmart, we assigned the Walton family the United States as their country code. This technique was applied for 911 GUOs in our sample, and we successfully assigned country codes to 42 percent of these firms. The remaining unassigned firms were not included in our analysis.

A3: Additional Variables

Purchasing Power Parity (PPP) country-level deflators come from the World Bank World Development Indicators. These estimates are based on the 2011 International Comparison Program benchmark estimates in most cases, but are supplemented by annual conversion factors for 47 countries through Eurostat and OECD data.

The share of labor[28] as an input into the gross output of a sector is obtained for each country from the World KLEMS or OECD STAN.[29] Industry labor shares are defined as specifically as possible: at the two-digit ISIC[30] level at its finest level of detail, or at the ISIC section or range of sections where specification at the two-digit was not provided. Out of 3,471 country-sectors, 197 are missing labor share data. Data for the estimation of labor share is from 2012 or from the next most recent complete year of data available.

[28] Calculated at the sector level as the compensation of engaged persons divided by the sum of the compensation of employed persons and capital compensation. "Engaged persons" are defined as employed and self-employed individuals; their compensation is calculated under the assumption that both types of workers receive the same wage. Capital compensation is the value-added of a sector less labor compensation.

[29] World KLEMS was the first choice for this data, as data prepared by national statistical agencies under this methodology follows SNA 2008 and assures a higher degree of international comparability (Jorgenson 2016). If industry labor shares were not available for a country in World KLEMS, data was sought from the OECD STAN database which is primarily based on member countries' SNA 2008 national accounts and is supplemented with data from other sources (national business surveys/censuses etc.) to estimate missing values (OECD 2018).

[30] ISIC divisions correspond with NACE codes at the 2-digit level.

Table A.3: Data sources for labor share by country and sector

Country	Year	Number of ISIC divisions (2-digit NACE) available (89 total)	Data Source (release partner)
Australia	2012	86	World KLEMS (Australian Bureau of Statistics)
Austria	2012	86	World KLEMS (EU KLEMS)
Belgium	2012	88	OECD STAN
Bulgaria	2012	78	World KLEMS (EU KLEMS)
Canada	2008	88	World KLEMS (Statistics Canada)
China	2010	79	World KLEMS (RIETI)
Cyprus	2012	76	World KLEMS (EU KLEMS)
Czech	2012	89	World KLEMS (EU KLEMS)
Germany	2012	81	World KLEMS (EU KLEMS)
Denmark	2012	89	World KLEMS (EU KLEMS)
Spain	2012	89	World KLEMS (EU KLEMS)
Estonia	2012	89	World KLEMS (EU KLEMS)
Finland	2012	89	World KLEMS (EU KLEMS)
France	2012	81	World KLEMS (EU KLEMS)
United Kingdom	2012	86	World KLEMS (EU KLEMS)
Greece	2012	89	World KLEMS (EU KLEMS)
Croatia	2012	83	World KLEMS (EU KLEMS)
Hungary	2012	89	World KLEMS (EU KLEMS)
India	2012	88	World KLEMS (Reserve Bank of India)
Ireland	2011	88	OECD STAN
Iceland	2012	78	OECD STAN
Israel	2012	62	OECD STAN
Italy	2012	89	World KLEMS (EU KLEMS)
Japan	2009	86	World KLEMS (RIETI)
Korea (the Republic of)	2012	88	World KLEMS (Korea Productivity Center)
Lithuania	2011	88	OECD STAN
Luxembourg	2012	84	World KLEMS (EU KLEMS)
Latvia	2012	51	World KLEMS (EU KLEMS)
Netherlands (the)	2012	83	World KLEMS (EU KLEMS)
Norway	2012	88	OECD STAN
Poland	2012	79	World KLEMS (EU KLEMS)
Portugal	2012	86	World KLEMS (EU KLEMS)
Romania	2012	86	World KLEMS (EU KLEMS)
Russian Federation (the)	2012	86	World KLEMS (GGDC and HSE)
Slovakia	2012	89	World KLEMS (EU KLEMS)
Slovenia	2012	83	World KLEMS (EU KLEMS)
Sweden	2012	84	World KLEMS (EU KLEMS)
Taiwan	2010	84	World KLEMS (Asia KLEMS)
United States	2012	89	World KLEMS (EU KLEMS)

