American Economic Journal: Macroeconomics 2024, 16(2): 1–22
https://doi.org/10.1257/mac.20220036

How to Construct Nationally Representative Firm-Level Data from the Orbis Global Database: New Facts on SMEs and Aggregate Implications for Industry Concentration†

By Şebnem Kalemli-Özcan, Bent E. Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yeşiltaş*

We construct nationally representative firm-level longitudinal data for European countries using financial statements from the Orbis database. We validate our data by comparing its coverage and firm size distribution to official statistics. We showcase two applications to show the importance of firm representativeness in understanding macroeconomic outcomes. First, we show that small and medium-sized firms account for a large share of aggregate economic activity. Second, we document that firm representativeness is important for calculating industry concentration trends over time as the share of economic activity accounted by top firms in an industry changes with the firm samples used. (*JEL* C81, D22, G30, L11, L25, L60)

Empirical research based on firm-level data is increasingly important in macrofinance. For this research, one needs firm-level datasets that combine firms' real activity, such as employment and production, with their financing, debt, and assets. Most datasets cover only real activity (national censuses) or financial activity (financial reporting by publicly listed companies). The Orbis global database, from Bureau van Dijk (BvD)—a Moody's Analytics company—is the largest cross-country firm-level database that combines both, encompassing firms' financial statements and their real activity in terms of sales, employment, and investment. This database includes public and private firms' balance sheets and income statements as well as detailed information on firms' locations, industries, and domestic and foreign owners and subsidiaries, which allows researchers to observe global real and financial interconnections between the firms. In spite of the extensive use of the Orbis database for research, firm-level data downloaded from this database are not nationally representative, either due to the way the data is delivered to researchers

* Kalemli-Özcan: University of Maryland, NBER, and CEPR (email: kalemli@umd.edu); Sørensen: University of Houston and CEPR (email: bent.sorensen@mail.uh.edu); Villegas-Sanchez: ESADE, Universitat Ramon Llull, and CEPR (email: carolina.villegas@esade.edu); Volosovych: Erasmus University Rotterdam (email: volosovych@ese.eur.nl); Yeşiltaş: Koç University (email: syesiltas@ku.edu.tr). Simon Gilchrist was coeditor for this article. We thank three anonymous referees and numerous colleagues and seminar participants that have greatly contributed to improving the final version of the paper.

† Go to https://doi.org/10.1257/mac.20220036 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

and/or due to shortcuts the researchers adopt for quick downloads of an immense database that covers millions of firms over time. We provide a guide for researchers on how to download and organize the data such that it ends up being nationally representative or comes close to being so.[1]

To show the importance of such cross-sectional and dynamic intertemporal representation, we show two applications using a large set of European countries. First, we use our data to investigate the importance of small and medium-sized firms (SMEs) in the economy-wide production and employment in each of the countries in our sample. We show that SMEs account for a large part of real economic activity. Notice that the official statistics show this fact, but only for a select set of private sectors and not consistently over time, whereas we confirm this fact for all sectors of a given economy. Second, we investigate industry concentration trends in Europe and document a declining trend. We show that for nonrepresentative samples, one may find increasing industry concentration over time. If a nonrepresentative sample is used, the firms identified as "top firms" in a given industry and the share of economic activity they account for over time are different from the case where a representative sample of firms is used.

Before turning to these applications, we validate our dataset. For the validation exercise in terms of firm representation, we need to show that our firms cover a large part of the aggregate economy and are representative in terms of the firm size distribution given by official sources. We focus on the manufacturing sector for our validation exercise because official aggregate data provided by the Structural Business Statistics (SBS) of Eurostat—which we use for comparison and are provided by each country's national statistical office—are only consistently publicly available for the manufacturing sector across European countries over time. The exercise consists of comparing the aggregated total output, firm size distribution, and presence of foreign companies in our database to official aggregate statistics provided in Eurostat SBS. To validate the foreign companies' aggregate output, we use official data collected by the OECD. The coverage varies by country and over time, but for the European countries used in the applications, our firms cover close to 60 percent of manufacturing output at the beginning of our sample and more than 70 percent at the end of our sample.[2]

For our first application, we take the firms in our data and group them into employment size bins. Then we aggregate the output of those firms and divide by the total output of all our firms. We do this for each of our countries. This exercise confirms that SMEs, defined as firms with 20–250 employees (consistent with the Eurostat definition), account for more than half of aggregate employment and gross output in almost all our countries. This finding will also be important for our second application, which is to show the importance of using representative firm-level data for understanding trends in industry concentration in Europe. Recent research has found that industry concentration, defined as the market shares of the top 4 or

[1] See the online Appendix, which was originally circulated under the title "How to Construct Nationally Representative Firm Level Data from the ORBIS Global Database," NBER Working Paper 21558.

[2] In the data section, we provide a fair account of the advantages and disadvantages of using Orbis as well as the differences in coverage across countries over time.

top 8 firms in a given two-digit industry, has increased in the United States (see Grullon, Larkin, and Michaely 2019 and Crouzet and Eberly 2018)[3] but declined in Europe (Gutiérrez and Philippon 2017; Döttling, Gutiérrez, and Philippon 2017; Covarrubias, Gutiérrez and Philippon 2020).[4] Bajgar et al. (2019) challenge the findings of this second set of researchers and argue that industry concentration in Europe has increased once the role of business groups is taken into account by considering consolidated company statements. We find a declining concentration trend in Europe as in Gutiérrez and Philippon (2017) and explain why there are conflicting results in the literature.[5]

In a nonrepresentative sample of firms, the degree of industry concentration has increased only for the subset of firms that solely file consolidated firm statements, as also shown by Bajgar et al. (2019).[6] However, when a representative sample is used, there is a declining industry concentration trend even if one uses firms with consolidated statements; hence, the financial account type does not matter for results in a representative sample of firms. Why is this the case? We dig deeper and show that the increase in the filing of consolidated statements is due to the increased presence of foreign-owned firms and changing regulation, which is a natural part of the European integration process since the 2000s. Although foreign firms do not account for a large fraction of aggregate output in these countries (at most 30 percent overall),[7] increasing integration with the largest foreign-owned firms can drive trends in industry concentration if only consolidated statements are considered, because multinational companies are required to file those types of statements. In fact, we show that the increasing concentration trend among firms reporting consolidated statements is purely driven by foreign-owned firms.

The standard measure of concentration is the share of gross output by the top 8 firms in a country-sector-year over total gross output in the same cell. We find that when selecting the top 8 firms (ranked by firm output), different concentration trends appear if we focus on firms reporting unconsolidated accounts versus firms reporting consolidated accounts. Industry concentration decreases for firms

[3] Grullon, Larkin, and Michaely (2019) show that over the last two decades, over 75 percent of US industries experienced an increase in concentration levels with significant heterogeneity across sectors. Crouzet and Eberly (2018) find that the retail sector accounts for a large share of the increase in the observed aggregate business concentration. Increasing industry concentration trends were documented together with increasing profit margins and firm markups (see De Loecker, Eeckhout, and Unger 2020). Rossi-Hansberg, Sarte, and Trachter (2021) show that the positive trend observed in US concentration becomes negative when focusing on measures of local concentration. Amiti and Heise (2021) find that higher import competition caused a decline in the market shares of the top 20 US firms.

[4] The increase in average firm markups in Europe has also been limited (as shown by Díez, Fan, and Villegas-Sánchez 2021), which is attributed to better antitrust regulation by Covarrubias, Gutiérrez, and Philippon (2020). Besley, Fontana, and Limodio (2021) find that average firm profit margins are higher and concentration is lower in nontradable sectors when the antitrust measures are stronger. See Van Reenen (2018) for a review.

[5] Notice that Gutiérrez and Philippon (2017); Döttling, Gutiérrez, and Philippon (2017); and Covarrubias, Gutiérrez and Philippon (2020) find decreasing industry concentration trends in Europe using different data sources, namely KLEMS, ECB-Compnet, and Orbis.

[6] Consolidated financial statements are financial statements of an entity with multiple divisions or subsidiaries that can be located in the same country or in different countries.

