OECD Science, Technology and Industry Working Papers
2019/02

Can firm micro data match
macro trends? Comparing
MultiProd and STAN

**Matej Bajgar,
Giuseppe Berlingieri,
Sara Calligaris,
Chiara Criscuolo**

<https://dx.doi.org/10.1787/2b0ac915-en>

OECD Science, Technology and Industry Working Papers

The release of this working paper has been authorised by Andrew Wyckoff, OECD Director for Science, Technology and Innovation and by Laurence Boon, OECD Economics Directorate. OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to Directorate for Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

A previous version of this document was presented and discussed by the OECD Committee for Industry, Innovation and Entrepreneurship (CIIE) and its Working Party on Industry Analysis (WPIA).

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CIIE/WPIA(2017)6/FINAL

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 811181. The European Commission is not responsible for any use that may be made of the information produced by the project.

© OECD 2019 You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

Can Firm Micro Data Match Macro Trends? Comparing Multiprod and STAN

Matej Bajgar, Giuseppe Berlingieri, Sara Calligaris and Chiara Criscuolo

Abstract

Better understanding about the drivers of aggregate productivity and wage inequality requires data that offer a representative picture of the underlying firm-level heterogeneity but are, at the same time, able to reproduce patterns observed in aggregate data. The OECD MultiProd project aims to generate such data by collaborating with a network of national experts who apply a harmonised statistical code to representative business microdata across a large number of countries. This paper compares the project's output to the OECD STAN database to test to what extent MultiProd data can be taken as reflecting the aggregate economies in question, and if they are able to reproduce patterns observed in aggregate data across years, industries and countries. The results suggest that (1) MultiProd captures a major part of gross output, value added and employment in most of the countries covered; and (2) MultiProd reproduces aggregate patterns relatively well, with median correlations over time, across industries and across countries between 0.75.

Table of Contents

OECD Science, Technology and Industry Working Papers.....	2
Can Firm Micro Data Match Macro Trends? Comparing Multiprod and STAN.....	3
1. Introduction	5
2. Data sources: Why they may give different answers.....	7
3. Coverage: How much do the microdata capture?	10
4. Correlations: Can the microdata capture aggregate patterns?	14
5. Conclusion.....	17
Endnotes.....	18

Figures

Figure 3.1. Total output, value added and employment relative to STAN, by SNA A38 industry	11
Figure 3.2. Total value added relative to STAN, by SNA A38 industry	11
Figure 3.3. Total output, value added and employment relative to STAN, by country.....	12
Figure 3.4. Total output, value added and employment rel. to STAN, by country over time	13
Figure 4.1. Correlations between MultiProd and STAN, by variable and type of variation	15
Figure 4.2. Correlations in labour productivity across years, by SNA A38	15
Figure 4.3. Correlations in labour productivity over time, by country.....	16
Figure 4.4. Correlations in labour productivity across industries, by country.....	16

1. Introduction

The role of firm-level dynamics is essential in explaining aggregate performance. In the last decade, developed economies in Europe and elsewhere have faced an increasing heterogeneity in firm performance, even within narrowly defined sectors (Berlingieri et al., 2017a). In the light of this heterogeneity, analysing industry averages does not offer a complete picture: different economies and industries might display the same averages but very different underlying distributions, with different policy implications. For instance, low average productivity can be explained by too few firms at the top (lack of innovation) or by too many firms at the bottom (weak market selection) – two different situations that call for very different policy responses.

The Directorate for Science, Technology and Innovation (STI) at the OECD aims to contribute to this debate by providing new policy relevant evidence through the MultiProd project. This project provides a comprehensive picture of productivity patterns across many countries over the last two decades.¹ It is part of a larger effort within STI that draws on existing official national sources of confidential firm-level data to provide non-confidential harmonised micro-aggregated statistics and analyses that are comparable across countries. In particular, the administrative sources and official surveys include business registers, tax data and production surveys. MultiProd relies on a distributed microdata approach, and it produces results at the level of cells defined, for example, by country, industry and firm size categories. The results are then collected, checked and analysed at the OECD.

In order to correctly analyse and interpret results from the MultiProd project, it is important to understand whether they are representative for the entire population of firms. Do the underlying microdata capture most of the employment and output observed in aggregate data sources? And can they replicate patterns observed in aggregate data over time, as well as across countries and industries? In principle, there are ample reasons for why this might not be the case, starting from the fact that macroeconomic statistics and MultiProd, although both based on micro-data, have a different primary focus: the former aims to measure economic aggregates, the latter to capture firm heterogeneity. The different aims imply different choices in the construction of the data, for example, with respect to imputation, data cleaning, weighting and assignment of firms to industries or variable definitions.

