

Online Appendix to “Industry Concentration in Europe and North America”

Matej Bajgar¹, Giuseppe Berlingieri^{2,3,4}, Sara Calligaris²,
Chiara Criscuolo^{2,4} and Jonathan Timmis⁵

¹CERGE-EI; Charles University;

²OECD;

³ESSEC Business School;

⁴Centre for Economic Performance (LSE);

⁵World Bank

A Data

In this section of the Online Appendix, we describe the main data sources used in the paper.

A.1 MultiProd

One of the two main data sources used in the analysis is the OECD “MultiProd” project, a vast distributed microdata project undertaken in the last few years within the Directorate for Science, Technology and Innovation of the OECD. This section gives an overview of the project and its main data sources, and refers to a series of external technical documents that provide the full details on the methodology and representativeness of the data.

A.1.1 The project and the distributed microdata approach

Researchers and policy analysts have increasingly recognized the importance of microdata for understanding the complexity of economic dynamics and the heterogeneity in economic outcomes across firms. With this goal in mind, the MultiProd project studies firm-level productivity patterns across countries, sectors and time. It collects micro-aggregated data and moments of the productivity distribution that allow for a cross-country analysis of: i) productivity heterogeneity; ii) allocative efficiency; iii) misallocation; iv) aggregate

productivity growth; and v) the link between productivity and wages. The harmonized micro-aggregated data are crucial for investigating on a cross-country basis the extent to which different policy frameworks can shape firm productivity, and for examining the way resources are allocated to more productive firms.

While the availability of confidential micro-level data has expanded considerably for individual countries over the past decades, confidentiality concerns and other administrative issues still pose serious obstacles to the transnational access to official microdata. To circumvent these obstacles, the MultiProd project relies on the so-called “distributed microdata approach”. Such decentralized method consists in collecting statistical moments of the distribution of firm characteristics (employment, productivity, wages, age, etc.) by a centrally written, but locally executed, routine that is flexible and automated enough to run across different data sources in different countries. The advantages of this novel data collection methodology are manifold. It puts a lower burden on national statistical agencies, and limits running costs for such endeavors. Importantly, it also overcomes the confidentiality constraints of directly using national micro-level administrative databases, while at the same time achieving a high degree of harmonization and comparability across countries, sectors and over time. The approach was pioneered in the early 2000s in a series of cross-country projects on firm demographics and productivity (Bartelsman et al., 2005, 2009). The OECD currently follows the distributed microdata approach in three ongoing projects: MultiProd, DynEmp and MicroBerd.¹

This paper exploits a small portion of the output contained in the MultiProd data, and in particular the statistics reported by quantiles of the sales distribution. The full details of the methodology used in the project, as well as the main characteristics of the final MultiProd dataset, are available in Berlingieri et al. (2017) and Desnoyers-James et al. (2019).

A.1.2 Country coverage and data sources

This study relies on the version 1.1 of the MultiProd database (February 2019), which in general includes 20 countries (namely, Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Sweden and Switzerland). For the purpose of this study, we focus on 12 European countries during the period 2000-2014. While, for some countries, the data are available also in earlier years, the coverage is too limited to investigate patterns of concentration on a cross-country basis. The sample of countries is

¹The DynEmp (Dynamics of Employment) project provides harmonized micro-aggregated data to analyze employment dynamics (Criscuolo et al., 2014, 2015) and MicroBerd provides information on R&D activity in firms from official business R&D surveys (OECD, 2020).

not fully balanced over the period, with a shorter time coverage in particular for Ireland and Portugal.

Table A.1 reports the countries included in this study, the time coverage and the main data sources. More detailed country-specific information on the data sources can be found in Desnoyers-James et al. (2019).

Table A.1: Data coverage and sources

Country	Years	Data Sources
Belgium	2000-2014	Central Balance Sheet Office, National Social Security Office and VAT data
Denmark	2000-2012	Accounts Statistics for Non-Agricultural Private Sector, General Enterprise Statistics (Accounts statistics, VAT statistics, employment statistics, Business Register)
Finland	2000-2012	Structural business statistics data (surveys, corporate tax records, and Statistics Finland's Business Register)
France	2000-2014	FICUS/FARE, DADS, LIFI, SIRUS (Contours des entreprises profilées)
Germany	2003-2013	Four production surveys (JBU, KSE, IEU and AFiD-Panel Structure survey in the services sector)
Hungary	2000-2012	Corporate Income Tax data (CIT) of National Tax and Custom Administration
Ireland	2006-2014	CIP (Census for Industrial Production) for manufacturing, ASI (Annual Services Inquiry) for services, Business Register
Italy	2001-2014	ASIA (Business Register), Indagine sulle grandi imprese (SCI), Database Commercio Estero (COE), Balance sheet data for limited companies
Netherlands	2001-2014	Productiestatistieken (PS) and Algemeen Bedrijvenregister (BR)
Norway	2000-2012	Accounts statistics (incorporated firms) and business register
Portugal	2004-2012	Integrated Business Accounts System (IBAS): administrative compulsory sources and business annual fiscal reports
Sweden	2002-2012	SBS administrative data (tax data)

A.1.3 Representativeness

The primary source of data for the MultiProd database is administrative data covering the universe or near-universe of businesses with positive employment. For four countries in the sample (Germany, Ireland, Italy, the Netherlands) administrative data on the full population of firms do not exist, so the program relies on production surveys (PS) combined with a business register (BR). The former contain all the variables needed for

the analysis of productivity but may be limited to a sample of firms; the latter contains a more limited set of variables (mainly employment, sector of activity, age and ownership) but for the entire population of firms.

One of the big challenges of conducting cross-country analysis of firm-level data when only production surveys are available is that comparing selected samples of firms might yield a partial and biased picture of the economy. Whenever available, business registers, which contain the whole population of firms, are therefore used to compute a population structure by year-industry-size class.² This structure is then used to construct variable specific weights to re-weight the data contained in the production surveys in order to construct micro-aggregated data that are as representative as possible of the whole population of firms, and hence comparable across countries.³

Table A.2 details the sampling and census thresholds that characterize the data (by default MultiProd restricts the sample to firms with at least one employment unit). The sampling threshold is the minimum size for a firm to be included in the target population of the administrative or the survey data, the census threshold is the minimum size above which the data are exhaustive and contain the universe of businesses. The re-weighting strategy outlined above is relevant whenever the two thresholds differ. This is the case for Ireland (in services, and manufacturing in 2006-2007), Italy, and Netherlands. It would also be needed for the service sector in Germany, but unfortunately the MultiProd project has not yet gained access to the German business register.⁴

The ex-post re-weighting strategy adopted in MultiProd is not needed or not effective whenever the sampling and census thresholds coincide, and whenever the sampling threshold is above one, as there is no firm to represent the left tail of the size distribution. The first case clearly applies to the seven countries for which we have the universe of firms. It also applies to the manufacturing sector in Germany, where the data cover the universe of businesses above 20 employees. The second case is relevant for the Netherlands where the sampling threshold is 10, and hence small businesses are not represented.