[7] The average 30 percent share is obtained when we consider all countries from our sample that are also covered by the official OECD statistics on the multinational activity. But there is extensive variation across countries. For example, in Luxembourg, foreign companies account for more than 50 percent of aggregate output, while the same ratio in Germany or Italy is around 20 percent, with the United Kingdom and Spain taking about the average values.

reporting unconsolidated statements, while it is increasing for those reporting consolidated statements. Most importantly, when combining firms reporting either consolidated or unconsolidated financial statements, the overall message is that industry concentration has declined in Europe since 2001. A priori, there is no reason for focusing on a certain set of accounting statements as opposed to combining all statements as long as one is careful about not double counting the same firm reporting both statements. Focusing on a selected set of statements will lead to focusing on a selected set of firms such as listed firms, business groups, and foreign firms, and will give misleading trends in concentration. This practice of selecting certain groups will also deliver biased results due to changing regulation. For example, we show a sharp increase in concentration around 2007, which coincides with a change in the European accounting legislation.[8]

When we use representative samples of firms for our European countries, regardless of the use of different types of financial statements, we find a declining industry concentration trend. This is robust to different denominators that measure total output (i.e., total sample of Orbis firms, the sample of top 100 firms in Orbis and official aggregate sector data from STAN-OECD). The key in getting robust concentration trends is the correct selection of the "top firms." There is a nonnegligible number of large firms that can account for large shares of activity in a given industry and report unconsolidated accounts. These firms would be left out of studies focusing on the evolution of concentration measures by business groups by construction, as they are not part of such groups but they are nevertheless large. For policies such as antitrust regulation, it is essential to know about these firms. Of the total number of country-sector-year triplets, in 52 percent of these triplets, the top 8 firms report a mixture of consolidated and unconsolidated financial statements (i.e., within the top 8 firms, some firms report consolidated sales, while other firms report unconsolidated sales), enough to change who is in the top 8 group.[9]

The remainder of this paper is organized as follows. Section I describes the dataset and the construction of representative samples, with details reported in an online Appendix. Section II explains our data validation exercise. Section III presents our application to SMEs. Section IV presents our application to industry concentration. Section V concludes.

I. Data

The Orbis database provides harmonized financial and ownership information at the firm level across countries. The Orbis database covers more than 100 countries and over 400 million firms as of January 2022. BvD collects data from over 160 different

[8] To promote convergence of accounting standards at the global level, the European Commission introduced new accounting rules on the requirements to report consolidated and unconsolidated accounts in line with international standards adopted by International Financial Reporting Standards.

[9] To account for the concern that we are mixing consolidated information from the headquarters and unconsolidated information from subsidiaries and, therefore, double counting output, we use information on global ultimate owner, domestic ultimate owner, and immediate shareholder and checked whether the top 8 firms within the same country-industry-year triplet shared an owner. We find a negligible number of triplets in which the top 8 firms share an owner, and therefore we are confident that we are not double counting sales in the numerator of the concentration measure.

government and commercial information providers and harmonizes the data in a standard "global" format to facilitate worldwide company comparisons. The Orbis database includes both private and publicly listed firms, and it is not a census survey. The financial and balance sheet information comes from national business registers, governed by country-specific legal and administrative filing requirements. Although most countries oblige limited liability companies to register once they are formed, requirements in terms of firm size for reporting balance sheet items varies across countries.[10]

There is a common misconception that firm-level financial data from national statistical offices always have better coverage than Orbis. For countries where the laws require every firm to file to the national business register, this is not the case, because BvD uses the same sources as the statistical offices. For countries where the law requires only large firms to file financial statements, it might be the case that national statistical offices run different administrative surveys with better coverage of small firms.

For most European countries, reporting to the national business registers is mandatory; however, it might still be the case that researchers will not get full coverage from their Orbis download for a given country if they use a single vintage of the database or a direct download from the Orbis web platform.[11] For example, Bajgar et al. (2020), using a single download from the 2017 vintage, find that Orbis is tilted toward larger, older, and more productive firms, even within each size class. In order to have consistent coverage of small and large firms over time and by industry, one has to follow the approach we advocate and use the historical vintages, download data vintage by vintage, and match the firm data over time using unique firm identifiers. If a single vintage is used, firms will be missing in a longitudinal sense because Orbis drops nonreporting firms from the database after a certain period of time. In addition, some variables such as "value added" and "intermediate inputs" may be missing from some downloads, such as those from the commonly used Wharton Research Data Services (WRDS) (see WRDS 2010, 2023), which does not systematically cover all variables.[12] Single-vintage data will often overrepresent larger firms and underrepresent smaller firms due to survivorship bias.

Some researchers opt for reweighing the data, using weights based on official *aggregate* statistics in order to increase the representativeness of small firms.[13] If

[10] See Table A.6.1 in online Appendix A.6 for a list of the BvD information providers as well as for the details on filing requirements by country.

[11] Many researchers have found that the Orbis web browser interface displays a large number of unique firm identifiers, but the actual financial or real variables, when downloaded, turn out to be missing, especially going back in time. There are several reasons for this. First, there is a reporting lag in the BvD products of roughly two years, meaning that a firm's filing in 2017 will appear fully on the media issued/accessed in 2019. Second, depending on the BvD product, certain companies are erased from the database if there is no reporting done for some time, even if the firm continues operating (but not reporting). Third, there is a download cap imposed by BvD on web interfaces, and this cap often causes missing data rather than termination of the download job. Fourth, BvD collection efforts have improved over time. In addition to these technical considerations, certain cleaning and checking procedures have to be implemented.

[12] We have updated the comparison of three different data access methods: WRDS, Orbis Online, and Orbis Historical. A detailed account is provided in online Appendix Section A.1.1, which shows that Orbis Online is the most problematic due to firm attrition and download limitations. Meanwhile, WRDS provides information for the most recent eight years without download limitations, but still experiences firm attrition in the early years.

[13] For example, the OECD has a large project (MultiProd) to construct nationally representative samples working directly with national statistical offices that have access to business registries (see Berlingieri et al. 2017). However, in several countries, the statistical office only has access to production surveys and in those cases, a reweighting scheme is applied.

our guidelines are followed, there is no need to reweigh the data to obtain nationally representative firm-level datasets.[14] In addition, to produce the correct propensity score weights, the nonrepresentative data need to be matched first to a fully representative data at the *firm level*, such as census; otherwise, reweighing with ad hoc country-sector weights will introduce more errors, as argued by Haltiwanger et al. (2017). Based on our guidelines, BvD has recently developed a new product, labeled the "Historical Product," which links several vintages/disks of the Orbis data through firm identifiers to obtain firm-level longitudinal datasets for many countries, as we have done "manually." Although this new product avoids many of the issues involved by combining numerous vintages/disks, it requires a certain methodology to process the historical data. We provide the guide and programs to process this historical data at http://econweb.umd.edu/~kalemli/orbis.html. It is reassuring that the coverage ratios from our "manual procedure" and those from Orbis Historical Product are similar (Diez, Fan, and Villegas-Sanchez 2021 and Gourinchas et al. 2020 use Orbis Historical, follow our cleaning procedure, and report coverage ratios for financial variables and entry rates).

II. Validation

In order to validate the representativeness of our dataset, we proceed in two steps. In the first step, we measure output-based coverage ratios for the manufacturing sector (see Eurostat 1995–2014). We.[15] We proxy output by firm's operating revenue. We compute the ratio of the value of total output produced by firms in our sample relative to the value of total output from the official Eurostat-SBS data for the manufacturing sector. We do this exercise for the European countries that we use in our applications.

Table 1 shows how much of gross output, as reported by Eurostat-SBS data, is covered by the firms operating in the manufacturing industry in each of the selected European countries over this period. Some cells will be missing due to missing Eurostat-SBS data. With the exception of Finland, most countries show close to or above 60 percent coverage ratios, especially since 2001. Table 2 shows the overall European coverage for the manufacturing sector.

The second step of our validation exercise is to show that our firms are not only covering a large part of aggregate economy, but they are also representative; that is, our data can mimic the official firm-size distribution provided by Eurostat-SBS data. Table 3 presents the size distribution based on both gross output and employment in the manufacturing sector. We focus again on the manufacturing sector because official statistics do not provide firm size distribution consistently over time for all the sectors for all our countries. As an example year, we picked 2006, as shown

[14] We describe the methodology for preparing micro data based on Orbis in two self-contained appendixes. Online Appendix A deals with treatment of firm financial information, while online Appendix B deals the foreign ownership information. In particular, online Appendixes A.1 to A.4 describe the organization of Orbis and our advice on data download strategies. Online Appendix A.5 describes how to put together the financial data in panel form, while online Appendix B.1 explains the methodology used to put together foreign ownership data in panel form.

[15] Online Appendix C.1 provides details on the official aggregate datasets we use for comparison purposes.