The purpose of this report is to investigate to what extent the microdata-based analysis in MultiProd reflects the aggregate economies in question, and how well it fits the patterns observed in aggregate data. For this purpose, it compares statistics calculated with MultiProd data against the benchmark of the OECD STructural ANalysis (STAN) Database, based primarily on the 2008 System of National Accounts (SNA08) statistics and an ISIC Rev. 4 industry breakdown. The analysis focuses on manufacturing and non-financial market services, which together account for most of the economy and are the sectors studied by MultiProd research to date.

It is important to highlight that the results of this exercise should be considered as indicative and work in progress. As explained in more detail below, construction of aggregate (SNA08) statistics involves many adjustments, which imply that even administrative data covering the whole population of firms will not exactly reproduce the aggregate statistics. An observed difference between the two sources does not imply that one of the sources is “wrong”, rather it reflects the fact that they are designed for different purposes.

Furthermore, given the on-going nature of the MultiProd project, the data are constantly improved and updated by conducting automatic and manual data checks, extending to new and possibly more representative data sources, as well as adding new countries. This report analyses the version 1.1 of the MultiProd database and is aimed at uncovering possible areas of improvement to inform the work underway for the version 2.0 of the database.

Keeping these caveats in mind, our comparisons suggest that the MultiProd results are, in general, consistent with aggregate patterns. In most countries, the microdata used by MultiProd have a good coverage, representing typically 80-100% of total gross output, value added and employment in manufacturing and non-financial market services, with a fairly stable coverage over time. Statistics calculated in the MultiProd project are also generally successful in replicating aggregate patterns observed in the STAN data at the 2-digit industry (SNA A38) level. The median correlations between MultiProd and STAN within country-industry over time, for a given country and year across industries, and for a given industry and year across countries are very high. They are above 0.75 (and in several cases close to 1) for all variables examined (gross output, value added, employment, labour productivity and wages), for all types of variation and for most countries.

Distributed analysis of representative official microdata and analysis of commercial firm-level data sets each have advantages and should be seen as complementary. Besides distributed microdata analysis, commercial firm-level datasets represent an alternative data source for cross-country analysis of firm-level performance. A companion report examines how well one such dataset – Orbis – can account for economic aggregates, as well as various moments of firm distribution (see Bajgar et al., forthcoming). The comparison of the results obtained in the two reports reveals that overall the MultiProd dataset offers greater coverage and superior ability to fit aggregate patterns for a vast majority of countries and industries. Nevertheless, the coverage of Orbis is reasonably good for some countries, and Orbis has an important advantage in its flexibility: unlike MultiProd, it allows cross-country regression analysis at the firm level, and constructing new indicators in Orbis is easier than re-running a new version of a distributed code in 20+ countries.

Different types of analysis do not only call for different data, but they also assign different emphasis to the ability of the microdata to replicate macroeconomic trends. For example, a decomposition of aggregate productivity is unlikely to give meaningful results if the aggregate productivity being decomposed does not closely track the productivity observed in aggregate data. In contrast, poor ability to replicate macro trends does not preclude a meaningful within-firm regression analysis, although selection issues in commercial databases – due to the fact that the data might not cover a random sample of small firms but only the better performing ones – call for caution in both the analysis (e.g. selecting firms only above a certain size threshold) and interpretation (e.g., if a dataset does not cover small firms, estimated effects of a policy change should be interpreted as effects on medium and large firms rather than effects on firms in general).

The rest of the paper is structured as follows. Section 2. introduces the MultiProd and STAN database and outlines why these sources may lead to different results. Section 3. evaluates the coverage of MultiProd data against the benchmark of STAN. Section 4. investigates to what extent MultiProd can reproduce industry-level trends observed in STAN. Section 5. concludes.

2. Data sources: Why they may give different answers

This section introduces the two main sources used in the report: the OECD MultiProd and the OECD STAN databases. It compares them, and subsequently outlines why aggregate figures and industry-level trends constructed with these datasets may differ from each other.