The re-weighting is also not used in Hungary, where the administrative data present a revenue-based census threshold of around EUR 150,000. Due to simplified tax regimes (single-entry bookkeeping), very small firms are not included in the Corporate Income Tax data. The threshold is small and since it is revenue-based, the coverage of the administrative data increases over time (given the relatively high inflation rate until 2012). Hence, for incorporated businesses with positive employment, the data are almost ex-

²The population structure is defined at the 4-digit industry level and for eight size classes.

³The BR also allows for: i) a more precise treatment of entry and exit; ii) the calculation of more precise sectoral modes and conversion tables in case of changes in the sectoral classification over time.

⁴A new data collection is under way, and, in the new version of MultiProd, the German results will be based on the combination of both production surveys and the business register (URS Neu).

Table A.2: Sampling and Population Thresholds

Country	Sampling Threshold	Census Threshold	Business Register
Belgium	1 employee	1 employee	Not needed
Denmark	1 employee	1 employee	Not needed
Finland	1 person engaged	1 person engaged	Not needed
France	1 employee	1 employee	Not needed
Germany	Manufacturing: 20 employees; Services: EUR 17 500	Manufacturing: 20 employees (for JBU and IEU)	No (not needed in manuf.)
Hungary	Double-entry bookkeeping companies. Since 2004, mandatory for companies with turnover higher than HUF 50 million	Double-entry bookkeeping companies. Since 2004, mandatory for companies with turnover higher than HUF 50 million	Not needed
Ireland	CIP: 3 persons engaged up to 2007, no threshold from 2008 onwards; ASI: 2 persons engaged	CIP: all firms with at least 10 persons engaged are surveyed up to 2007, at least 3 persons engaged from 2008 onwards; ASI: all firms with more than 20 persons engaged	Yes
Italy	1 person engaged (limited companies)	1 person engaged (limited companies)	Yes
Netherlands	10 persons engaged	50 persons engaged up to 2009; from 2010, there is a limited number of sectors for which only firms with more than 100 persons engaged are census.	Yes
Norway	1 employee (limited companies)	1 employee (limited companies)	Not needed
Portugal	1 person engaged	1 person engaged	Not needed
Sweden	1 employee	1 employee	Not needed

Note: In case of variable-specific coverage, the table refers to the coverage of employment and turnover.

haustive. Finally, for Ireland, the re-weighting strategy does not play a large role in the manufacturing sector, for which, starting from 2008, only firms with one or two persons engaged are re-weighted. The service sector presents a sampling threshold of two persons engaged, so in principle firms with only one person engaged are not represented. However, since employment is measured in terms of persons engaged (i.e., including working proprietors or unpaid family workers) rather than employees, it is very unlikely that firms with positive employment are not covered.⁵ In any case, Section B.1 shows that all the results are robust to excluding both the countries in which the left tail of the size distribution is/might be missing (Germany, Hungary, Ireland, the Netherlands), and the countries where the ex-post re-weighting strategy is adopted to make the survey data representative and comparable across countries (Ireland, Italy, and Netherlands).

To summarize, the data included in the study cover: the universe of businesses with more than one employment unit in seven countries (Belgium, Denmark, Finland, France, Norway, Portugal, Sweden); the near universe in Hungary; and, data that are representative for the entire population of businesses in Italy and essentially Ireland from 2008. The two main exceptions are Germany, for which a business register is not available (and there is a 20 employee threshold in manufacturing), and the Netherlands, for which businesses below 10 persons engaged are not included. To verify that the MultiProd data are indeed representative, Bajgar et al. (2019) investigates to what extent the micro-aggregated data available in MultiProd reflect the patterns observed in aggregate sectoral data from national accounts. The exercise shows that the microdata used by MultiProd have a good coverage, representing typically about 80% of total gross output, value added and employment in manufacturing and non-financial services. In addition and most importantly, the coverage is reassuringly stable over time. The results also show that MultiProd well reproduces aggregate patterns, with median correlations over time, across sectors and across countries between 0.75 and 1.⁶

A.2 Orbis-Worldscope-Zephyr

The second main data source – used independently of MultiProd – consists of matched Orbis-Worldscope-Zephyr data. Our concentration measure based on this source uses information on business group structure to aggregate subsidiary sales within the same region and industry. Doing so requires both group and subsidiary financial information,

⁵If anything, the coverage might actually be larger, because firms with two persons engaged might not employ any worker, and the coverage in MultiProd is typically limited to firms with positive employment.

⁶Any comparison between micro-level data and aggregate sectoral data from national accounts has to be taken with some caution because the two sources of data are designed for different purposes. The construction of aggregate (SNA08) statistics involves many adjustments, which imply that even administrative data covering the whole population of firms will not exactly reproduce aggregate statistics in National Accounts.

as well as ownership information detailing parent-subsidary linkages. This section gives an overview of the main data sources of financial information of business groups and their subsidiaries and the corresponding ownership information, and it provides additional detail on how concentration indicators are constructed from this data.

A.2.1 Historical Orbis financial data

Orbis is the largest cross-country firm-level database that is available and accessible for economic and financial research. The Historical Orbis database used in this paper is designed to mirror alternative approaches to obtaining reliable historical data, such as merging of multiple CDs containing snapshots at one point in time.

To ensure that we capture all of each group’s subsidiaries, we use broad information for firms of all sizes, in all business sectors. Specifically, we include all industries excluding public services – 2-digit NACE rev.2 codes 1-82.

We undertake a number of cleaning steps, closely following the suggestions by Kalemli-Özcan et al. (2019), Gal (2013), and Andrews et al. (2016). These include dropping duplicate observations and removing outliers identified as implausible changes or ratios.

Financial information within Orbis is available at the business group-level (“consolidated” financial data aggregated across subsidiaries of the firm) and individual firm-level (“unconsolidated” information referring to an individual firm). As we discuss later, our approach fundamentally relies upon unconsolidated data of the individual subsidiaries within a group.

The business-group level “consolidated” information is only used in cases where the total subsidiary sales exceed group sales (due to inter-company transactions) or when the headquarters firm’s unconsolidated data is missing.⁷ In the former case, total subsidiary sales are scaled down to match group sales. This affects about 14% group-region-industry-year observations belonging to relevant business groups.⁸ In the latter case we manually search financial statements to fill in the missing information for the largest 1000 groups, which we discuss later, and otherwise impute headquarters firm sales from the residual between group and total subsidiary sales, changing about 2% observations of relevant business groups.