Table 1—Coverage of the Manufacturing Sector Based on Gross Output

Year	AT	BE	CZ	DE	EE	ES	FI	FR	GB	GR
2001	0.47	0.78	0.73	0.50	0.92	0.78	0.36	0.79	0.60	N/A
2002	0.61	N/A	0.70	0.51	0.93	0.80	0.37	0.82	0.62	N/A
2003	0.59	0.81	0.77	0.57	0.93	0.79	0.39	0.79	0.66	0.92
2004	0.47	0.80	0.84	0.65	0.97	0.79	0.41	0.83	0.62	0.73
2005	0.45	0.80	0.86	0.90	0.95	0.78	0.41	0.82	0.70	0.69
2006	0.67	0.78	0.77	0.73	0.94	0.83	0.40	0.84	0.71	0.66
2007	0.72	0.79	0.81	0.77	0.96	0.81	0.45	0.87	0.69	0.68
2008	0.77	0.78	0.74	0.64	0.95	0.85	0.49	0.90	0.67	0.64
2009	0.72	0.74	0.77	0.60	0.92	0.87	0.46	0.89	0.81	0.51
2010	0.78	0.78	0.81	0.60	0.92	0.90	0.47	0.92	0.84	0.47
2011	0.80	0.77	0.79	0.57	0.95	0.85	0.50	0.96	0.89	0.45
2012	0.76	0.77	0.75	0.48	0.96	0.83	0.51	0.95	0.83	0.40
Average	0.65	0.78	0.78	0.63	0.94	0.82	0.44	0.87	0.72	0.62
Year	HU	IT	LV	NO	PL	PT	RO	SE	SI	SK
2001	0.88	0.65	0.51	N/A	0.57	0.63	0.59	0.73	0.77	0.54
2002	0.97	0.71	0.54	0.68	0.57	0.71	0.71	0.74	0.88	0.62
2003	0.84	0.70	0.60	0.68	0.59	0.73	0.83	0.71	0.92	0.83
2004	0.89	0.73	0.73	0.72	0.71	0.75	0.96	0.74	0.89	0.88
2005	0.88	0.77	0.79	0.69	0.67	0.91	0.95	0.75	0.91	0.78
2006	0.90	0.79	0.80	0.75	0.70	0.93	0.77	0.76	0.91	0.78
2007	0.88	0.79	0.80	0.76	0.69	0.92	0.89	0.78	0.91	0.76
2008	0.81	0.90	0.85	0.69	0.60	0.94	0.90	0.74	0.92	0.99
2009	0.90	0.86	0.80	0.88	0.74	0.92	0.97	0.87	0.91	0.95
2010	0.84	0.87	0.83	0.82	0.69	0.93	0.93	0.84	0.99	0.94
2011	0.77	0.89	0.87	0.81	0.60	0.93	0.91	0.81	0.83	0.98
2012	0.86	0.86	0.83	0.81	0.62	0.92	0.93	0.71	0.97	0.91
Average	0.87	0.79	0.75	0.75	0.65	0.85	0.86	0.77	0.90	0.83

Notes: The country codes are AT (Austria), BE (Belgium), CZ (Czech Republic), DE (Germany), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HU (Hungary), IT (Italy), LV (Latvia), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Each country-year cell represents the sum of manufacturing firms' gross output reported in Orbis data as a share of total official manufacturing gross output reported in Eurostat-SBS data.

in this table. Both panels show that the firm size distribution in our data is very close to the official one provided by Eurostat-SBS data, in terms of both output and employment. The table shows that most of the gross output and employment in the manufacturing sector is accounted for by SMEs, both in our data and in the official data. Some exceptions are Finland, the United Kingdom, and Slovakia with an underrepresentation of large firms; Greece with an overrepresentation of medium firms; and Italy and Slovenia with a slight underrepresentation of small firms.

III. Application: New Facts on SMEs Based on Orbis

Having validated our data for the manufacturing sector, we provide new facts on the size distribution of firms in all sectors, based on Orbis information. Official Eurostat-SBS statistics do not include all sectors of economic activity across countries. Column 1 in Appendix Table A.1 reports the relative importance, in output terms, of the sectors not included in Eurostat-SBS data as a percentage of the total based on Orbis information. The percentages vary from country to country but are not negligible. In most countries, official statistics are missing for sectors representing close to or more than 20 percent of total economic output. Similarly, column 2

Table 2—European Coverage of the Manufacturing Sector Based on Gross Output

Year	EU-unweighted (1)	EU-weighted (2)	EU-wide (3)
2001	0.65	0.64	0.65
2002	0.69	0.66	0.70
2003	0.73	0.69	0.68
2004	0.76	0.71	0.71
2005	0.77	0.78	0.79
2006	0.77	0.76	0.76
2007	0.79	0.77	0.78
2008	0.79	0.76	0.76
2009	0.81	0.78	0.76
2010	0.81	0.79	0.77
2011	0.80	0.79	0.76
2012	0.78	0.75	0.72

Notes: The columns of this table present EU averages of output-based coverage ratios that are constructed for each country-year cell presented in Table 1. EU averages are constructed following three alternative ways: column 1 presents simple EU-unweighted means, column 2 shows weighted averages where the corresponding country GDP acts as weight, and column 3 sums the gross output across all the firms included in the sample and compares to the corresponding gross output totals across the EU countries, assuming no borders between countries.

Table 3—Firm Size Distribution in the Manufacturing Sector: 2006

	AT	BE	CZ	DE	EE	ES	FI	FR	GB	GR	HU	IT	LV	NO	PL	PT	RO	SE	SI	SK
Panel A. Based on gross output																				
Orbis-Amadeus																				
1 to 19 employees	0.03	0.05	0.04	0.06	0.15	0.13	0.08	0.05	0.03	0.14	0.01	0.12	0.14	0.11	0.02	0.13	0.09	0.09	0.07	0.09
20 to 249 employees	0.23	0.30	0.32	0.23	0.67	0.4	0.38	0.23	0.23	0.55	0.14	0.49	0.62	0.4	0.32	0.44	0.32	0.26	0.32	0.33
250+ employees	0.74	0.66	0.64	0.70	0.18	0.47	0.54	0.72	0.74	0.31	0.85	0.40	0.25	0.49	0.67	0.43	0.59	0.65	0.61	0.58
Eurostat-SBS																				
0 to 19 employees	0.07	0.08	0.10	0.06	0.12	0.13	0.06	0.09	0.08	0.26	0.07	0.20	0.12	0.13	0.09	0.14	0.08	0.09	0.12	0.05
20 to 249 employees	0.32	0.27	0.31	0.22	0.60	0.38	0.21	0.27	0.29	0.26	0.21	0.41	0.54	0.36	0.28	0.42	0.31	0.28	0.32	0.23
250+ employees	0.60	0.65	0.59	0.72	0.28	0.49	0.73	0.63	0.63	0.48	0.72	0.38	0.33	0.51	0.62	0.43	0.60	0.63	0.56	0.72
Panel B. Based on employment																				
Orbis-Amadeus																				
1 to 19 employees	0.08	0.13	0.05	0.05	0.15	0.25	0.16	0.1	0.01	0.13	0.02	0.13	0.19	0.18	0.03	0.26	0.12	0.17	0.08	0.06
20 to 249 employees	0.36	0.41	0.41	0.32	0.58	0.49	0.43	0.34	0.35	0.57	0.26	0.55	0.56	0.47	0.40	0.54	0.38	0.34	0.38	0.33
250+ employees	0.56	0.45	0.53	0.63	0.27	0.26	0.41	0.56	0.64	0.30	0.72	0.32	0.25	0.35	0.57	0.20	0.50	0.48	0.55	0.61
Eurostat-SBS																				
0 to 19 employees	0.17	0.18	0.19	0.15	0.17	0.31	0.14	0.19	0.19	0.50	0.20	0.40	0.18	0.20	0.21	0.32	0.12	0.18	0.18	0.10
20 to 249 employees	0.38	0.37	0.37	0.32	0.53	0.43	0.34	0.34	0.38	0.28	0.36	0.38	0.52	0.42	0.38	0.49	0.40	0.33	0.36	0.34
250+ employees	0.46	0.45	0.44	0.53	0.30	0.26	0.51	0.47	0.43	0.22	0.44	0.22	0.30	0.38	0.41	0.19	0.48	0.49	0.45	0.55

Notes: The country codes are AT (Austria), BE (Belgium), CZ (Czech Republic), DE (Germany), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HU (Hungary), IT (Italy), LV (Latvia), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Panel A reports the share of gross output accounted for by each corresponding size bin in Orbis-Amadeus and Eurostat-SBS databases, respectively. Panel B reports the share of employment accounted for by each corresponding size bin in Orbis-Amadeus and Eurostat-SBS databases, respectively. Due to rounding, the totals might not add to 1.

shows that while Eurostat-SBS data reports information on the breakdown of employment and output by country-sector-size, there are many cells that lack this information. Again based on our sample in Orbis, we report that, for example, in Spain, Eurostat-SBS data provide the breakdown by sector and size for firms representing 41 percent of total output, leaving out 59 percent of total output. The lack

Table 4—Firm Size Distribution in the Aggregate Economy: 2006

	AT	BE	CZ	DE	EE	ES	FI	FR	GB	GR	HU	IT	LV	NO	PL	PT	RO	SE	SI	SK
Panel A. Based on gross output																				
1 to 19 employees	0.16	0.14	0.15	0.15	0.35	0.24	0.16	0.15	0.07	0.18	0.06	0.18	0.35	0.26	0.06	0.31	0.24	0.23	0.17	0.27
20 to 249 employees	0.32	0.36	0.37	0.24	0.47	0.37	0.29	0.32	0.26	0.43	0.29	0.40	0.44	0.33	0.36	0.37	0.35	0.31	0.38	0.32
250+ employees	0.52	0.50	0.48	0.61	0.18	0.40	0.54	0.53	0.67	0.39	0.65	0.42	0.22	0.41	0.58	0.32	0.41	0.46	0.45	0.42
Panel B. Based on employment																				
1 to 19 employees	0.17	0.24	0.13	0.08	0.31	0.28	0.23	0.20	0.03	0.15	0.04	0.15	0.32	0.30	0.06	0.39	0.29	0.30	0.13	0.15
20 to 249 employees	0.33	0.36	0.41	0.20	0.45	0.37	0.33	0.34	0.25	0.42	0.27	0.40	0.42	0.37	0.33	0.36	0.34	0.31	0.37	0.35
250+ employees	0.50	0.40	0.46	0.73	0.23	0.34	0.44	0.46	0.73	0.43	0.70	0.45	0.27	0.32	0.61	0.26	0.37	0.39	0.50	0.50

Notes: The country codes are AT (Austria), BE (Belgium), CZ (Czech Republic), DE (Germany), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HU (Hungary), IT (Italy), LV (Latvia), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Panel A reports the share of gross output accounted for by each corresponding size bin in Orbis-Amadeus. Panel B reports the share of employment accounted for by each corresponding size bin in Orbis-Amadeus database. Due to rounding, the totals might not add to 1.

of consistent information across countries, sector, and size bins precludes the direct comparison of the Orbis database size distribution to that in Eurostat-SBS data, and therefore we provide information based solely on the Orbis-Amadeus database.