2.1. OECD MultiProd

MultiProd is an OECD project, which relies on the so-called *distributed microdata approach*. It applies a harmonised statistical routine to confidential microdata in more than 20 countries to produce a database of statistics calculated at the country, year, and ISIC Rev.4 SNA “A38” industry level.² Other firm characteristics such as size, age group, demographics (e.g. entrants, exiting firms) and productivity quantiles are also considered. MultiProd relies primarily on administrative data or production surveys, which provide the firm-level variables needed for analysing productivity and wages (e.g. output, value added, employment, capital, investment and wage and salaries). As the data might not cover the universe of firms in all countries, they are complemented, where necessary, with business registers, which contain a more limited set of variables but cover the entire firm population. The population structure captured by the business register is used to re-weight the production surveys to obtain representative statistics. The data, methodology and output of the MultiProd project are described in detail in Berlingieri et al. (2017b), whereas additional details about the version of the MultiProd dataset used in this report can be found in Desnoyers-James et al. (2019).

The version of MultiProd output used here covers 20 countries (Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Sweden and Switzerland). The analysis in this paper largely focuses on statistics calculated for the ISIC Rev.4 SNA A38 industries. For confidentiality reasons, the output for Luxembourg is currently available only at a more aggregate 7-sector level, so Luxembourg is not included in the present analysis. Additionally, the analysis excludes New Zealand and Chile, which are not covered (New Zealand) or are only partially covered (Chile) by the version of STAN used here. Therefore, the present paper covers the following 17 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Sweden and Switzerland. The sample time period varies across countries and spans from 1994 to 2014, but most often from early 2000s to 2014. MultiProd collects data for all sectors of the economy, whenever available. However, for the purposes of this analysis, the sample is restricted to manufacturing and non-financial market services, excluding “Coke and refined petroleum”, “Real estate” and “Scientific R&D”.³ The data are generally representative of the entire population of firms with at least one employment unit (i.e. employee in most cases). However, there are some exceptions, such as: i) the data for Austria and the Netherlands cover only firms with at least 10 employees, and the data for Germany cover only firms with at least 20 employees; ii) the data for Finland exclude firms with less than one full-time equivalent in terms of persons engaged; iii) the data for Japan include only manufacturing. For a detailed description of the thresholds see Desnoyers-James et al. (2019).

Statistics calculated by the MultiProd code cover a broad range of topics, including distribution of firm size and firm growth; distribution of labour and multi-factor productivity levels and growth; aggregate productivity; aggregate productivity decompositions developed by Olley and Pakes (1996), Petrin and Levinsohn (2012) and Melitz and Polanec (2015); and characteristics of firms at the productivity frontier. Initial research based on MultiProd output has focused on the growing dispersion of productivity and wages across firms, as well as on the relationship between the two (Berlingieri et al., 2017a; OECD, 2018), and on how the link among firm productivity, size, and wages differs across sectors (Berlingieri et al., 2018a,b).

The variables used in this paper include (industry-level) total gross output, value added, employment, but also aggregate (weighted) average labour productivity and wages, which can be obtained by dividing, respectively, the total value added and the total wage bill by the total employment. The employment variable covers the number of persons engaged or the salaried employees.⁴ For the majority of countries, the underlying data provide information on employees or both (only employees for Australia, Belgium, Canada, Denmark, France, Hungary, Japan, Norway, Sweden and Switzerland; both for Austria, Germany, Ireland, Italy and New Zealand), the only exception being the Netherlands, Portugal and Finland. All variables are calculated separately for each country, SNA A38 industry and year. For the comparisons with STAN, only observations appearing in both MultiProd and STAN are kept. This is also the case when the values are aggregated to the economy-wide level, to ensure that any differences between datasets are not due to sectoral coverage.

2.2. OECD STAN

The OECD STAN database describes industrial performance for a large set of OECD countries.⁵ The target industry list covers ISIC Rev.4 2-digit Divisions (A88 list) and standard aggregates with some additional detail for certain activities and some customised aggregates for analytical purposes. Thus, in general, information is available at the level of SNA A38 industries (for a few countries and years, more aggregate statistics may only be available). It is primarily based on countries' latest published National Accounts (SNA08) by industry statistics complemented by other sources such as business surveys or censuses to make estimates for missing detail and earlier vintages of National Accounts to extend series back in time.