Our approach requires good coverage of both business-group and individual subsidiaries’ financial information. We expand the coverage of Orbis data by using available unconsolidated information to infer missing years in the consolidated information of the same firm and vice versa⁹. About 6% of observations belonging to relevant business groups

⁷This scaling approach implicitly assumes that each subsidiary has a similar share of inter-company sales as a proportion of their total sales. Subsidiary-specific data on inter-company sales is not available.

⁸I.e., business groups that appear among the top 8 in at least one world region, industry and year.

⁹To ensure unconsolidated trends are a reliable proxy, we require unconsolidated sales to be at least

are affected by this step. We have tested the accuracy of this approach by examining its ability to predict our actual Orbis observations, randomly setting half our actual observations to missing.¹⁰ Our approach closely replicates actual Orbis observations: with a 95% correlation of between the predicted series and actual sales.

A.2.2 Worldscope financial data

We further expand the coverage of our consolidated data by combining Worldscope and Orbis data using firm International Securities Identification Number (ISIN) numbers, which uniquely identify listed firms. For some countries, such as the US, and some years, Worldscope is able to improve the coverage of Orbis.¹¹

Worldscope is a cross-country firm-level commercial database of listed firms (Orbis also includes non-listed) representing about 95% of global stock market capitalization. We apply the same cleaning rules to the Worldscope data as noted above for Orbis data. Worldscope reflects consolidated financial data and contains very similar values to Orbis consolidated data – for observations present in both datasets.¹²

We supplement Orbis data with Worldscope in two ways: adding missing years for firms that are already in Orbis and adding missing firms that never have financial data in Orbis.¹³ Overall, adding Worldscope affects about 0.5% observations belonging to relevant business groups. We tested the accuracy of this approach by examining its ability to predict our actual Orbis observations, randomly setting half our actual observations to missing. Our approach closely replicates actual Orbis observations: we find a 99% correlation between the predicted series and actual sales.

50% of consolidated firm sales and therefore represent the bulk of firm activity (for the firm on average). Furthermore, we censor the unconsolidated trends at the 5th and 95th percentiles, so we do not use extreme growth periods. Note that we do not apply thresholds to using consolidated trends to fill in unconsolidated sales, since if unconsolidated sales are only a small proportion of consolidated sales any errors are unlikely to impact the group structure substantially.

¹⁰Specifically, we randomly set half the actual consolidated Orbis observations to missing that also have unconsolidated information. We then use the unconsolidated sales growth rates to fill in the missing numbers and see how close the filled in and actual observations are.

¹¹We find that this leads to approximately 10,000 additional firm observations each year in North America and 5,000 in Europe.

¹²For the observations that are both in Orbis and in Worldscope, it is possible to compare the consolidated information in the two datasets. Most firms have very similar sales data in both Orbis and Worldscope, with 50% of firms having sales within approximately 5-7% in each data source and sales growth within 9-10% in each data (considering 25th and 75th percentiles). This supports the appropriateness of combining the two sources.

¹³For filling in missing years we use sales growth rates (rather than levels) from Worldscope, to avoid jumps when the data source changes. In addition, to remove possible cases of matching the wrong firms in the two data sets, we also only use growth rates for firms that are sufficiently similar, with sales on average within 35% in the Worldscope and Orbis in the years when we observe them in both.

A.2.3 Orbis-Zephyr ownership information

We use Orbis-Zephyr ownership information to identify subsidiaries that are part of the same business group. The ownership information underlies our concentration measure, which aggregates subsidiary sales within the same region and industry. Our primary source of ownership information is Historical Orbis, which we also supplement with data on the Zephyr database of Mergers and Acquisitions (M&As). Both datasets are provided by Bureau Van Dijk (BvD) and share a common firm identifier, which allows us to merge them. Orbis contains comprehensive information on ownership linkages between firms, which has been extensively used in the academic literature (e.g., Cravino and Levchenko, 2017), however the data primarily start in 2007. The Zephyr M&A database allows us to measure earlier changes in ownership, enabling the construction of a firm-level ownership time-series starting as early as 2000 for some firms.¹⁴

We use each subsidiary's global ultimate owner in Historical Orbis to identify subsidiaries that are part of the same business group. The global ultimate owner is available at each year-end from 2007 to 2016 and is calculated by BvD. To calculate ultimate owners, BvD use the tree of ownership linkages for each firm, every year. They identify each firm's immediate shareholders, then their shareholders' shareholders and so on. We use the ultimate owner defined using the 50.01% minimum ownership percentage, since this is a common definition of control of another firm and hence whether the subsidiary financials are consolidated into the parent accounts.¹⁵

We improve the coverage of ultimate ownership data, particularly for earlier years, using ownership changes identified from the Zephyr M&A database and Orbis historical ownership linkages. The historical ownership linkage information in Orbis is particularly large and available since the early 1990s, covering approximately 116 million subsidiaries. Zephyr contains deal-level information on M&As from 1997 for European firms, from 2000 for North America and global deals from 2003. Given our focus on Europe and North

¹⁴Whilst ultimate ownership data starts in 2007, for some firms it is not available until later years. Common approaches to correcting for this in the literature are either to assume that firms without an Orbis ultimate owner are independent or to take data from a recent year - assuming ownership has not changed over time. Both of these approaches are problematic. With increasing coverage of ownership over time in Orbis, the former approach will falsely equate missing data with independence and lead to an overstatement of ownership changes over time. The latter approach will lead to an understatement of ownership changes over time and will typically overstate the number of markets and countries in which a firm operates.

¹⁵A majority of voting rights is a sufficient condition for control under international accounting practices. For example, International Accounting Standards 27 and 28 require consolidation for majority ownership stakes (exceeding 50%), whereas minority stakes and joint ventures are not consolidated. Whilst we do not observe voting rights, we assume these are reflected in shareholder ownership percentages, such that a majority (50.01%) of shares reflects a majority of voting rights. We reflect majority owned subsidiaries (exceeding 50%) as part of the same business group, whereas we treat joint ventures and partially owned firms as separate entities to their non-majority owners.

America, this constrains our analysis to 2000. Zephyr contains about 1.4million M&A deals from 2000 to 2016. The Zephyr M&A database allows us to roll back ultimate ownership information, until there is a deal representing a change in majority ownership. We use the vendor to measure the immediate owner before the deal transaction. Similarly, the Orbis historical ownership linkages also allows us to measure the immediate owner for earlier years (using the 50.01% ownership threshold). We follow the same exercise as BVD to find the ultimate owner, calculating the immediate owner’s majority shareholder and so on.