A4: Descriptives

Table A.4: Descriptives for all firms in NACE sectors 25 and 62 in 2014

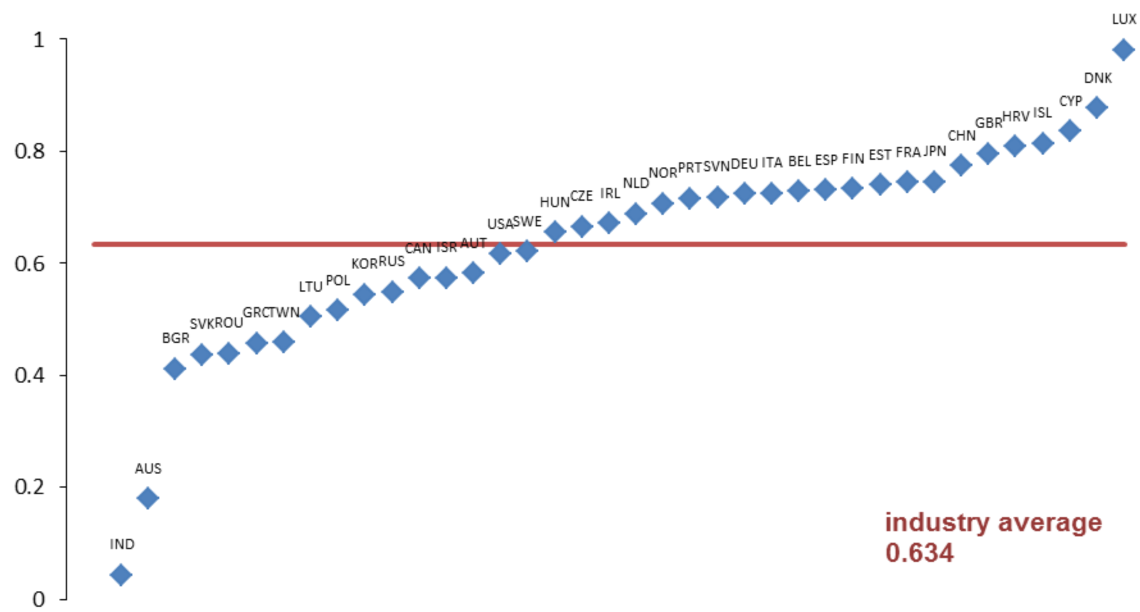
		Number of employees				Firm revenue (PPP-adjusted)				Tangible Fixed Assets (PPP-adjusted)			
	Number of firms	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**
Manufacturing Sector: NACE 25 - Manufacture of fabricated metal products except machinery and equipment													
Foreign	1,791	129.8	47	19,850	1	$41,237	$9,259	$3,749,334	$2	$8,364	$1,585	$2,546,534	$1
Domestic	27,222	31.5	10	14,187	1	$8,582	$1,224	$8,901,959	$2	$2,026	$193	$943,201	$1
Services Sector: NACE 62 - Computer programming, consultancy, and related activities													
Foreign	2,982	166.6	33	15,516	1	$63,109	$6,829	$22,304,144	$2	$3,723	$116	$1,227,034	$1
Domestic	16,884	35.8	4	43,726	1	$8,310	$441	$7,753,017	$1	$1,261	$25	$3,452,168	$1

Note: Sample includes all firms with adequate data to perform the TFP Index Method calculation.
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