STAN appears to be the most suitable aggregate point of comparison for MultiProd. Its important advantage in this context over other industry-level datasets (EU KLEMS; the World Input Output Database, Timmer et al., 2015) is that the MultiProd code is based on industry definitions and industry-specific price deflators from STAN. The differences between MultiProd and STAN are, therefore, not due to differences in the industry classifications or deflators used. This also means that all countries in MultiProd (except New Zealand and Chile) are well covered by STAN.

The STAN variables used in this paper are the ones listed above for MultiProd. To match MultiProd as closely as possible, the default variable to measure employment is the number of employees, and the number of persons engaged is used for countries for which the former measure is not available in STAN (Australia, Japan, Switzerland), or for which MultiProd also relies on persons engaged (Finland, Netherlands, Portugal).

2.3. Why different data sources may lead to different results

There are plenty of reasons why the results based on MultiProd and STAN might differ.

To begin with, there are conceptual differences between the two data sources. For instance, for the “Wholesale and retail” sector, the measure currently used in MultiProd for output is total sales, while gross output in National Accounts excludes “*Purchases of goods and services purchased for resale*”, only taking into account the trade margins. Similarly, STAN measures value added in basic prices, but business surveys often measure value added in factor costs (i.e. net of taxes and subsidies on production other than taxes on products).⁶

Many differences are also due to the different focus of the datasets, with STAN aimed at measuring economic aggregates and MultiProd at capturing firm heterogeneity. For example, official aggregate statistics, on which STAN is based, often rely on imputation for missing observations, whereas MultiProd tends to drop imputed values to avoid artificially low heterogeneity across firms. Both types of data sometimes rely on weighting, but they apply different weights. Importantly, such methodological choices play a greater role in countries where the underlying firm data represent a sample rather than a census and, as such, cover a smaller share of firms. The results will tend to differ more between the two data sources in these countries. Adjustments in national accounts intended to balance the Gross Domestic Product obtained through the expenditure-based, output-based and income-based approaches can also drive differences between MultiProd and STAN.

Additionally, individual firms may be assigned to different industries in each case: while official aggregate statistics are likely to rely on the industry a firm reports in a given year, MultiProd keeps the industry constant over time (equal to the most recent mode industry) for each firm to facilitate industry-specific analysis tracking individual firms over time. The change in industry classifications from ISIC rev. 3.1 and ISIC rev. 4 is also treated differently in each data source. In addition, industry misclassification may lead to some ‘orphan’ firm appearing in sectors that are not actually covered by the underlying data sources; hence the coverage of the average sector might be artificially low to the extent that these outlier sectors are kept in the analysis (since they are not empty). Differences in variable definitions and data cleaning may also generate differences between the data; for instance, MultiProd typically keeps only firms with at least one unit of employment.

As outlined in this section, there are strong reasons for why aggregate patterns produced by each of the datasets *could* differ. To what extent this is indeed the case is an empirical question that the remainder of this paper tries to answer.

3. Coverage: How much do the microdata capture?

A first step towards understanding how well the firm-level data approximate aggregate statistics involves investigating if both data sources add up to similar total gross output, value added and employment. To this end, this section describes the overall coverage of microdata used in the MultiProd project and its variation over industries, variables, countries and time.

The average coverage for each SNA A38 industry is illustrated in Figure 3.1. Each bar represents the ratio between, respectively, the total gross output, value added or employment observed in MultiProd and STAN, on average across countries and years. The graph is based only on those country-industries that are present in both data sources for manufacturing and non-financial market services. The average coverage is fairly high, between 80% and 100% in most industries for all variables.⁷

Importantly, in order to reflect more closely the data that is used in the reports when relying on MultiProd data, the analysis presented here is based on weighted microdata. Thus, it shows the aggregate gross output, value added and employment across the firms in the sample once appropriate weights are applied. For a version of Figure 3.1 that is based on unweighted data, see Figure in the Annex.

The average coverage figures hide a significant variation over countries and years. For value added, Figure 3.2 displays the same values as the previous graph but underlying the dispersion in each SNA A38. For most industries, the coverage is between 70% and 100% in the majority of countries and years, and fewer than 10% of the 4100 country-industry-year observations in the sample show coverage of less than 50%. Only very few observations (<1%) have coverage below 25%, and most of these are due to just three country-industries: “Pharmaceuticals” and “Transport equipment” in Norway and “Telecommunications” in France. “Hotels and restaurants” tend to have a somewhat lower coverage than other industries. This is because a relatively large share of output in “Hotels and restaurants” is due to self-employed individuals (with zero employees), who are not captured by MultiProd for most countries.⁸

Figure 3.1. Total output, value added and employment relative to STAN, by SNA A38 industry

Note: Only countries-industries covered by both datasets in a given year are included. Gross output omitted for "Wholesale and retail" sector.