Table 4 presents the size distribution based on gross output and employment in the aggregate economy based on the information provided in our Orbis sample of firms in 2006. Panels A and B show that most of the gross output and employment are accounted for by SMEs in the entire economy, mimicking the same fact we show for the manufacturing sector. Notice that this is a new fact shown by our data because, as explained, official statistics on the firm size distribution are not available for all sectors and countries.

IV. Application: Trends in Industry Concentration

Our second application regards the importance of firm representativeness in measuring the industry concentration trends in Europe.[16] To illustrate the importance of using a representative dataset, we report industry concentration measures using different sets of firms. We focus on the importance of the account type (i.e., consolidated versus unconsolidated financial accounts) and firm type (domestic or foreign owned). We use the standard measure of industry concentration, which is the market share of the top 8 (henceforth, MS8) firms in a given two-digit industry $s2$, country k, and year t relative to the population of all firms in the $s2, k, t$ triplet. We also use market shares of the top 4 (henceforth, MS4) firms or Herfindahl industry concentration measures, obtaining similar results (see Figure A.4 in the Appendix).[17]

[16] We select the sample of countries that cover at least 50 percent of the output reported by official statistics for aggregate economy in all years over the period 2001–2012, namely Austria, Belgium, the Czech Republic, Germany, Estonia, Spain, Finland, France, the United Kingdom, Greece, Hungary, Italy, Latvia, Norway, Poland, Portugal, Romania, Sweden, Slovenia, and Slovakia. See Table D.1.1 in the online Appendix for output-based coverage ratios for aggregate economy.

[17] We follow Nomenclature of Economic Activities (NACE) Level 1 and Level 2 classifications to aggregate four-digit industries to one- and two-digit industry–level data. See online Appendix Table A.6.2 for the NACE Revision 2, Level 2 classification. Orbis assigns the company to a unique "primary" industry by the largest portion of its operating revenue; some companies might have multiple "secondary" codes (describing their additional activities). We use the primary code as the "sector" of a given company.

We start with three different samples to calculate market shares of top 8 firms for the period 2001–2012 to illustrate the importance of firm representativeness: the sample of firms using all types of financial accounts, the sample of firms reporting unconsolidated accounts, and the sample of firms reporting consolidated accounts. If a firm reports both types of accounts at the same time, we clean this double counting before calculating the market share measures. In each sample, we "designate" top 8 firms by looking at the largest eight firms based on operating revenue. In addition, we also calculate aggregate concentration measures for Europe, both an "EU-wide" measure and an "EU-country-weighted" measure. For the first measure, we use the market shares of the largest eight firms in the pool of EU countries in a given sector-year. For the second measure, we aggregate each country's own concentration measure based on market shares of the top 8 firms using as weights a given country's GDP.

Orbis classifies firms' balance sheet information according to four types of accounts corresponding to consolidation codes:[18]

- U1: only unconsolidated accounts are *available* in Orbis.
- C1: only consolidated accounts are *available* in Orbis.
- U2–C2: both unconsolidated and consolidated accounts are *available* in Orbis.[19]

Consolidated statements are financial statements of an entity with multiple divisions or subsidiaries. It is obvious that using only unconsolidated or only consolidated accounts will change the total sales/revenue of the top 8 firms, biasing the concentration measures. Using unconsolidated accounts only might underrepresent the true level of concentration if sales of the same consolidated group are recorded across different business units. This is especially important in the case of EU-wide concentration measures with foreign subsidiaries distributed across different EU countries. It is also misleading to use only consolidated accounts because many large companies do not report consolidated accounts, and hence this practice will bias the selection of top 8 firms. In fact, an overwhelming majority of companies in Orbis report unconsolidated accounts. An additional problem specific to the European setting is that regulatory changes after 2007

[18] In addition, Orbis contains companies with the account type "LF" with limited financial information and "NF" with no financial items at all. Also, there are entities with no recent accounts ("NRF") or no recent limited financials ("NRLF"), where "no recent" refers to last three years. By default, the Orbis media gives preference to the consolidated accounts, which can be changed via the search settings. See Figure D.4.1 of online Appendix D.4 for an example of how Orbis registers the multiple accounts of different types of Koç Holding Inc., the largest business group in Turkey.

[19] See Appendix Section A for full details on how we deal with different types of duplicates and double counting. All firms reporting under different consolidation codes share the same BvD ID number and differ on their last letter code reflecting C1, U1, C2, or U2, respectively, depending on the type of reporting consolidation code. To avoid double counting of sales, we eliminate duplicates based on BvD ID keeping the consolidated accounts when both consolidated and unconsolidated are reported (i.e., we drop the unconsolidated sales of headquarters). In spite of this approach, we cannot rule out the possibility that the sales of affiliates are double counted (i.e., included in the consolidated sales of the headquarters and separately, as unconsolidated sales of the affiliate). We deal with this concern by using the ownership information.

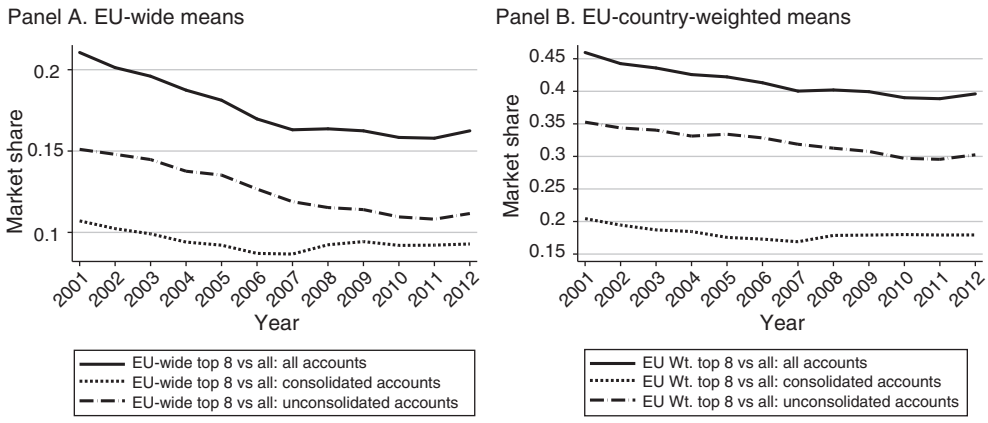

FIGURE 1. MARKET SHARE OF TOP 8 FIRMS

Notes: The figure plots market concentration in the total economy for the European countries in our sample from 2001 to 2012. Panel A plots "EU-wide" market shares over the period: we treat the European Union as a single market and define market shares in all the EU countries in a given sector-year. We aggregate over sectors using sectoral value added as weights. In Panel B, we plot "EU-country-weighted" average market shares, using a given country's GDP as weight. In this panel we calculate each country's concentration measure separately, adding up sectors with sector value added and then add up countries with GDP weights.

push company reporting to consolidated accounts.[20] We checked our data and confirmed that cases that we detected to be switchers took place mostly between 2007 and 2009, when the regulatory change came into effect. We correct for this to make sure that we are not identifying switchers as different firms as explained in Appendix Section B.

Figure 1 shows our main results. Panel A in Figure 1 shows the EU-wide concentration measures, while panel B in Figure 1 shows the EU-country-weighted counterpart. Each figure shows concentration trends based on market shares of top 8 firms for the three samples we use as described above. Clearly, when one uses all firms and all accounts, the concentration is declining in Europe. However, when distinguishing between firms reporting unconsolidated accounts and those reporting consolidated accounts, we can observe opposing trends, especially after 2007 in both panels. When we use consolidated accounts, it seems like concentration declined during the period 2000–2007 and increased from 2007 onward, consistent with the results in Bajgar et al. (2019). The results are even more pronounced when we do not control the companies switching accounts due to the regulatory change, as shown in Figure A.1 in the Appendix.