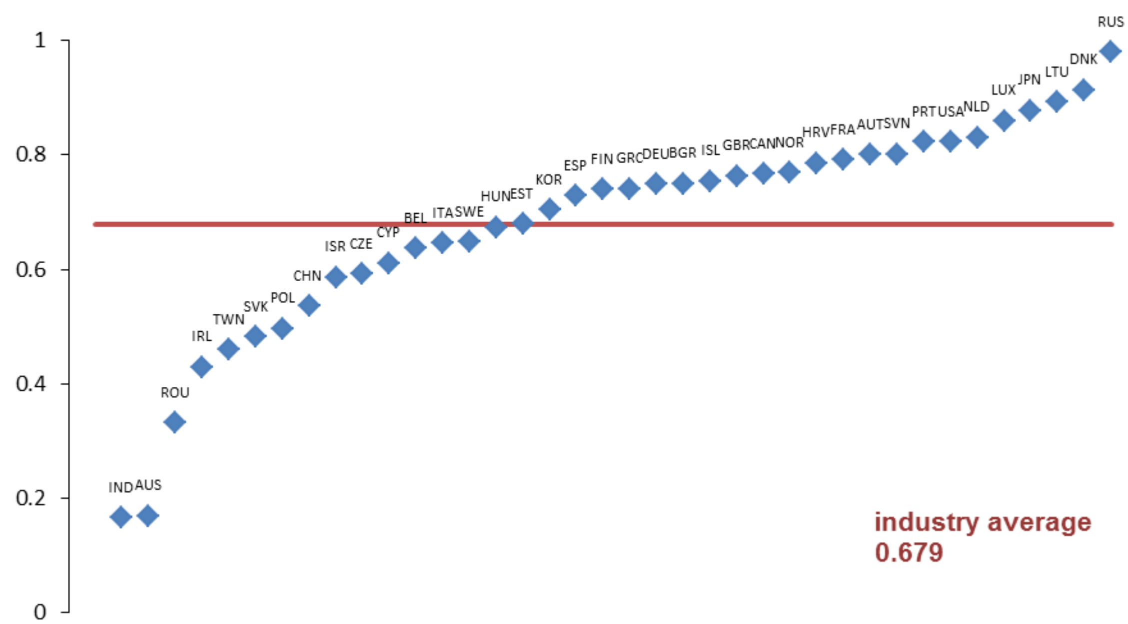
Appendix B: Supplemental tables and figures

Figure B.1: Labor shares for Select Industries

Manufacture of fabricated metal products, except machinery and equipment (NACE 25)



Computer Programming, consultancy and related activities (NACE 62)



Source: Authors' calculations from KLEMS and OECD STAN databases.

Table B.1: Country-Subsector Coverage in Each Estimation Method

	Count of NACE four-digit Subsectors in Each Estimation Method				
Country Code	The Labor Productivity Sample (2014)	The Index Method Sample (2014)	The OLS Sample (at least 3 years)	The Olley-Pakes Sample (at least 3 years)	The Levinsohn-Petrin Sample (at least 3 years)
AUS	214	3	3	3	
AUT	92	75	76	22	3
BEL	127	124	126	125	104
BGR	301	166	292	230	265
BIH	38		38	32	36
BLR	7				
CHN	72	68	68	20	
CYM			1	1	
CZE	249	130	161	117	156
DEU	425	235	333	126	109
DNK	16	14	18	14	
ESP	427	422	425	415	419
EST	133	98	110	93	95
FIN	289	177	186	173	163
FRA	178	106	178	169	141
GBR	256	244	247	245	
GRC	34	33	34	22	
HKG	22				
HRV	210	194	210	191	210
HUN	346	328	334	330	85
IRL	26	22	22	20	
ISL	16	10	11	10	
ISR	1	1	1	1	
ITA	543	510	512	510	508
JPN	260	258	261	46	1
KAZ	7		7		
KOR	216	214	214	201	197
LTU	176	42	34		
LUX	1	1	1	1	
LVA	254	77	265	8	8
MKD	165		165	129	150
MNE	6		7	3	4
NLD	20	16	18	1	
POL	44	28	29	23	23
PRT	324	317	318	309	279
ROU	362	339	388	352	368

RUS	462	385	462		
SRB	200		219	185	197
SVK	204	189	204	188	203
SVN	182	170	177	175	175
SWE	242	7	10	8	4
UKR	361		360	91	87
USA	12		12	12	
VNM	1		1		
Total	7521	5003	6538	4601	3990

