

NBER WORKING PAPER SERIES

HOW TO CONSTRUCT NATIONALLY REPRESENTATIVE FIRM LEVEL DATA FROM  
THE ORBIS GLOBAL DATABASE:  
NEW FACTS AND AGGREGATE IMPLICATIONS

Sebnem Kalemli-Ozcan  
Bent Sorensen  
Carolina Villegas-Sanchez  
Vadym Volosovych  
Sevcan Yesiltas

Working Paper 21558  
<http://www.nber.org/papers/w21558>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2015, Revised December 2019

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Carolina Villegas-Sanchez acknowledges financial support from Banco Sabadell, Universitat Ramon Llull and AGAUR-Generalitat de Catalunya.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Sebnem Kalemli-Ozcan, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

How to Construct Nationally Representative Firm Level Data from the Orbis Global Database:  
New Facts and Aggregate Implications

Sebnem Kalemli-Ozcan, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych,  
and Sevcan Yesiltas

NBER Working Paper No. 21558

September 2015, Revised December 2019

JEL No. E0,F0,O1

**ABSTRACT**

We construct representative firm-level longitudinal data for twenty-seven European countries using financial statements from the Orbis global database, providing a “how-to” guide on the construction. We validate our dataset by comparing its aggregate coverage to official statistics and present three new facts. First, smaller firms (SMEs) account for the largest share of economic activity. Second, industry concentration has increased among firms that report only consolidated statements, but decreased overall. Third, the increased concentration is accounted for by foreign-owned firms. Documenting these facts requires nationally representative data both in cross-sectional and time-series dimensions.

Sebnem Kalemli-Ozcan  
Department of Economics  
University of Maryland  
Tydings Hall 4118D  
College Park, MD 20742-7211  
and CEPR  
and also NBER  
kalemli@econ.umd.edu

Vadym Volosovych  
Finance Group, Department of Business Economics  
Erasmus University Rotterdam  
Room E2-31  
P.O. Box 1738  
3000 DR Rotterdam, The Netherlands  
and Tinbergen Institute and ERIM  
volosovych@ese.eur.nl

Bent Sorensen  
Department of Economics  
University of Houston  
3623 Cullen Blvd  
Houston, TX 77204-5019  
and CEPR  
besorensen@uh.edu

Sevcan Yesiltas  
Koc University  
Department of Economics and Finance  
Rumelifeneri Yolu 34450 Sariyer, Istanbul  
Turkey  
syesiltas@ku.edu.tr

Carolina Villegas-Sanchez  
Department of Economics, Finance and Accounting  
ESADE Business School  
Avenida de Torreblanca, 59  
08172 Sant Cugat - Barcelona  
SPAIN  
SPAIN  
carolina.villegas@esade.edu

An online appendix is available at <http://www.nber.org/data-appendix/w21558>

# 1 Introduction

Research in macroeconomics, finance, and international economics based on firm-level data is increasingly popular. The Orbis global database, from Bureau van Dijk (BvD)—a Moody’s Analytics company—is the largest cross-country firm-level database encompassing firms’ financial statements and their production activity. It includes public and private firms’ balance sheets and income statements. The database also includes detailed information on firms’ domestic and foreign owners and subsidiaries, which allows researchers to observe global interconnections between the firms through ownership. The Orbis database is used in many research papers, but data samples downloaded from this database are often not nationally representative. In the online appendix,<sup>1</sup> we provide detailed instructions on how to download and organize data from Orbis, based on our multi-year experience in working with this data source. Using these guidelines, we put together a representative dataset for the EU countries plus Norway and present three new facts about firm activity.

We validate our dataset by aggregating our micro data by country and comparing summary statistics on firm size distribution and presence of foreign companies to aggregate official statistics. While our data is available for all sectors, the official data is available for a subset of sectors making a consistent comparison for the aggregate economy difficult. Therefore, we evaluate the representativeness of our data using two levels of aggregation. First, in the online appendix, we show that our aggregated data by firm size for the manufacturing sector is representative when compared to the firm-size distribution reported for this sector in the (official) Structural Business Statistics (SBS) database from the Eurostat. Second, we show the size distribution for the total economy, where the total economy is defined based on an aggregation of overlapping sectors in our data and in the SBS data. For this level of aggregation, our “aggregated” data covers more than 70 percent of total economic activity and has firm-size distributions that match the official data. Thus, we feel confident in using our data for all the sectors of a given country to document the firm-size distribution and not only for the sectors provided in the official statistics.

Using our dataset, we document three new facts about firm activity. First, using all the

---

<sup>1</sup>The online appendix was originally circulated under the title “How to Construct Nationally Representative Firm Level Data from the ORBIS Global Database,” an NBER Working Paper No: 21558.

sectors of a given economy and defining small and medium size enterprises (SMEs) as firms with 20–250 employees (consistent with the Eurostat definition), we find that SMEs account for more than half of aggregate employment and gross output in most countries. Second, the degree of industry concentration—defined as market shares of the largest firms—has increased only for the subset of firms that solely file consolidated firm statements. Industry concentration has decreased when we use the universe of all firms to calculate these measures, regardless of the account type. Third, the observed increase in industry concentration in the subset of firms that only report consolidated statements is due to foreign-owned firms. Although foreign firms do not account for a large fraction of aggregate output in these countries (at most 30 percent overall),<sup>2</sup> a small number of the largest foreign firms can drive trends in industry concentration.