[20] The Council of the European Union adopted an International Accounting Standards (IAS) Regulation that requires listed companies, including banks and insurance companies, to prepare their consolidated accounts in accordance with IAS from 2005 onward. Member states may defer application until 2007 for those companies that are listed both in the European Union and elsewhere and that currently use generally accepted accounting principles (GAAP) that are adopted by the United States (or use other GAAP) as their primary basis of accounting, as well as for companies that have only publicly traded debt securities. Further details on this regulation are available at https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/financial-reporting_en.

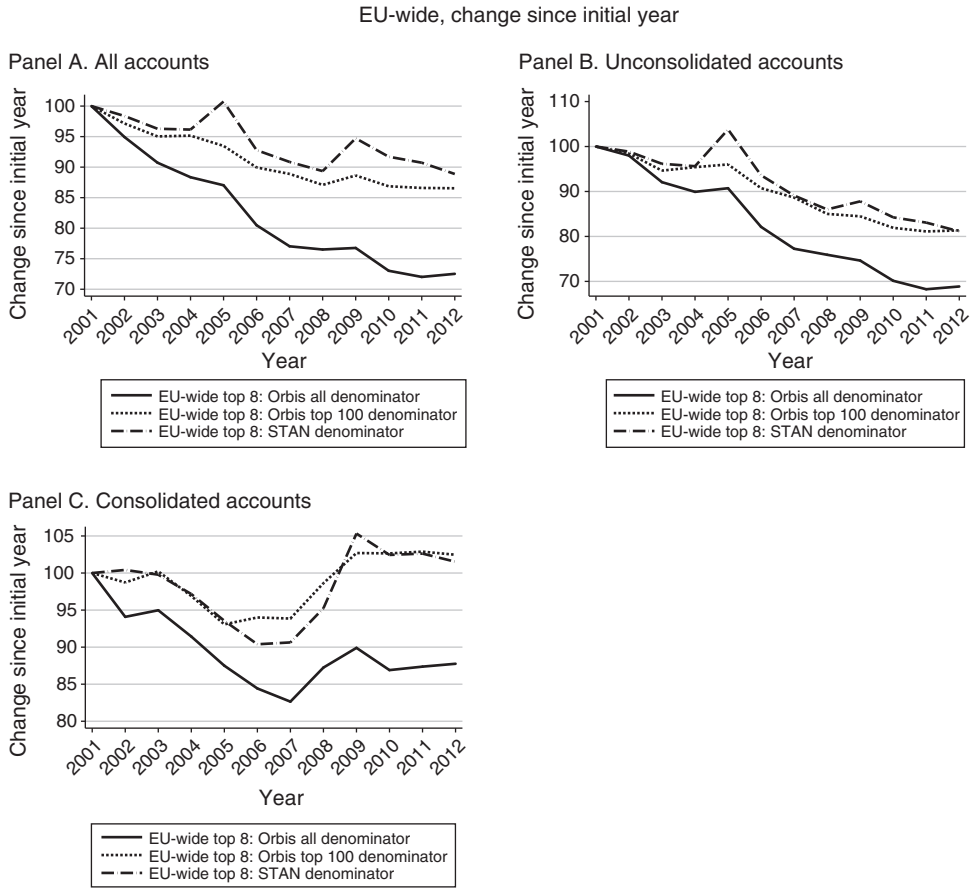

Figure 2. Market Share of Top 8 Firms: Different Denominators for "All" Firms

Notes: The figure plots market concentration in total economy for the European countries from 2001 to 2012. All panels plot "EU-wide" market shares as *percentage* change since the initial year (= 100) over this period: we treat the European Union as a single market and calculate market share of top 8 firms in all the EU countries in a given sector-year as their aggregated output relative to (i) the aggregated output of all firms (solid line); (ii) the aggregated output of top 100 firms in the pool of EU countries (short-dashed line), and (iii) the EU-wide gross output reported by the OECD STAN database for the same sector-year (long-dashed-dot line). We aggregate over sectors using sectoral value added as weights.

These concentration measures might be impacted by the fact that Orbis, although representative, does not cover the universe of firms in an economy. To show that this does not impact the declining concentration trend result, Figure 2 shows EU-wide concentration measures using three different denominators to calculate the market shares. Each denominator is a different proxy for total output of all the firms in an economy: Orbis-total, Orbis-100 (that is, the output from the top 100 firms in Orbis) and gross output aggregate reported in the OECD Structural Analysis (STAN) database (available at http://stats.oecd.org/Index.aspx?DataSetCode=STAN08BIS; see OECD 2000–2014b). To make it comparable to the results in Bajgar et al. (2019), we

report EU-wide means relative to a baseline year over the period 2001–2012.[21] Panel A shows results for the sample of all accounts, while panels B and C show results for unconsolidated and consolidated accounts, respectively. The declining trend during the entire sample period for the full sample as well as the unconsolidated accounts sample is clear in panels A and B. In fact, there is almost no difference between the normalization by Orbis-100 sales and STAN gross output data. In fact, we prefer not to use the STAN denominator as the sector classification given that the STAN A64 classification does not fully correspond to the two-digit industry classification in NACE Rev. 2.[22] Nevertheless, concentration in Europe declined by close to 10 percent when considering the full sample and by 20 percent when considering the sample of firms reporting unconsolidated accounts, regardless of whether STAN or Orbis information is used.[23]

Panel C shows a decreasing trend up to 2007 and an increasing trend thereafter, when we are to use the consolidated accounts sample. Concentration among firms reporting consolidated accounts increased by 2.5 percent between 2001 and 2012 in Europe. As we already argued, this is driven by two factors: (i) by not including unconsolidated accounts, important large firms will be missed in the top 8 group, and (ii) the regulatory change in 2007 pushed a certain set of firms to switch from reporting unconsolidated to consolidated. As we will show next, these firms who switch reporting, captured increasingly more so by the consolidated reporting, happen to be foreign firms, and they report consolidated as required by the regulation.

To show the importance of omitting large private firms that report unconsolidated accounts, we undertook an additional analysis. This analysis will also show that finding different concentration trends is not about Orbis data but about using "only" consolidated accounts. As shown in Figure 3, using consolidated accounts in Orbis or in Compustat Global (which consists only of publicly held companies; see WRDS 1995–2014) will also result in increasing concentration trends across European countries, especially in the post-2007 period due to changes in the regulation. Ali, Klasa, and Yeung (2008) argue that Compustat-based industry concentration measures are poor proxies for actual industry concentration. The correlation between the Compustat and US census–based Herfindahl indexes is only 13 percent.

In order to dig deeper and understand what drives the increasing concentration trends when using consolidated accounts, we separate foreign and domestic top firms.[24] The literature has shown that foreign/multinational companies are large and operate through a network of subsidiaries and affiliates in several countries (e.g., Kalemli-Ozcan, Sorensen, and Volosovych 2014).

[21] Figure A.2 in the Appendix shows the levels and the fact that the decreasing trend for unconsolidated accounts is robust to the denominator being top 50, top 100 or top 1,000 both for EU-wide and EU-country-weighted measures.

[22] For the exercise in Figure 2, we selected the overlapping sectors in both classifications.

[23] Bajgar et al. (2019) show increasing concentration trends when using data from their MultiProd project. This project uses business registry information when available and uses production survey data reweighted when official representative data are not available. Because we find decreasing concentration trends when using official output from STAN in the denominator, we conclude that the main difference is due to differences in the selection of the top firms that contribute to the numerator (different weighting schemes can influence the selection of top firms).

[24] Any firm whose equity is owned by foreigners in excess of 10 percent is defined as a foreign-owned firm, following the balance-of-payments definition of the IMF.