We also undertake an extensive array of cleaning steps to identify and correct potential issues in the ownership data –focused towards identifying missing linkages amongst the largest firms. First, we correct ultimate owners that are in fact majority owned by another firm, since by definition they cannot be an ultimate owner. Second, we remove temporary (one or two year) changes in ultimate owner that reverse themselves, as this seems highly unlikely to occur in reality. Third, to detect missing linkages, we examine large firms (sales exceeding 100 million Euros) that change from having no subsidiaries to a large number of subsidiaries from one year to the next. Fourth, to identify missing links we examine large firms that never have any subsidiaries and the opposite case, large groups of subsidiaries that never have a parent with financials. Fifth, we identify missing links where there are ownership changes amongst firms that have very similar names and are so very likely part of the same group all along (e.g., ABC Motors acquired by ABC Motors Thailand). Finally, we manually check 300 of the largest groups, using the subsidiary structure listed in their financial statements each year to cross-check against the resulting ownership data. Overall, these steps involve manually checking around 1000 of the largest firms (sales exceeding 1 billion Euros), semi-automated procedures based on name-matching algorithms for another 4000 large firms (exceeding 100 million Euros) and automatically correcting another 5000 firms.

A.2.4 Resulting Sample and Further Manual Checks

The combined Orbis-Worldscope-Zephyr data allows measures of concentration at the level of 2-digit NACE rev.2 industries and at world regions. One caveat is that our data currently have poorer coverage of subsidiary-level information for US firms. Note this is not the case for US business-group level information, since Orbis and Worldscope appear to cover close to the universe of listed firms.

We also undertake extensive and time-consuming manual checks of the resulting trends in sales for 1000 of the largest groups. The starting point is to graph trends in sales for each of the largest 8 groups in every industry and world region in our sample, which we

then manually investigate for anomalous jumps.¹⁶ We investigate anomalies by comparing to data available from financial statements and SEC filings. The resulting corrections relate to differences in ownership and financial information, for instance due to timing of identifying M&As or divestments, dual stock market listings, years of missing subsidiary data, typos and so on.

We report concentration metrics only for the well-covered economies and industries (see Bajgar et al., 2020). We present concentration metrics for manufacturing and non-financial market services (excluding 19 – Manufacture of coke and refined petroleum products, 68 – Real estate activities and 70 – Activities of head offices; management consultancy activities).¹⁷ We present analyses for Europe and North America, based on years 2000-2014 for 21 economies. The metrics for Europe are based on Belgium, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, the Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden and the UK, and those for North America reflect the US and Canada.

¹⁶Since our empirical analysis is in differences, we focus on issues identified in changes rather than levels.

¹⁷For North America we also exclude 14 – Manufacture of wearing apparel, 15 - Manufacture of leather and related products and 30 - Manufacture of other transport equipment.

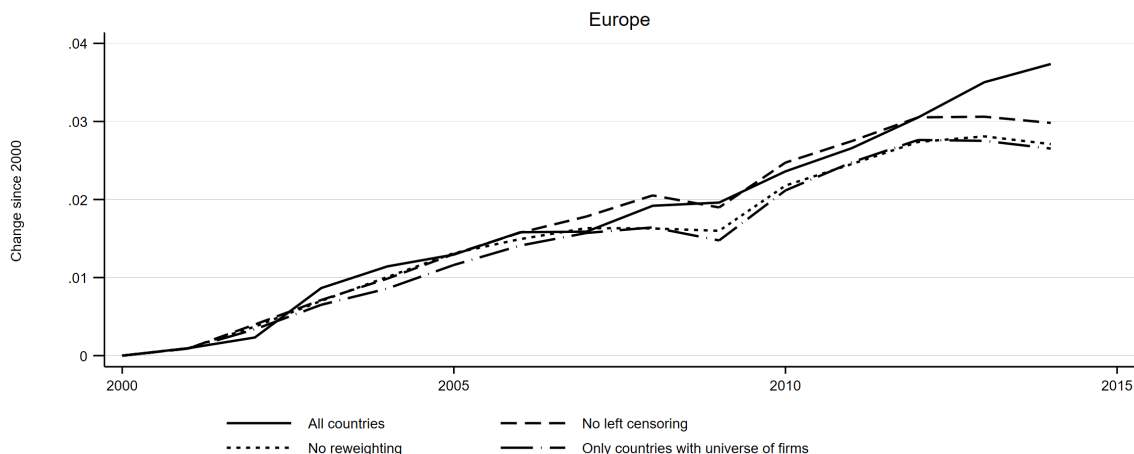
B Robustness exercises

This section describes a series of robustness exercises to demonstrate the robustness of our results to different data or methodological choices.

B.1 Concentration in European countries, MultiProd data

In Figure B.1, we show that the average patterns of industry concentration across European countries in the MultiProd data are a result neither of specific trends in countries with some data limitations, nor of the re-weighting strategy adopted to make the data representative and comparable across countries.

Figure B.1: Share of sales accounted for by 10% largest firms - Different samples



Note: Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of sales of the firms in the top decile of the sales distribution in each country and 2-digit industry. The graph reports the cumulative weighted average change in industry concentration, with equal country weights and industry weights given by each industry's share in the total sales across all industries within a given country. Countries in the full sample are BEL, DEU, DNK, FIN, FRA, HUN, IRL, ITA, NLD, NOR, PRT and SWE. Countries with no left censoring include BEL, DNK, FIN, FRA, ITA, NOR, PRT and SWE. Countries with no re-weighting include BEL, DEU, DNK, FIN, FRA, HUN, NOR, PRT and SWE. The countries with the universe of firms available include BEL, DNK, FIN, FRA, NOR, PRT and SWE.

The first line replicates our baseline results for the full sample of countries, combining manufacturing and non-financial market services. The second line corresponds to a sample that excludes any country that has a sampling threshold that censors the left tail of the size distribution (Germany, Hungary, Ireland, the Netherlands).¹⁸ The third line corresponds

¹⁸Note that we adopt a conservative strategy and drop also countries for which the underlying data come from official administrative data covering the near universe of firms. This is for instance the case of Hungary where, due to simplified tax regimes, the official Corporate Income Tax data do not cover firms with a turnover below around EUR 150,000.