Source: OECD MultiProd v1.1 and STAN databases.

Figure 3.2. Total value added relative to STAN, by SNA A38 industry

Source: OECD MultiProd v1.1 and STAN databases.

The coverage for each country is shown in Figure 3.3. The values are aggregated to country-level taking into account only those industries which are present in both datasets, so that the results are not affected by different industry coverage (apart for potential industry misclassification).

This graph confirms a good overall coverage for all variables, in the range of 80-100% for three quarters of the countries. The coverage is somewhat lower in Switzerland, where only a sample of firms with less than 50 employees is available (the population is available for larger firms), and in the Netherlands, where only firms with at least 10 employees are included in the data.

For capturing aggregate trends, the stability of coverage over time is as important as its level. Figure 3.4 displays the coverage for each country over time. The coverage seems to lie within a relatively tight interval in most countries. Upward trends are noticeable in Hungary, Norway and, for gross output, Finland and Ireland. Some downward trend is visible for gross output in Italy, and for value added and employment in Finland. Denmark and Norway further show a temporary coverage dip around 2009, possibly reflecting some adjustments made in the national accounts during the economic downturn.

Overall, this section suggests that microdata underlying MultiProd offer a good and stable coverage for most countries and industries. Is the coverage good enough to allow MultiProd to replicate patterns observed in aggregate data? The next section turns to this question.

Figure 3.3. Total output, value added and employment relative to STAN, by country

Note: Manufacturing and non-financial market services only (excluding “Coke and refined petroleum”, “Real estate” and “Scientific R&D”). Figures for gross output exclude “Wholesale and retail”.

Source: OECD MultiProd v1.1 and STAN databases.

Figure 3.4. Total output, value added and employment rel. to STAN, by country over time

Note: Manufacturing and non-financial market services only (excluding "Coke and refined petroleum", "Real estate" and "Scientific R&D"). Only industries present in the data in all years are included. Figures for gross output exclude "Wholesale and retail".

Source: OECD MultiProd v1.1 and STAN databases.

4. Correlations: Can the microdata capture aggregate patterns?

This section examines whether MultiProd data are able to capture patterns observed in aggregate data. It explores correlations between values in MultiProd and STAN calculated over observations defined at the level of country, SNA A38 industry and year.

The main results are shown in Figure 4.1, which displays distributions of correlations between the two datasets. It contains three sets of correlation coefficients: (i) a set of correlations calculated over time for each combination of country and SNA A38 industry; (ii) a set of correlations calculated across SNA A38 industries for each country in each year; and (iii) a set of correlations across countries for each SNA A38 industry in each year.⁹ They test the ability of MultiProd data to capture, respectively, time trends, differences between industries, and cross-country variation. The correlations are calculated for gross output, value added, employment, labour productivity, and average wages. The labour productivity and average wages are calculated by dividing, respectively, the total value added and the total wage bill of each industry by its total employment. This graph and the subsequent graphs in this section exclude “Coke and refined petroleum”, “Real estate” and “Scientific R&D”.

Overall, MultiProd is quite successful in matching patterns observed in STAN, with median correlation above 0.75 for all variables and all types of variation (Figure 4.1).¹⁰ The correlations are also high for most of individual SNA A38 industries. Focusing on correlations in labour productivity over time, the median correlation is well above 0.5 for all industries except “Transportation and storage”, “Media” and “Administrative services” (Figure 4.2).

Figure 4.3 and Figure 4.4 capture the ability of MultiProd data to match aggregate patterns at country level. They focus again on labour productivity. Correlations over time (Figure 4.3) are very high for some countries (Germany, France, Sweden), with the median above 0.8 for most countries and in the 0.3-0.6 range for Norway, Australia, Canada, Ireland and Portugal. Correlations across industries (Figure 4.4) are generally higher than those over time, ranging from almost 1 for Portugal, Austria, Switzerland and Denmark to around 0.7 for Norway and Hungary.