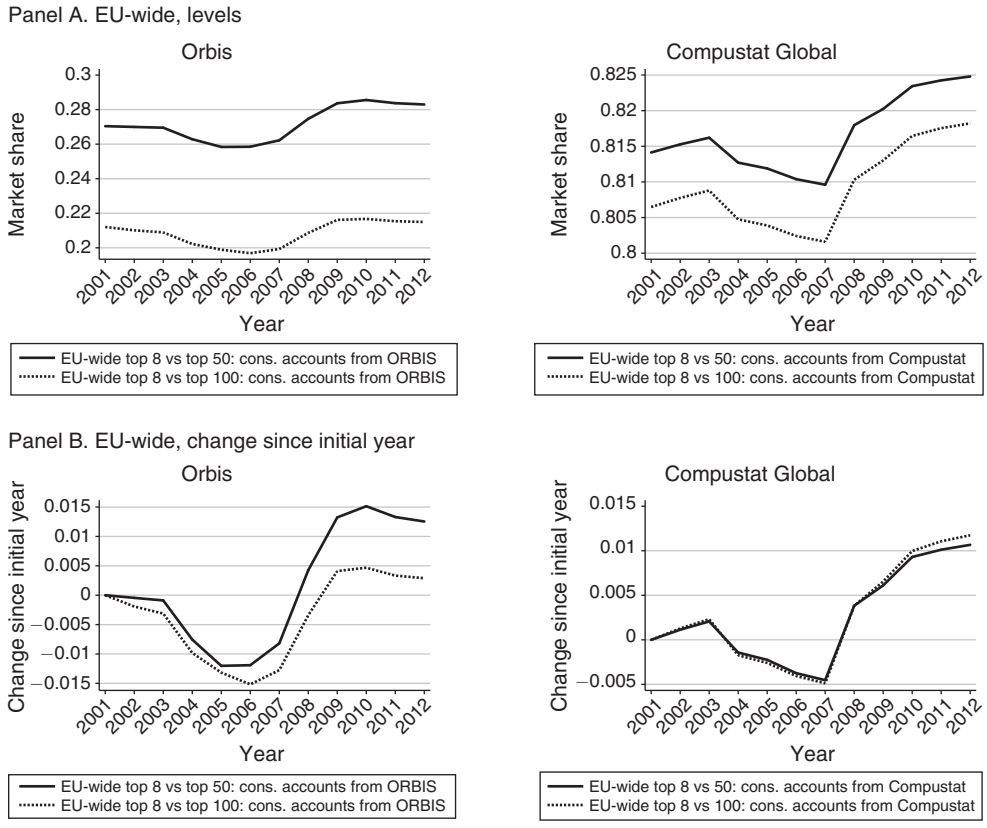

Figure 3. Market Share of Top 8 Firms: Orbis versus Compustat Global

Notes: The figure plots market concentration in total economy for the European countries in the sample of consolidated accounts (from our sample, shown in the top-left and bottom-left panels, and from Compustat-Global database, shown in the top-right and bottom-right panels) in the period 2001–2012. All panels plot "EU-wide" market shares over this period: we treat the European Union as a single market and define market share of top 8 firms in all the EU countries in a given sector-year as their aggregated output relative to (i) the aggregated output of top 50 firms (solid line) or (ii) the aggregated output of top 100 firms (short-dashed line). We aggregate over sectors using sectoral value added as weights.

First, to validate our data, we aggregate the activity of the foreign firms in our panel to the country-year level and compare these aggregated ownership numbers to the OECD data on the activities of foreign affiliates of multinationals from the Activity of Multinational Enterprises (AMNE) and the Activities of Foreign Affiliates (AFA) databases (see OECD 2000–2014a).[25] Figure B.3.1 of online

[25] AMNE (available on the OECD data portal: https://stats.oecd.org/Index.aspx?DataSetCode=AMNE_IN) covers 28 OECD host countries from 2008 onward, although the coverage varies by country and over time. We base our comparison on the manufacturing sector because the earlier OECD data, reported in the AFA database, consistently provides information only from manufacturing sector (see http://stats.oecd.org/Index.aspx?DataSetCode=AFA_IN3). The OECD Foreign Affiliates Statistics database provides detailed data on the activities of foreign affiliates operating in the *services* sector, albeit for a smaller sample of 25 OECD countries. OECD obtains their data from the Eurostat that conducts annual surveys on the activities of foreign-controlled enterprises and foreign affiliates abroad controlled by residents of the compiling country. Surveys are conducted in most cases by the national sta-

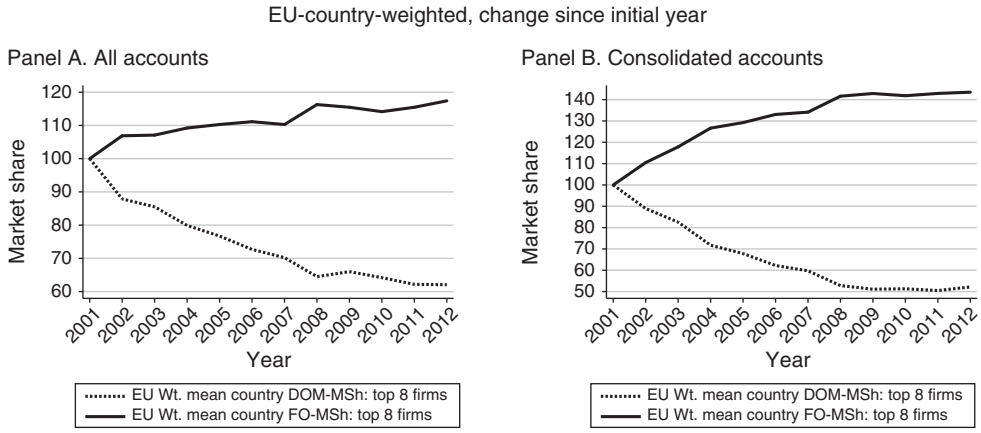

FIGURE 4. MARKET SHARE OF FOREIGN AND DOMESTIC FIRMS WITHIN TOP 8 FIRMS

Notes: The figure plots market concentration of foreign (FO) and domestic (DOM) firms within top 8 firms in the total economy for the European countries since 2001. Both panels plot "EU-country-weighted" market shares (MSh) as percentage change since the initial year (= 100) over this period: we calculate each country's concentration measure separately, adding up sectors with sector value added and then adding up countries with GDP weights. In both panels, being a foreign firm is determined based on official threshold of more than 10 percent foreign ownership in the initial year.

Appendix B.3 graphically presents this comparison. The figure demonstrates that our data capture very well the share of multinational activity in total activity reported by the official statistics.[26]

To compute the concentration measures across the two groups, foreign and domestic, we keep the denominator fixed (i.e., it includes all firms in our sample), and we change the numerator by distinguishing between the sales of the top 8 foreign firms and the top 8 domestic firms. Figure 4 shows the results. We report EU-country-weighted averages to have a clear interpretation of foreign firms. BMW-Germany is considered a domestic firm in Germany, while BMW-Spain is considered a foreign-owned company in Spain. In an EU-wide measure, they are both domestic. We show results first using all accounts and then using consolidated accounts, as changes from the initial year.[27] As mentioned, the two panels share the same denominator (sales of all firms in the sample).[28]

Both panels show that the increase in industry concentration is driven by foreign firms, regardless of the account type used. Hence, any nonrepresentative sample that gives larger weight to these foreign firms will register an increasing

tistical office or central bank of each country. While the key variables in the survey are common across countries, the target sample varies across countries. See online Appendix B.2 for a description of the issues considered to maximize comparability across samples.

[26] Online Appendix B.3 provides these statistics by country and other details of our validation exercise. Our advice on how to download and clean the foreign firms' data is described in detail in online Appendix B.1.

[27] We classify foreign-owned and domestically owned firms based on the ownership information on the first year of the sample—i.e., 2001—and follow their market shares over time. Similar results are found based on the time-varying ownership information.

[28] Appendix Figure A.3 shows that the results are robust to considering the output of the top 100 firms in the denominator.

concentration trend over time. Using consolidated accounts is just one example of such nonrepresentation because—especially in Europe with the regulatory changes in 2007—using consolidated accounts will end up giving a higher weight to foreign firms.

V. Conclusion

We construct nationally representative firm-level longitudinal data for European countries using financial statements from the Orbis database. We provide a detailed "guide" on the construction for other researchers, and we validate our data by comparing its coverage and firm size distribution to official statistics for the manufacturing sector.

To show the importance of firm representativeness, we showcase two applications, where representation is critical for understanding macro outcomes. First, we show that SMEs account for a large share of aggregate economic activity, regardless of the sector and country. This is a new fact because one can only obtain official statistics on firm size distribution for certain sectors. Hence, before our paper, we did not know whether the larger role of SMEs in aggregate economic activity was specific to certain sectors.[29] Second, we document that such firm-level representation is important for obtaining correct industry concentration trends over time in Europe. In a representative firm sample, one obtains a declining concentration trend. In a nonrepresentative sample—focusing on large firms, foreign firms, or firms reporting using certain financial accounting types—the industry concentration is rising. In our application, which is for Europe, we show that all these examples of nonrepresentation give more weight to large foreign firms, whose market shares clearly have gone up as part of the easing of cross-border regulations during the European integration process. These firms cannot represent economy-wide trends though, as we show that using nationally representative firm-level information industry concentration trends in Europe have declined.

Appendix

A. *Dealing with Duplicates*

BvD assigns three generic variables—i.e., "BvD Account Number," "BvD ID Number," and "Consolidation Code"—to a given firm available in the data. The Consolidation Code variable can take different values, corresponding to the type of financial statements reporting available to BvD. The codes reported by the variable Consolidation Code are as follows: C1 indicates that BvD has information on the firm's consolidated accounts only. U1 indicates BvD has information on the firm's

[29] Of course, one can obtain the firm size distribution of every sector by using the micro data from national census surveys for each country and redefining the size categories to match the definition of SMEs. The official aggregate statistics from Eurostat do not do this—that is, they do not tell us the role of firms with fewer than 250 employees in aggregate economic activity for every sector and/or for the entire country. Note that even the Small Business Administration in the United States with detailed data on SMEs only reports this statistic for the total of private sector in the United States.

unconsolidated accounts only. C2 indicates that BvD has information on the firm's both consolidated and unconsolidated accounts and those associated to C2 are the consolidated ones. U2 indicates BvD has information on the firm's both consolidated and unconsolidated accounts and those associated to U2 are the unconsolidated ones. LF indicates the firm reports limited financial information. The variable BvD Account Number is composed of three parts: the first two letters at the beginning of the string stand for the country code ("BE" for Belgium, "US" for the United States, "GB" for the United Kingdom, and so on), and the last character of the string refers to the type of consolidation code, based on values reported in the variable Consolidation Code. The remaining numeric part in the middle of the string and the first two letters at the beginning of the string (identifying the country code) constitute the variable "BvD ID Number."