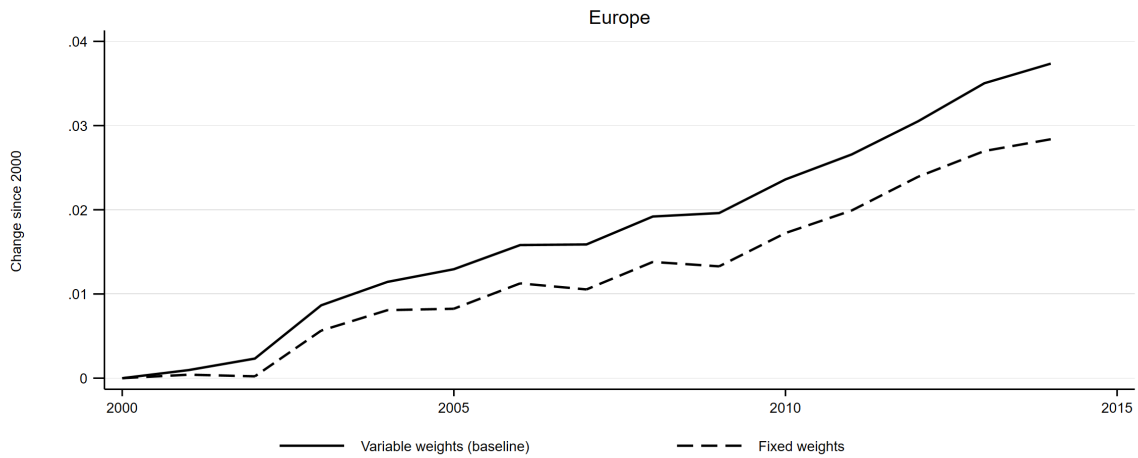
to a sample excluding the countries where the ex-post re-weighting strategy is adopted (Ireland, Italy, the Netherlands). The re-weighting strategy is adopted in MultiProd to ensure that, whenever the data come from a stratified random survey, the collected aggregate moments are representative for the entire population of businesses, and hence comparable with countries for which the universe of firms is available. Finally, the last line corresponds to the most conservative sample, consisting only of countries for which administrative data for the universe of firms are available.

Figure B.1 reveals that the concentration trends are very similar across the four different country samples, with the exemption of the last two years, when concentration in the full sample continued to grow but it remained flat in the more restrictive samples. However, the trend is less robust in the last two years mainly because of the unbalanced nature of the underlying sample, with only five countries available in 2014 (Belgium, France, Ireland, Italy, the Netherlands). Overall, the results demonstrate that the observed increase in concentration is not driven by countries where a universe of firms is not directly observed.

We then show that the average patterns of industry concentration across European countries are not primarily a result of a change in industrial composition. In the baseline results we allow for the industry weights to vary over time, while Figure B.2 compares the baseline results with the case in which industry weights are kept constant to their initial level. The overall trend in industry concentration is somewhat smaller, as a result of the larger change in industry concentration observed in market services combined with the on-going process of structural change characterized by the usual fall in the share of the manufacturing sector. However, even when the industrial composition is kept constant over time, the average increase in industry concentration remains sizable and very close in magnitude to the change in concentration obtained with Orbis data.

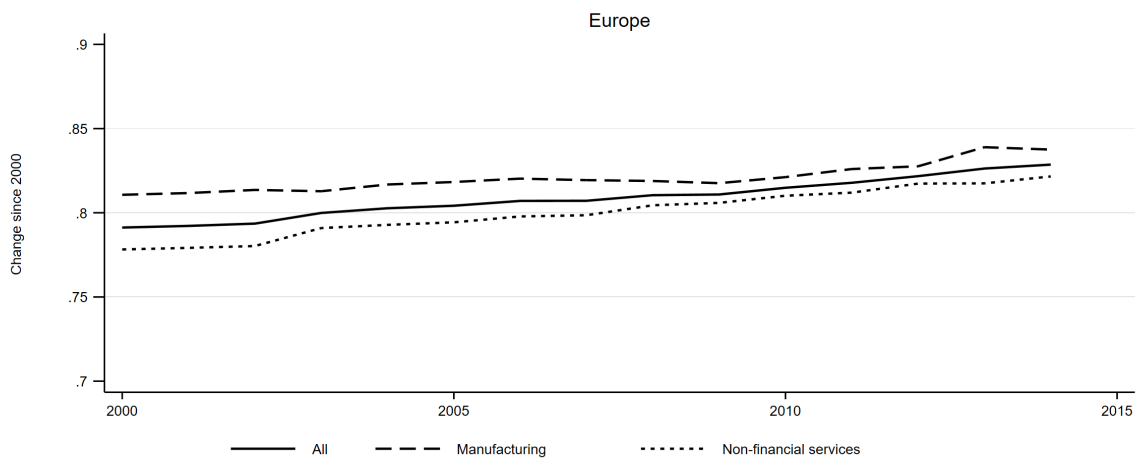
Finally, Figure B.3 displays the average levels of industry concentration, showing that market services appear to be less concentrated on average but with a level that has been approaching that of manufacturing by the end of the period. The 10% largest firms capture on average more than three quarter of total sales in each industry-country pair, a share that is larger than what is captured by the 8 largest business groups in the whole European region (see Figure B.7). The higher level of concentration is expected given that the analysis is at the country level and that the median national industry contains over 200 firms: the concentration measure available in the MultiProd data is closer to the CR20 measure constructed with Orbis data.

Figure B.2: Share of sales accounted for by 10% largest firms - Impact of industry weighting



Note: The countries included are BEL, DEU, DNK, FIN, FRA, HUN, IRL, ITA, NLD, NOR, PRT and SWE. Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of sales of the firms in the top decile of the sales distribution in each country and 2-digit industry. The graph reports the cumulative weighted average change in industry concentration, with equal country weights and industry weights given by each industry's share in the total sales across all industries within a given country. The baseline trends allow for weights to vary over time, while in the 'Fixed weights' case the weights are kept constant to their initial level.

Figure B.3: Share of sales accounted for by 10% largest firms - Levels



Note: The countries included are BEL, DEU, DNK, FIN, FRA, HUN, IRL, ITA, NLD, NOR, PRT and SWE. Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of sales of the firms in the top decile of the sales distribution in each country and 2-digit industry. The graph reports the cumulative weighted average change in industry concentration, with equal country weights and industry weights given by each industry's share in the total sales across all industries within a given country. To display levels, each line has been shifted up to match the weighted average concentration level across the given countries and industries in 2007.

B.2 Concentration in Europe & North America, Orbis-Worldscope-Zephyr data

In this section we examine the robustness of the baseline concentration metrics to key methodological choices.¹⁹ First, we examine the impact of our approach that aggregates subsidiary data compared to using either subsidiary or group data in isolation. Second, we examine the impact of scaling the total subsidiary data to match group consolidated accounts. Third, we show the key role that multinationals play in explaining our concentration trends and we provide the trends for the average *levels* of industry concentration.²⁰ We focus on European metrics in this section due to the better coverage of unconsolidated subsidiaries and because the increasing concentration in the United States is less contentious as it has now been documented by a number of studies using different data and methodologies.