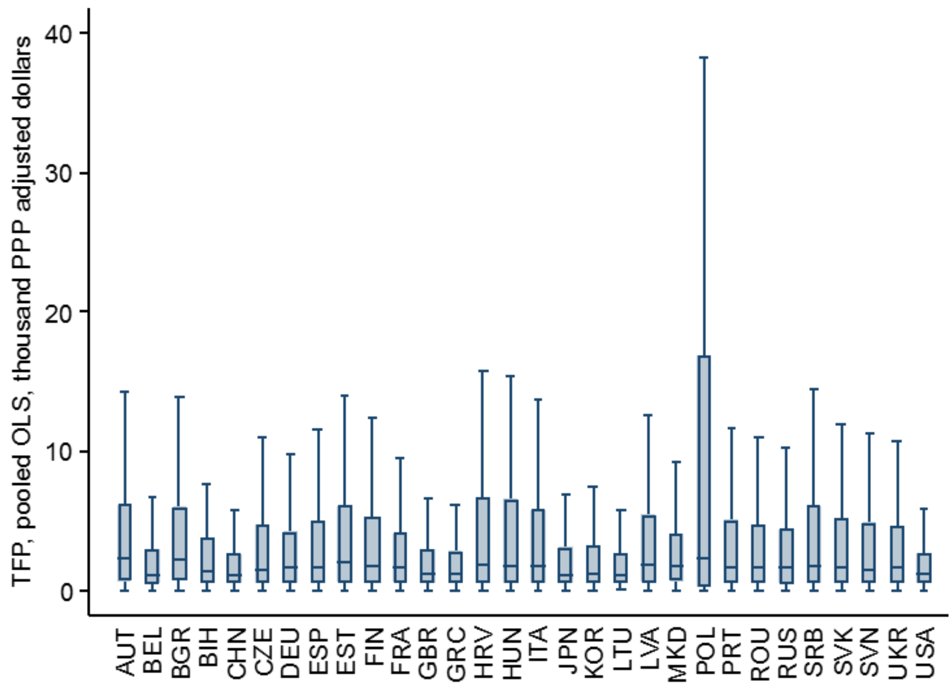
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Table B.2: Four-digit NACE subsectors contained in two-digit sector codes 25 and 62