Figure 4.1. Correlations between MultiProd and STAN, by variable and type of variation

Note: Manufacturing and non-financial market services only (excl. “Coke and refined petroleum”, “Real estate” and “Scientific R&D”). Left panel: correlations across years, calculated separately for each country-SNA A38 pair. Middle panel: correlations across industries, calculated separately for each country-year pair. Right panel: correlations across countries, calculated separately for each SNA A38-year pair. The graph plots the dispersion over country-SNA A38s (left), country-years (middle) or SNA A38-years (right) of these correlations.

Source: OECD MultiProd v1.1 and STAN databases.

Figure 4.2. Correlations in labour productivity across years, by SNA A38

Note: Manufacturing and non-financial market services only (excluding “Coke and refined petroleum”, “Real estate” and “Scientific R&D”). Correlations are calculated across years, separately for each country-SNA A38 pair. The graph plots, for each SNA A38 industry, the dispersion over countries of these correlations.

Source: OECD MultiProd v1.1 and STAN databases.

Figure 4.3. Correlations in labour productivity over time, by country

Note: Manufacturing and non-financial market services only (excluding “Coke and refined petroleum”, “Real estate” and “Scientific R&D”). Correlations are calculated across years, separately for each country-SNA A38 pair. The graph plots, for each country, the dispersion over SNA A38 industries of these correlations.

Source: OECD MultiProd v1.1 and STAN databases.

Figure 4.4. Correlations in labour productivity across industries, by country

Note: Manufacturing and non-financial market services only (excluding “Coke and refined petroleum”, “Real estate” and “Scientific R&D”). Correlations are calculated across industries, separately for each country-year pair. The graph plots, for each country, the dispersion over years of these correlations.

Source: OECD MultiProd v1.1 and STAN databases.

5. Conclusion

Better understanding about the drivers of aggregate economic trends, requires data that offer a representative picture of the underlying firm-level heterogeneity but are, at the same time, able to reproduce the patterns observed in the aggregate data that they are striving to explain. The OECD MultiProd project aims to generate such data by collaborating with a network of national experts who apply a harmonised statistical code to representative business microdata across a large number of countries. This paper compares the project's output to the OECD STAN database to test to what extent MultiProd data can be taken as reflecting the aggregate economies in question, and if they are able to reproduce patterns observed in aggregate data across years, industries and countries.

The results suggest that the MultiProd output captures a major part of gross output, value added and employment in most of the countries covered, and for both manufacturing and non-financial market services. They further show that MultiProd reproduces aggregate patterns relatively well, with median correlations over time, across industries and across countries between 0.75 and 1. Finally, they suggest that MultiProd, and similarly structured projects more generally, represent a promising source of information to shed new light on the micro drivers of the observed macroeconomic trends.

Endnotes

¹ See further details here: <http://www.oecd.org/sti/ind/multiprod.htm>

² While there are 88 2-digit Divisions defined in ISIC Rev.4, the A38 list consists of a set of aggregate sectors designed for use in National Accounts (SNA) statistics. See Part 4, pages 274-277, of https://unstats.un.org/unsd/publication/SeriesM/seriesm_4rev4e.pdf.

³ Only manufacturing industries are available for Japan.

⁴ “Employees” are defined as all persons who, by agreement, work for another resident institutional unit and receive remuneration. “Persons engaged” are instead defined as the total number of persons who work in or for the establishment, including working proprietors, active business partners and unpaid family workers, as well as persons working outside the establishment when paid by and under the control of the establishment. For the majority of countries MultiProd has information on employees or both, the only exception being the Netherlands, Portugal (only persons engaged) and Finland (person engaged expressed in full time equivalent).

The distinction between employees and persons engaged is relevant, for instance, when computing multi-factor productivity and wages: when computing multi-factor productivity, we should ideally take into account all persons involved in the production process, i.e., “persons engaged”; when computing wages as the ratio between cost of labour and employment, we would instead like to keep only salaried workers, i.e., “employees”. Therefore, whenever available, both measures are exploited as appropriate.

⁵ For more information, see <http://oe.cd/stan> and <https://doi.org/10.1787/na-data-en>.

⁶ Value added in factor costs corresponds to the sum of labour costs and the gross operating surplus.

⁷ “Wholesale and retail” shows coverage in terms of output much higher than 100% (omitted from the graph to preserve scale); this is because gross output in National Accounts only considers the trade margins on goods for resale rather than total sales.

⁸ As explained in Section 2.1, MultiProd excludes firms with less than one employment unit. For countries where employment is measured in terms of the number of employees, this means that self-employed individuals are dropped from the sample.