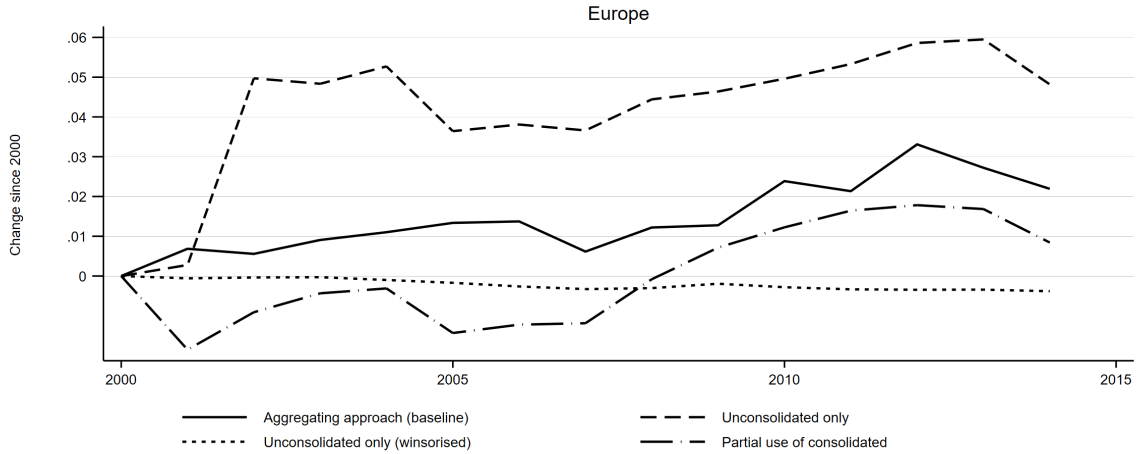
How one treats subsidiary financial information and whether one considers subsidiaries individually or aggregates subsidiary sales (within the same region and industry) can lead to significantly different concentration trends (Figure B.4). When we treat each subsidiary as independent and use only Historical Orbis unconsolidated financial data (“Unconsolidated only”), we find a highly volatile concentration in the initial years – likely due to data coverage issues – and a moderately increasing trend in the subsequent years. A simple winsorizing strategy of the top and bottom 0.1% of firms in terms of sales solves the volatility problem but throw away a lot of relevant information and results in an essentially flat concentration pattern (“Unconsolidated only (winsorised)”). This should not be surprising. In an average industry in the sample, we observe sales for about 50,000 firms. Winsorizing the top 0.1% firms is then equivalent to muting the effect of the largest 50 firms, which is clearly problematic when trying to calculate the top 8 concentration. This problem is very similar to the top coding issue in the large literature that investigates top income inequality (see Atkinson et al., 2011, for a review). Removing information at the top is not the solution and this is why we run extensive manual and semi-automated checks for the largest firms to minimize measurement error at the very top (see Section A.2).

When we still treat subsidiaries as independent but expand coverage of unconsolidated data by also using available consolidated information (from Worldscope and Orbis) to infer missing years of the same firm (“Partial use of consolidated”), we find a largely flat concentration trend until around the financial crisis, with an increase thereafter of

¹⁹Also note the choice of the denominator in the concentration ratio, discussed in Section 3.4 of the paper.

²⁰Examining the role of multinationals in our data was inspired by their growing importance documented in Kalemlı-Özcan et al. (2019).

Figure B.4: Impact of data cleaning and aggregating over subsidiary accounts



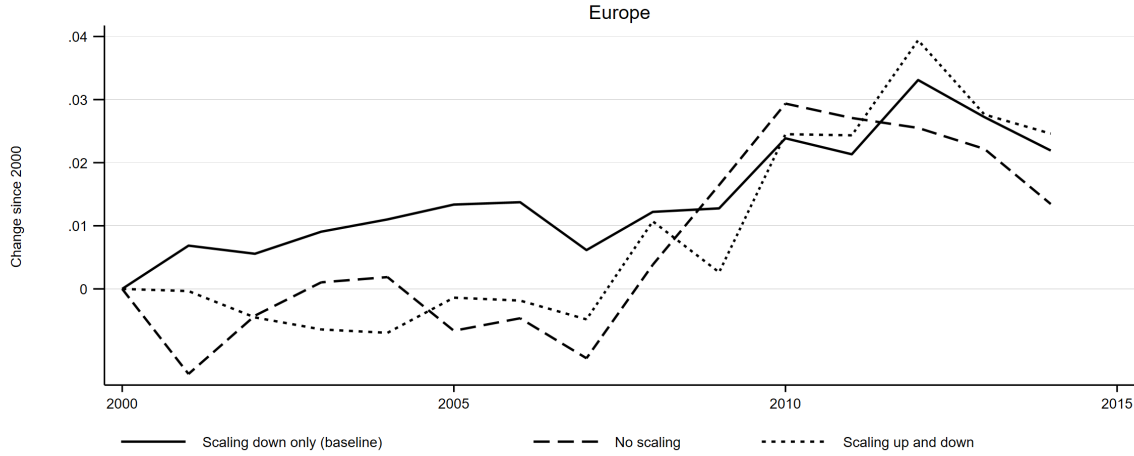
Note: The countries include BEL, DEU, DNK, EST, ESP, FIN, FRA, GBR, GRE, HUN, IRL, ITA, LVA, NLD, NOR, POL, PRT, SVN and SWE. Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of the top 8 business groups in each industry (CR8). The graph reports the cumulative weighted average change in industry concentration, with weights given by each industry’s share in the total sales across all industries within a given world region. The “Baseline” aggregates subsidiary sales within the same region and industry using corrected Orbis-Worldscope-Zephyr data. “Unconsolidated only” represents the simplest approach, focusing only on individual firms and using Historical Orbis unconsolidated financial data only. “Unconsolidated only (winsorised)” applies the previous approach together with a simple winsorizing strategy for the bottom and top 0.1% firms in terms of sales. “Partial use of consolidated” again focuses on individual firms but uses available consolidated information (from Worldscope and Orbis) to infer missing years of the same firm.

Source: Orbis-Worldscope-Zephyr and OECD-STAN.

2 percentage points. Note, however, that both of these approaches ignore the fact that, on average, a top 8 business group in our data consists of 20 subsidiaries operating in the same region and industry. When we account for this using our baseline aggregating approach, we document a smooth rise in concentration by a bit over 3 percentage points between 2000-2012, followed by a mild dip in the last two years.