In the dataset that we constructed using different vintages, we created our main company identifier ID_NUMBER, which is a copy of the BvD ID Number as well as our main account identifier CONSCODE2, which is a copy of the last letter of BvD Account Number. We fill CONSCODE2 with "C" or "U" according to the type of the firm's financial statements reporting available to BvD. Specifically, we fill CONSCODE2 with "C" if the financial information of the respective companies is available to BvD via C1 and/or C2 accounts. Similarly, we fill CONSCODE2 with "U" if the financial information of the respective companies is available to BvD via U1 and/or U2 accounts. Given that the original Consolidation Code is "LF" for the companies reporting limited financial information (for these companies, all financial variables except sales and total assets are missing), we extract the information on account type from the last letter of the variable "BvD Account Number" and fill CONSCODE2 of such companies with that information accordingly. These generic variables enable us to track *the same company* (the same ID_NUMBER) with multiple accounts of different types (U or C) in a given financial year. We tag those accounts as "duplicate accounts."

Duplicate accounts in Orbis data arise because companies may report both consolidated and unconsolidated statements in the same year, or firms may switch the type of financial statements they report over time, and when we combine information from different vintages, we may end up with different accounts coexisting in the same firm-year in the merged dataset. Such duplicate accounts appear in two cases:

Case 1: A company reports two accounts with the same ID_NUMBER and different consolidation codes and different values of financial and real variables for each consolidation code in the same year.

Case 2: A company reports two accounts with the same ID_NUMBER and different consolidation codes but the same values of financial and real variables for both consolidation codes.

To avoid duplicates (i.e., the same firm reporting under different consolidation codes) and have unique firm-year observations, we make the following choices. Among Case 1–type duplicates (i.e., firms reporting different financial values

under different consolidation codes), we give priority to those with consolidated accounts. To resolve Case 2–type duplicates (i.e., firms reporting the *same* financial values under different consolidation codes), we implemented the following filters:

(i) For the duplicates belonging to companies that continuously report unconsolidated accounts in the period they are available in BvD data, we give priority to those with unconsolidated accounts.

(ii) For the duplicates belonging to companies that continuously report consolidated accounts in the period they are available in BvD data, we give priority to those with consolidated accounts.

(iii) For the duplicates belonging to companies that report both consolidated and unconsolidated accounts with the *same* sales value, but not consistently over the period they are available in BvD data, we checked and verified that the volume of sales over time was consistent with the consolidation code. Therefore, we give priority to the consolidated code classification and reclassify the time series as consolidated.

B. *Dealing with Switchers*

After getting rid of duplicates, in terms of reporting both consolidated and unconsolidated accounts or in terms of reporting the same financial variables under different consolidation codes, we keep track of "switchers." These are firms that change the reporting account type over time, so the financial information we have combines consolidated and unconsolidated sales of the corresponding firm. Given the change in legislation around 2007 with more firms reporting consolidated statements following international financial standards, we decided to drop these firms to have a consistent time series and minimize the artificial increase in sales that might arise from the change in legislation around 2007. Notice the change in regulation toward the consolidation reporting in accordance with IAS was supposed to take effect from 2005 onward. However, member states could defer the application until 2007 for those companies that were listed both in the European Union and elsewhere and that were using the US GAAP (or other GAAP) as their primary basis of accounting, as well as for companies that had only publicly traded debt securities.

C. *Appendix Tables and Figures*

Table A.1—Fraction of "Missing" Output in Eurostat

Country	Missing sector (1)	Missing sector-size (2)
Austria (AT)	17	71
Belgium (BE)	18	66
Czech Republic (CZ)	7	55
Germany (DE)	18	68
Estonia (EE)	5	67
Spain (ES)	11	59
Finland (FI)	30	72
France (FR)	9	65
Great Britain (GB)	16	69
Greece (GR)	17	77
Hungary (HU)	9	61
Italy (IT)	13	56
Latvia (LV)	9	73
Norway (NO)	15	69
Poland (PL)	18	60
Portugal (PT)	9	68
Romania (RO)	6	62
Sweden (SE)	21	72
Slovenia (SI)	4	55
Slovakia (SK)	7	60

Notes: The table shows the importance of missing sectors in Eurostat-SBS database based on the information provided by Orbis-Amadeus database. Column 1 shows the share of unaccounted for output due to missing sectoral information in Eurostat-SBS database. Column 2 shows the share of unaccounted for output due to missing sector-size information in Eurostat-SBS database.

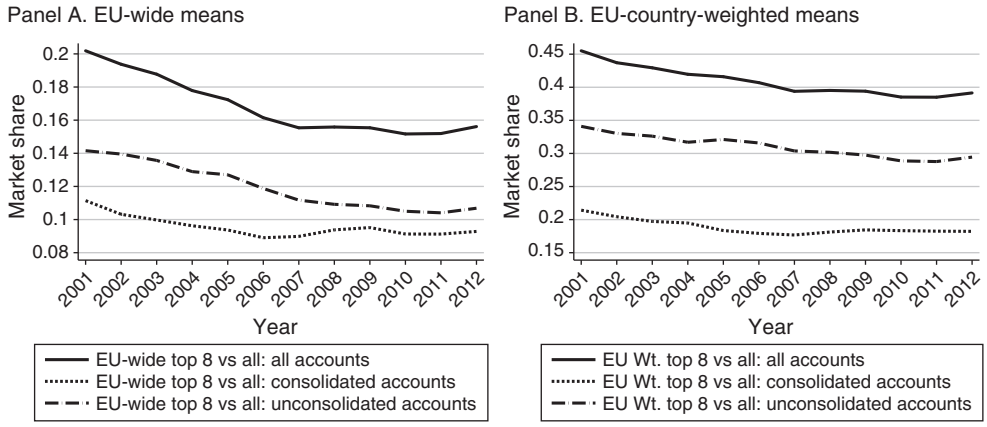

Figure A.1. Market Share of Top 8 Firms: Sample of All Firms Including Switchers

Notes: The figure plots market concentration in the total economy for the European countries in our sample from 2001 to 2012. Panel A plots "EU-wide" market shares over the period: we treat the European Union as a single market and define market shares in all the EU countries in a given sector-year. We aggregate over sectors using sectoral value added as weights. In panel B, we plot "EU-country-weighted" average market shares, using a given country's GDP as weight. In this panel, we calculate each country's concentration measure separately, adding up sectors with sector value added and then adding up countries with GDP weights.

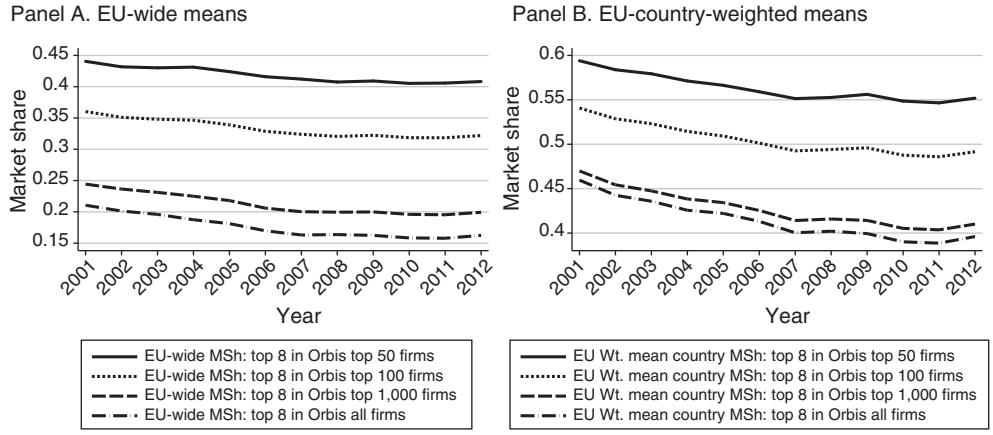

FIGURE A.2. MARKET SHARE OF TOP 8 FIRMS: DIFFERENT DENOMINATORS FOR TIME COVERAGE

Notes: The figure plots market concentration in total economy for the European countries in different subsamples of our sample from 2001 to 2012. Panel A plots "EU-wide" market shares over the period: we treat the European Union as a single market and define market shares in all the EU countries in a given sector-year. We aggregate over sectors using sectoral value added as weights. In panel B, we plot "EU-country-weighted" average market shares, using a given country's GDP as weight. In this panel, we calculate each country's concentration measure separately, adding up sectors with sector value added and then adding up countries with GDP weights. In both panels, we calculate market share of top 8 firms as their aggregated output relative to (i) the aggregated output of Orbis top 50 firms (solid line), (ii) the aggregated output of Orbis top 100 firms (short-dashed line), (iii) the aggregated output of Orbis top 1,000 firms (long-dashed line), and (iv) the aggregated output of Orbis all firms (long-dashed-dot line).