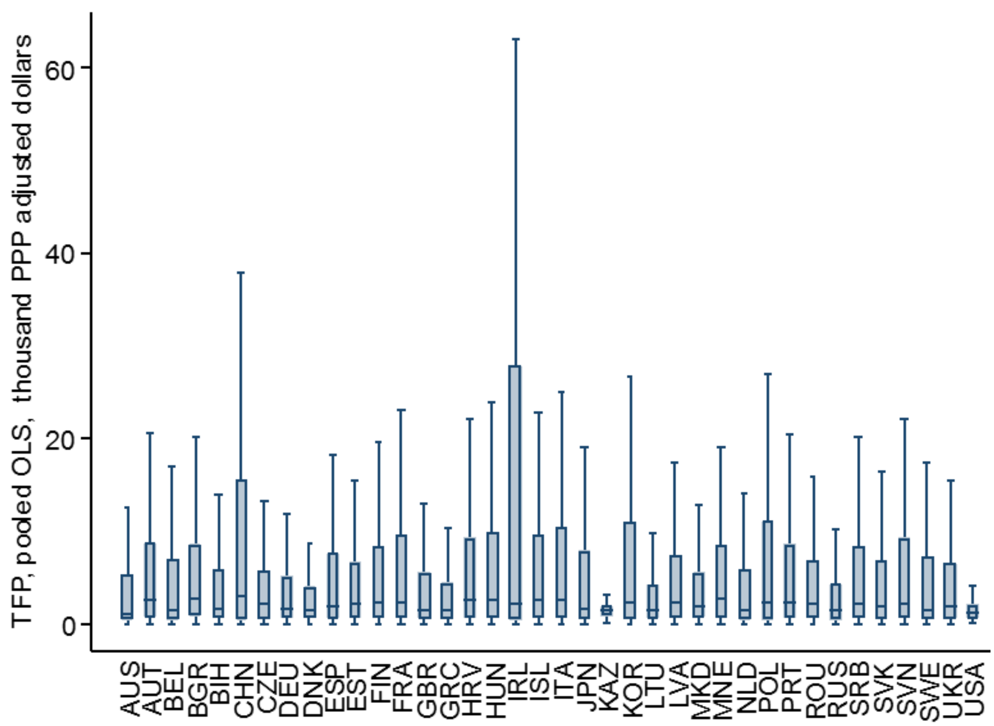
25 Manufacture of fabricated metal products, except machinery and equipment	
25.11	Manufacture of metal structures and parts of structures
25.12	Manufacture of doors and windows of metal
25.21	Manufacture of central heating radiators and boilers
25.29	Manufacture of other tanks, reservoirs and containers of metal
25.30	Manufacture of steam generators, except central heating hot water boilers
25.40	Manufacture of weapons and ammunition
25.50	Forging, pressing, stamping and roll-forming of metal; powder metallurgy
25.61	Treatment and coating of metals
25.62	Machining
25.71	Manufacture of cutlery
25.72	Manufacture of locks and hinges
25.73	Manufacture of tools
25.91	Manufacture of steel drums and similar containers
25.92	Manufacture of light metal packaging
25.93	Manufacture of wire products, chain and springs
25.94	Manufacture of fasteners and screw machine products
25.99	Manufacture of other fabricated metal products n.e.c.
62 Computer programming, consultancy and related activities	
62.01	Computer programming activities
62.02	Computer consultancy activities

Figure B.2: Dispersion of pooled OLS productivity estimates in Manufacturing and Services by county, 2012-2016.

Manufacturing (NACE 10-33)



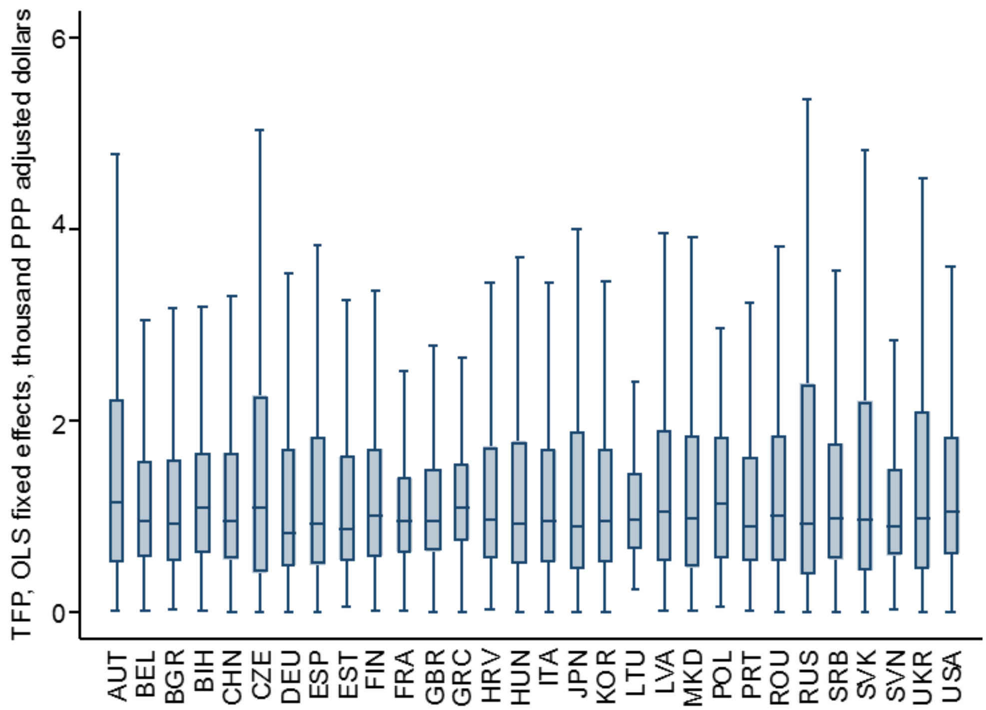
Services (NACE 41-64, 66-83, 85-99)