Another choice surrounds the treatment of cases where the sales of subsidiaries do not sum to the total consolidated sales of the business group. To some extent this is to be expected, because of intra-group sales where one part of the firm purchases inputs from another part. These intra-company transactions are removed from the group accounts but show up in each subsidiary’s sales. To prevent our results from being inordinately influenced by groups with a large share of intra-group sales, in our baseline specification we proportionally scale down the sales of all subsidiaries when subsidiary sales sum to more than the group. Omitting such scaling implies a somewhat weaker overall increase in concentration, with most of the increase concentrated in the period 2007-2010 (Figure B.5). There are also cases where subsidiary sales sum to less than the group sales, which

Figure B.5: Impact of scaling to match group consolidated accounts



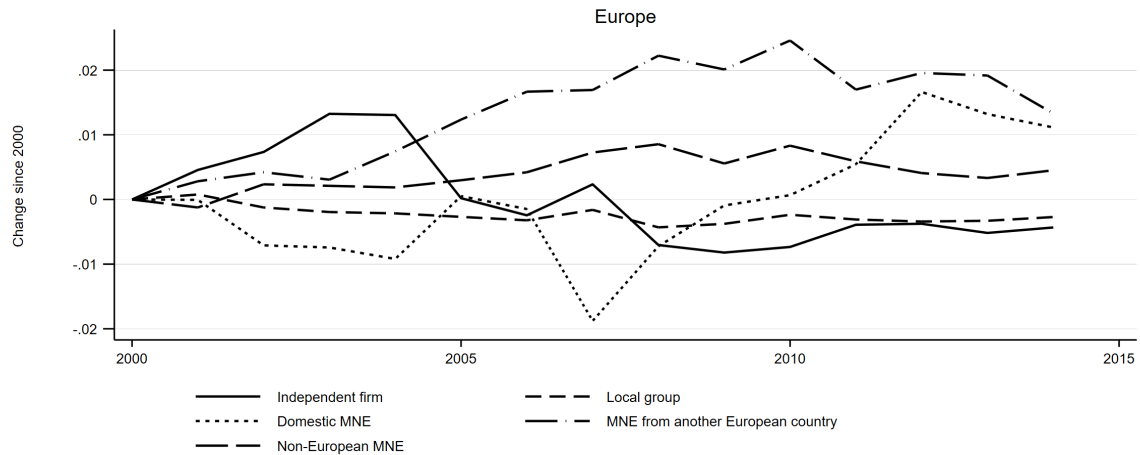
Note: The countries include BEL, DEU, DNK, EST, ESP, FIN, FRA, GBR, GRE, HUN, IRL, ITA, LVA, NLD, NOR, POL, PRT, SVN and SWE. Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of the top 8 business groups in each industry (CR8). The graph reports the cumulative weighted average change in industry concentration, with weights given by each industry’s share in the total sales across all industries within a given world region. Concentration metrics reflect the share of the top 8 firms in each industry (CR8). The graphs can be interpreted as the cumulated absolute changes in levels of sales concentration for the mean 2-digit sector within each region. The “Baseline” scales the total subsidiary sales down when they exceed the group consolidated sales. “No scaling” represents the approach without such correction. “Scaling down and up” involves scaling subsidiary sales in either direction to match group consolidated sales.

Source: Orbis-Worldscope-Zephyr and OECD-STAN.

likely reflects missing subsidiary data. It is, in principle, possible to correct these cases by scaling up the subsidiary sales so that they add up to the group sales. When we scale subsidiary sales up, as well as down, the overall concentration change over the sample period is very similar to our baseline, but the increase is concentrated in the middle of the sample period, with flat concentration prior to 2007 and a mild drop after 2012. The scaling up implicitly assumes that the missing data is distributed in the same industries and regions as the sample of subsidiary sales that we observe. In contrast, not scaling up implicitly assumes that the missing subsidiary data is located in regions and industries outside the sample. Given that our sample consists of countries with relatively good and consistent coverage in Orbis, we think that the latter assumption will more often hold in our data and consequently opt for not scaling up as our baseline approach.

We further illustrate the importance of multinationals (MNEs) for European concentration growth, reiterating the importance of considering subsidiary ownership structure. In Figure B.6 we decompose the European market share of the top 8 groups. Specifically, we show the market share of independent firms, local groups (with subsidiaries only in the home country), European MNEs in their headquarters country, European MNEs in other European countries and MNEs with headquarters outside Europe. We find that the

Figure B.6: Role of multinationals in concentration trends



Note: The countries include BEL, DEU, DNK, EST, ESP, FIN, FRA, GBR, GRE, HUN, IRL, ITA, LVA, NLD, NOR, POL, PRT, SVN and SWE. Included industries cover 2-digit manufacturing and non-financial market services. The figure disaggregates the market share of the top 8 firms in each industry (CR8) into the market share of (i) independent firms; (ii) local groups – i.e. with subsidiaries only in the home country; (iii) European MNEs in their headquarters’ country; (iv) European MNEs in other European countries; and (v) non-European MNEs – i.e., with headquarters outside Europe. The graphs can be interpreted as the cumulated absolute changes in levels of sales for the mean 2-digit sector within each region.

Source: Orbis-Worldscope-Zephyr and OECD-STAN.

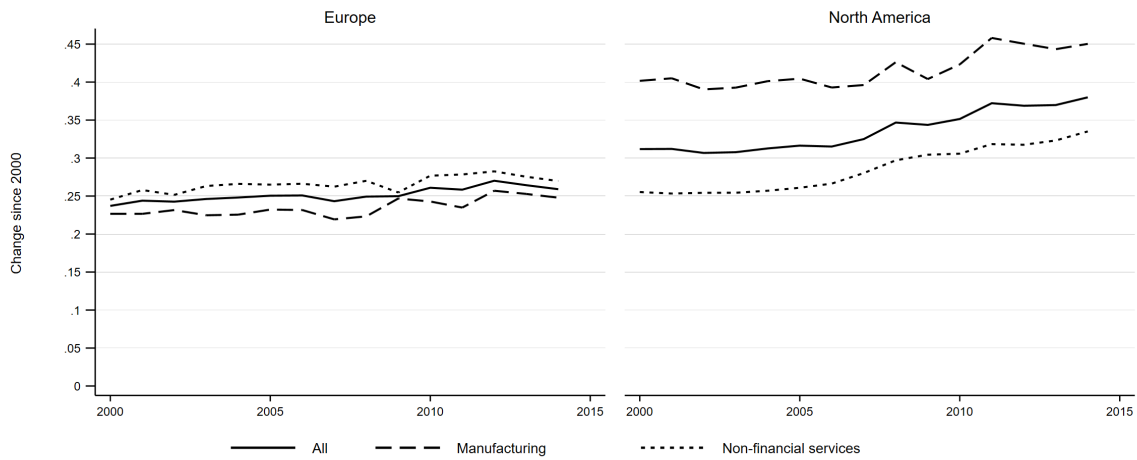
market shares of independent firms, local groups and non-European MNEs (in the top 8) have been largely flat over the sample period, with most of the overall concentration increase driven by European MNEs, both in their headquarters and non-headquarters countries. This growth of multinational sales in Europe is in line with the finding of Kalemli-Özcan et al. (2019), who highlight the contribution of MNEs to the increase in industry concentration in Europe.

Finally, Figure B.7 displays the average levels of industry concentration, by computing analogous weighted averages across industries. The results show that industries appear to be more concentrated on average in North America, where manufacturing displays the highest concentration levels. Despite several methodological differences, the values are broadly consistent with the sales and employment concentration levels for the United States in Autor et al. (2020) and Ganapati (2021).