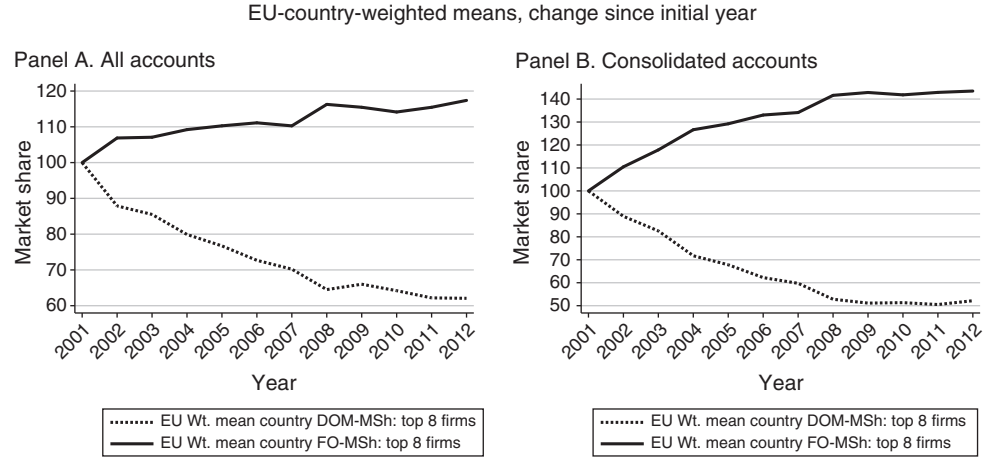

FIGURE A.3. MARKET SHARE OF FOREIGN AND DOMESTIC FIRMS WITHIN TOP 8 FIRMS, TOP 100 FIRMS IN DENOMINATOR

Notes: The figure plots market concentration of foreign and domestic firms within top 8 firms in the total economy for the European countries in different subsamples of top 100 firms in our sample since 2001. Both panels plot "EU-country-weighted" market shares as percentage change since the initial year ($=$ 100) over this period: we calculate each country's concentration measure separately, adding up sectors with sector value added and then adding up countries with GDP weights. In both panels, being a foreign firm is determined based on the official threshold of more than 10 percent foreign ownership in the initial year.

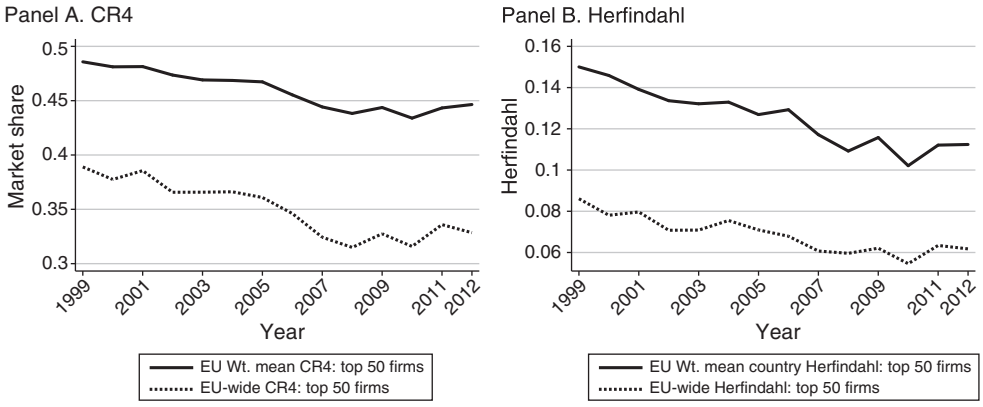

Figure A.4. Alternative Measures of Industry Concentration

Notes: This figure replicates Chart 9 and Chart 10 in Döttling, Gutiérrez, and Philippon (2017). We plot market concentration in the nonfinancial corporate sector for the EU KLEMS countries in a subsample of unconsolidated accounts of our sample from 1999 to 2012. The EU KLEMS countries are Austria (AT), Belgium (BE), Germany (DE), Spain (ES), Finland (FI), France (FR), Italy (IT), and the Netherlands (NL). We exclude utilities (KLEMS segments D35 to D39), financial firms (segments D64 to D66), real estate (segment D68), and industry segments D84 ("Public administration and defence; compulsory social security") and D99 ("Activities of extraterritorial organizations and bodies") from the nonfinancial corporate sector. We measure market concentration over this period by the market shares of top 4 firms (CR4) in top 50 firms, shown in panel A, and by Herfindahl index of top 50 firms, shown in panel B. In both panels, we plot market concentration both on an EU-wide level and on a country level: we plot "EU-wide" market shares over the period (short-dashed lines); we treat the EU KLEMS countries as a single market and define market shares in all the EU KLEMS countries in a given sector-year. We aggregate over sectors using sectoral value added as weights. We plot "EU-country-weighted" average market shares (solid lines): we calculate each country's concentration measure separately, adding up sectors with sector value added.

REFERENCES

Ali, Ashiq, Sandy Klasa, and Eric Yeung. 2008. "The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research." *Review of Financial Studies* 22 (10): 3839–71.

Amiti, Mary, and Sebastian Heise. 2021. "U.S. Market Concentration and Import Competition." Federal Reserve Bank of New York Staff Reports 968.

Bajgar, Matej, Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Jonathan Timmis. 2019. "Industry Concentration in Europe and North America." OECD Working Paper 18.

Bajgar, Matej, Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Jonathan Timmis. 2020. "Coverage and Representativeness of Orbis Data." Unpublished.

Berlingieri, Giuseppe, Patrick Blanchenay, Sara Calligaris, and Chiara Criscuolo. 2017. "The Multiprod Project: A Comprehensive Overview." OECD Science, Technology and Industry Working Paper 2017/04.

Besley, Timothy, Nicola Fontana, and Nicola Limodio. 2021. "Antitrust Policies and Profitability in Nontradable Sectors." *American Economic Review: Insights* 3 (2): 251–65.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon. 2020. "From Good to Bad Concentration? U.S. Industries over the Past 30 Years." *NBER Macroeconomics Annual* 34: 1–46.

Crouzet, Nicolas, and Janice Eberly. 2018. "Intangibles, Investment, and Efficiency." *AEA Papers and Proceedings 1*08: 426–31.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *Quarterly Journal of Economics* 135 (2): 561–644.

Díez, Federico J., Jiayue Fan, and Carolina Villegas-Sanchez. 2021. "Global Declining Competition?" *Journal of International Economics* 132: 103492.

Döttling, Robin, Germán Gutiérrez, and Thomas Philippon. 2017. "Is There an Investment Gap in Advanced Economies? If So, Why?" In *ECB Forum on Central Banking: Investment and Growth in Advanced Economies*, 129–93. Sintra, Portugal: European Central Bank.

Eurostat. 1995–2014. "Structural Business Statistics (SBS)." https://ec.europa.eu/eurostat/cache/metadata/en/sbs esms.htm.

Gourinchas, Pierre-Olivier, Sebnem Kalemli-Özcan, Veronika Penciakova, and Nick Sander. 2020. "Estimating SME Failures in Real Time: An Application to the COVID-19 Crisis." NBER Working Paper 27877.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2019. "Are U.S. Industries Becoming More Concentrated?" *Review of Finance* 23 (4): 697–743.

Gutiérrez, Germán, and Thomas Philippon. 2017. "Declining Competition and Investment in the U.S." Unpublished.

Haltiwanger, John, Ron S. Jarmin, Robert Kulick, and Javier Miranda. 2017. "High-Growth Young Firms." In *Measuring Entrepreneurial Businesses,* 11–62. Chicago: University of Chicago Press.

Kalemli-Ozcan, Şebnem, Bent Sørensen, and Vadym Volosovych. 2014. "Deep Financial Integration and Volatility." *Journal of the European Economic Association* 12 (6): 1558–85.

Kalemli-Özcan, Şebnem, Bent Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yeşiltaş. 2024. "Replication Data for: How to Construct Nationally Representative Firm-Level Data from the Orbis Global Database: New Facts on SMEs and Aggregate Implications for Industry Concentration." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E187541V1.

OECD. 2000–2014a. "AMNE and AFA Databases—Activity of Multinational Enterprises." https://www.oecd.org/sti/ind/amne.htm.

OECD. 2000–2014b. "STAN—STructural ANalysis Database." https://www.oecd.org/sti/ind/stanstructuralanalysisdatabase..

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2021. "Diverging Trends in National and Local Concentration." *NBER Macroeconomics Annual* 35: 115–50.

Van Reenen, John. 2018. "Increasing Differences between Firms: Market Power and the Macro-economy." Centre for Economic Performance, LSE CEP Discussion Paper 1576.

WRDS (Wharton Research Data Services). 1995–2014. "Compustat Global." https://wrds-www.wharton.upenn.edu/.

WRDS. 2010. "Bureau van Dijk (BvD)—Orbis." https://wrds-www.wharton.upenn.edu/.

WRDS. 2023. "Bureau van Dijk (BvD)—Orbis." https://wrds-www.wharton.upenn.edu/.