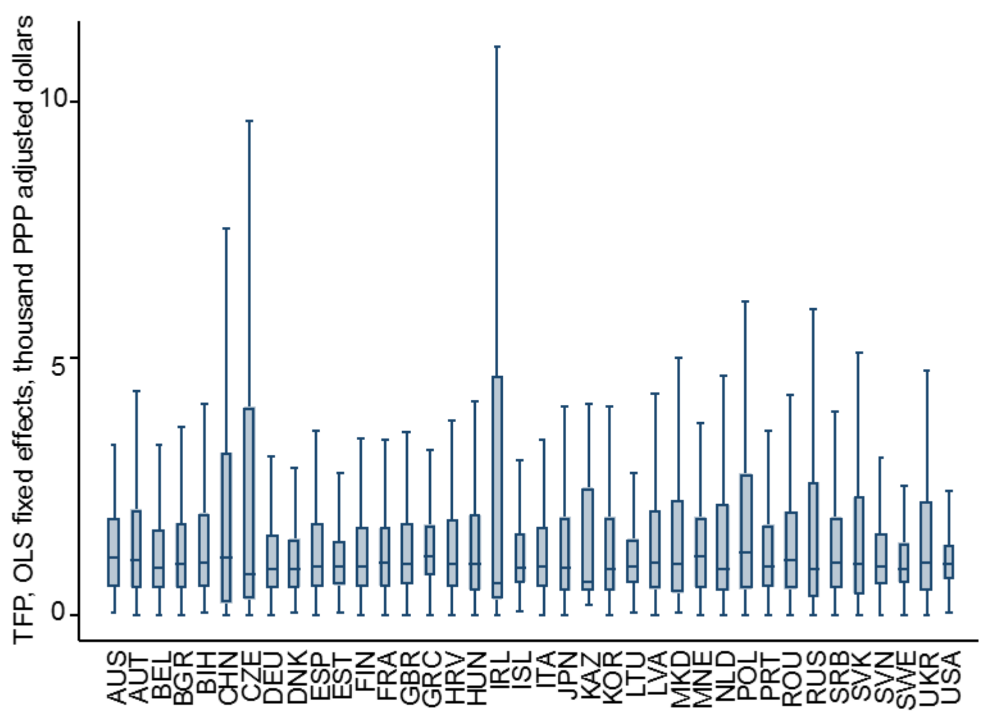
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure B.3: Dispersion of OLS fixed effects productivity estimates in Manufacturing and Services by county, 2012-2016.

Manufacturing (NACE 10-33)