⁹ For example, for the correlations over time, a separate correlation coefficient is obtained for each combination of a country and an industry. Each coefficient measures the success of MultiProd data in capturing the variation across years for a given country and industry.

¹⁰ Note that in a very few country-industry cells, the correlations between MultiProd and STAN are negative. This likely reflects noise in the data due to the fact that each individual correlation coefficient is calculated over a small number of observations (e.g. the number of sample years for a given combination of a country and an industry). Thus, even relatively minor differences between MultiProd and STAN can in rare cases result in negative correlations. For this reason, it is important to focus on the distribution of the correlation coefficients (e.g. median, 25th percentile...) rather than a few outliers.

References

- Andrews, D., C. Criscuolo and P. Gal (2016), “The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy”, *OECD Productivity Working Papers*, No. 5, <http://dx.doi.org/10.1787/63629cc9-en>.
- Bajgar, M., G. Berlingieri, S. Calligaris, C. Criscuolo,, “To Use or Not to Use (and How to Use): Coverage and Representativeness of Orbis Data”, *OECD Science, Technology and Industry Working Papers*, forthcoming.
- Berlingieri, G., P. Blanchenay and C. Criscuolo (2017a), “The Great Divergence(s)”, *OECD Science, Technology and Innovation Policy Papers*, No. 39, <http://dx.doi.org/10.1787/953f3853-en>.
- Berlingieri, G., P. Blanchenay, S. Calligaris and C. Criscuolo (2017b), “The MultiProd Project: A Comprehensive Overview”, *OECD Science, Technology and Industry Working Papers*, No. 4, <http://dx.doi.org/10.1787/2069b6a3-en>.
- Berlingieri, G., S. Calligaris, and C. Criscuolo (2018a), “The Productivity-Wage Premium: Does size still matter in a service economy?”, *OECD Science, Technology and Industry Working Papers*, No. 2018/13, <https://doi.org/10.1787/18151965>.
- Berlingieri, G., S. Calligaris, and C. Criscuolo (2018b), “The Productivity-Wage Premium: Does Size Still Matter in a Service Economy?”, *AEA Papers and Proceedings*, Vol. 108, pp. 328–33.
- Desnoyers-James, I., S. Calligaris, and F. Calvino (2019), “DynEmp and MultiProd: Metadata”, *OECD Science, Technology and Industry Working Papers*, forthcoming.
- Gal, P. (2013), “Measuring Total Factor Productivity at the Firm Level Using OECD-ORBIS”, *OECD Economics Department Working Papers*, No. 1049, <http://dx.doi.org/10.1787/5k46dsb25ls6-en>.
- Kalemli-Ozcan, S., B. Sorensen, C. Villegas-Sanchez, V. Volosovych and S. Yesiltas(2015), “How to Construct Nationally Representative Firm Level data from the ORBIS Global Database”, *NBER Working Papers*, <https://ideas.repec.org/p/nbr/nberwo/21558.html> (Accessed on 16 August 2017).
- Melitz, M. and S. Polanec (2015), “Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit”, *The RAND Journal of Economics*, Vol. 46/2, pp. 362-375, <http://dx.doi.org/10.1111/1756-2171.12088>.
- OECD (2018), *OECD Economic Outlook*, Volume 2018 Issue 2, OECD Publishing, Paris, https://doi.org/10.1787/eco_outlook-v2018-2-en.
- Olley, G. and A. Pakes (1996), “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, Vol. 64/6, p. 1263, <http://dx.doi.org/10.2307/2171831>.
- Petrin, A. and J. Levinsohn (2012), “Measuring Aggregate Productivity Growth Using Plant-Level Data”, *The RAND Journal of Economics*, Vol. 43/4, pp. 705-725, <http://dx.doi.org/10.1111/1756-2171.12005>.
- Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer and G.T. De Vries (2015), “An Illustrated User Guide to the World Input-Output Database: the Case of Global Automotive Production”, *Review of International Economics*, Vol. 23/3, pp. 575-605, <http://dx.doi.org/10.1111/roie.12178>.

Annex A. Results based on unweighted data

Figure A.1. Total output, value added and employment relative to STAN, by SNA A38 industry (unweighted data)

Note: Only countries-industries covered by both datasets in a given year are included. Gross output omitted for "Wholesale and retail" sector.

Source: OECD MultiProd and STAN databases.