B.3 Sectoral heterogeneity in concentration in Europe & North America, Orbis-Worldscope-Zephyr data

In this section, we report our baseline results using the Orbis-STAN dataset at the detailed 2-digit sectoral level. Table B.1 reports the cumulative change in industry concentration

Figure B.7: Share of sales accounted for by 8 largest business groups in Europe & North America - Levels



Note: The countries include BEL, DEU, DNK, EST, ESP, FIN, FRA, GBR, GRE, HUN, IRL, ITA, LVA, NLD, NOR, POL, PRT, SVN and SWE for Europe and CAN and USA for North America. Included industries cover 2-digit manufacturing and non-financial market services. Concentration metrics reflect the share of the top 8 business groups in each industry (CR8). The graph reports the cumulative weighted average change in industry concentration, with weights given by each industry’s share in the total sales across all industries within a given world region. To display levels, each line has been shifted up to match the weighted average concentration level across the given industries in 2007.

Source: Orbis-Worldscope-Zephyr and OECD-STAN.

between 2000 and 2014 expressed in percentage points, for each of the 2-digit (STAN A64) sectors included in our analysis. Examples of sectors that display similar patterns between Europe and the US include the increase in concentration in textiles, machinery equipment, wholesale trade, publishing, audiovisual & broadcasting, architectural and engineering services. The changes in concentration are typically higher in the US, including well-known examples such as retail trade or air transport in which concentration has increased markedly in the US, but barely so in Europe.

Table B.1: Change in CR8 Concentration in Europe & North America

A64 Sector	Europe	US
10: Manuf. of food, bev. and tobacco	-4.31	-1.29
13: Manuf. of text., apparel and leather	7.51	30.37
16: Manufacture of wood	3.36	-7.27
17: Manufacture of paper	-1.76	4.17
18: Printing and recorded media	-0.51	39.76
20: Manufacture of chemicals	9.65	0.69
21: Manufacture of pharmaceuticals	-2.37	-22.33
22: Manufacture of rubber	2.66	-0.36
23: Manufacture of other minerals	0.68	0.71
24: Manufacture of basic metals	-0.59	-0.38
25: Manufacture of fabricated metal	0.59	0.59
26: Manufacture of computers	-2.18	89.02
27: Manufacture of electrical eq.	0.91	21.86
28: Manufacture of machinery eq.	3.20	12.10
29: Manufacture of motor vehicles	13.85	-21.56
30: Manufacture of other transport	7.84	-45.20
31: Furniture, other manuf.	1.84	3.30
33: Repair of machinery and eq.	-3.50	-
45: Trade and repair of motor vehicles	9.35	4.03
46: Wholesale trade	9.39	15.16
47: Retail trade	1.38	25.08
49: Land transport	0.38	19.40
50: Water transport	12.93	-
51: Air transport	3.47	28.99
52: Warehousing	5.55	-
53: Postal and courier activities	-19.41	-
55: Accommodation and food services	0.15	2.05
58: Publishing	4.85	17.28
59: Audiovisual and broadcasting	7.15	12.84
61: Telecommunications	-12.39	-3.80
62: IT and information services	-3.00	-2.87
69: Legal, accounting and head office	-0.40	0.34
71: Architectural and engineering	6.59	8.45
72: Scientific R&D	5.53	1.44
73: Advertising and market research	3.98	1.15
74: Other professional and veterinary	2.79	9.62
77: Rental and leasing	1.65	1.08
78: Employment activities	-2.90	-4.85
79: Travel agency and related	-1.83	-
80: Security, buildings, office	2.44	-1.92

Note: The countries include BEL, DEU, DNK, EST, ESP, FIN, FRA, GBR, GRE, HUN, IRL, ITA, LVA, NLD, NOR, POL, PRT, SVN and SWE for Europe and CAN and USA for North America. Included industries cover manufacturing and non-financial market services. Concentration metrics reflect the share of the top 8 business groups in each industry (CR8). The table reports the cumulative change in industry concentration between 2000 and 2014 expressed in percentage points.

Source: Orbis-Worldscope-Zephyr and OECD-STAN.

References

- Andrews, D., Criscuolo, C., and Gal, P. N. (2016). The best versus the rest: The global productivity slowdown, divergence across firms and the role of public policy. OECD Productivity Working Papers 5, OECD Publishing.
- Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71.
- Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.
- Bajgar, M., Berlingieri, G., Calligaris, S., and Criscuolo, C. (2019). Can firm micro data match macro trends? Comparing MultiProd and STAN. OECD Science, Technology and Industry Working Papers 2019/02, OECD Publishing.
- Bajgar, M., Berlingieri, G., Calligaris, S., Criscuolo, C., and Timmis, J. (2020). Coverage and representativeness of Orbis data. OECD Science, Technology and Industry Working Papers 2020/06, OECD Publishing.
- Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2009). Measuring and analyzing cross-country differences in firm dynamics. In *Producer Dynamics: New Evidence from Micro Data*, pages 15–76. University of Chicago Press.
- Bartelsman, E., Scarpetta, S., and Schivardi, F. (2005). Comparative analysis of firm demographics and survival: Evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3):365–391.
- Berlingieri, G., Blanchenay, P., Calligaris, S., and Criscuolo, C. (2017). The MultiProd project: a comprehensive overview. OECD Science, Technology and Industry Working Papers 2017/04, OECD Publishing.
- Criscuolo, C., Gal, P. N., and Menon, C. (2014). The dynamics of employment growth: New evidence from 18 countries. OECD Science, Technology and Industry Policy Papers 14, OECD Publishing.
- Criscuolo, C., Gal, P. N., and Menon, C. (2015). DynEmp: a routine for distributed microdata analysis of business dynamics. *Stata Journal*, 15(1):247–274.
- Desnoyers-James, I., Calligaris, S., and Calvino, F. (2019). DynEmp and MultiProd: Metadata. OECD Science, Technology and Industry Working Papers 2019/03, OECD Publishing.

- Gal, P. N. (2013). Measuring total factor productivity at the firm level using OECD-Orbis. OECD Economics Department Working Papers 1049, OECD Publishing.
- Ganapati, S. (2021). Growing Oligopolies, Prices, Output, and Productivity. *American Economic Journal: Microeconomics*, 13(3):309–327.
- Kalemli-Özcan, Ş., Sørensen, B., Villegas-Sanchez, C., Volosovych, V., and Yeşiltaş, S. (2019). How to construct nationally representative firm level data from the Orbis Global Database: New facts and aggregate implications. Working Paper 21558, National Bureau of Economic Research.
- OECD (2020). The effects of R&D tax incentives and their role in the innovation policy mix. OECD Science, Technology and Industry Policy Papers 92, OECD Publishing.