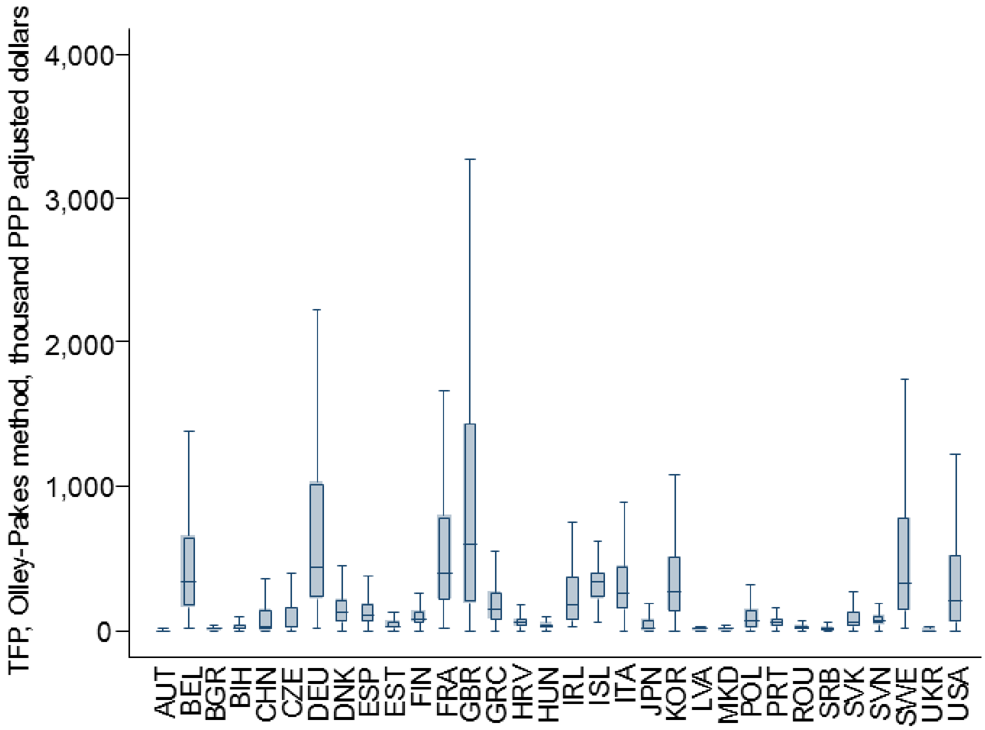
Services (NACE 41-64, 66-83, 85-99)



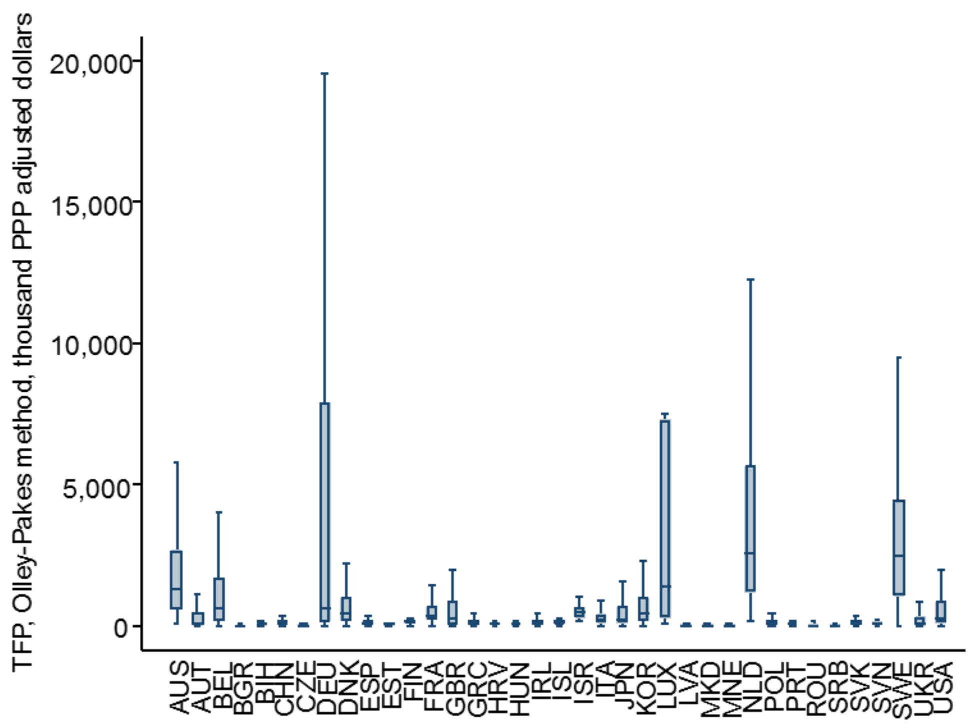
Source: Authors' estimates using data from Bureau van Dijk's Orbis database

Figure B.4: Dispersion of Olley-Pakes productivity estimates in Manufacturing and Services by country, 2012-2016.

Manufacturing (NACE 10-33)



Services (NACE 41-64, 66-83, 85-99)



Source: Authors' estimates using data from Bureau van Dijk's Orbis database