These facts will be missed when non-representative firm-level datasets are used. Recent research has found that industry concentration, defined as the market shares of the top 4 or top 8 firms in a given 2-digit industry, has increased in the United States ([Crouzet and Eberly \(2018\)](#)), but declined in Europe ([Gutiérrez and Philippon \(2017\)](#)). [Bajgar et al. \(2019\)](#) challenged the findings of [Gutiérrez and Philippon \(2017\)](#) and argue that industry concentration in Europe has increased once the role of business groups is taken into account by considering consolidated company statements. In fact, they did not find any declining trend. We suspect this is because their dataset is not fully representative across time, because we can show both an increasing trend and a decreasing trend, depending on which firm-accounts are used in the calculations. And the fact that SMEs account for half of gross output brings attention to the need for including this group of companies when considering macro questions in economics and finance.

The Orbis database is attractive for international micro (firm)-level research because of its large universe and cross-country comparability. The data is sourced from over 160 different government and commercial information providers (see [Table A.6.1](#) in [Appendix A.6](#) for a list of the BvD information providers). BvD organizes the data in a standard “global”

---

<sup>2</sup>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 UK and Spain taking about the average values.

format to facilitate company comparisons. The financial and balance sheet information in Orbis comes from business registers, governed by country-specific legal and administrative filing requirements, and coverage of small firms and financial variables varies from country to country due to different filing requirements with national business registers.<sup>3</sup> Although most countries oblige limited liability companies to register once they are formed, requirements in terms of lower bounds on firm-size for reporting and in terms of what balance sheet items to report varies across countries.

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.<sup>4</sup> 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 since Orbis drops non-reporting 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 (WDRS), which does not systematically cover all variables. The

---

<sup>3</sup>Appendix A.6 gives details on filing requirements by country.

<sup>4</sup>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 often this cap 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.

industry classification may also be misleading because these classifications change over time as firms expand or change their operations. Single-vintage data will often over-represent larger firms and under-represent smaller firms due to survivorship bias. Some researchers opt for re-weighting the data, using weights from national censuses in order to increase the representativeness of small firms.<sup>5</sup> If our guidelines are followed, there might be less of a need to re-weight the data to obtain nationally representative firm-level datasets.<sup>6</sup>

We describe the methodology for preparing micro data based on Orbis in two self-contained appendices. In particular, Appendices A.1 to A.4 describe the organization of Orbis and our advice on data download strategies. Appendix A.5 describes how to put together the financial data in panel form, while Appendix B.1 explains the methodology used to put together foreign ownership data in panel form.

We proceed as follows. Section 2 documents our new facts and Section 3 concludes.

## 2 New Facts

### 2.1 Fact 1: Small firms account for most of the aggregate economic activity

Table 1 shows how much of gross output, as reported by Eurostat, is covered by the firms in our dataset.<sup>7</sup> We measure output by firm operating revenue. Each cell is the ratio of the value of total output produced by “our” firms relative to the value of total output from the official data. BvD provides firm-level information on gross output for all sectors of each European country starting 1999,<sup>8</sup> however official data from Eurostat SBS provides

---

<sup>5</sup>See, for example, Bajgar et al. (2019).

<sup>6</sup>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>.

<sup>7</sup>Appendix C.1 provides details on the official aggregate datasets we use for comparison purposes.

<sup>8</sup>Data go further back in time in Orbis but the coverage is not good as the regulations for filing changed in 1999 requiring all firms to file with the registries if they are located in a EU country.

information on gross output (turnover) for only a subset of sectors.<sup>9</sup> For each country-year cell, we report the ratio of our aggregated gross output values to official values across those sectors for which the gross-output related variable is available in both datasets. Some ratios will be missing for some country-years due to missing Eurostat data.

As shown in Table 1, with some exceptions, our data account for more than 50 percent of the aggregate output in all countries and close to 80-90 percent in most countries. In some countries, notably Germany, the filing requirements have changed recently and hence in recent vintages coverage has improved.

Table 2 presents the size distribution based on gross output and employment in the aggregate economy.<sup>10</sup> The patterns do not vary much by year, so we randomly choose to display the results for 2006. The table shows the fraction of economic activity accounted for by firms belonging to each size class with separate countries in the columns. In Panel A, we aggregate our firm-level data to relevant size bins based on gross output and employment. This panel shows that most of the gross output and employment are accounted for by SMEs, using the typical official definition of SMEs as firms with 20–250 employees.

To make sure that our comparisons are “apples-to-apples,” we create a “hypothetical aggregate” economy based on aggregating the sectors where official data by size class is provided by Eurostat. We report the results in Table 2, Panel B. As can be seen, our firm-level data match the official size distribution for this limited set of sectors well.

## 2.2 Fact 2: Industry concentration trends depend on the representativeness of the sample

Our second fact regards industry concentration and it illustrates the importance of using nationally representative firm-level data in establishing stylized facts. Based on datasets that cover different sets of firms, different papers find different results for European countries’ industry concentration trends. For example, [Gutiérrez and Philippon \(2017\)](#) find declining industry concentration, whereas [Bajgar et al. \(2019\)](#) find increasing concentration using the

---

<sup>9</sup>These sectors are listed in Table C.2.3 of Appendix C.1.

<sup>10</sup>Table D.1.5 of Appendix D.1 shows the validation exercise done for the manufacturing sector, for which official data on firm-size distribution is available.

Table 1: COVERAGE OF THE AGGREGATE ECONOMY BASED ON GROSS OUTPUT

Year	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU
1999	0.21	0.60		0.39	0.29	0.04	0.76	0.63	0.35	0.54	0.46	0.95		0.79
2000	0.31	0.65		0.46	0.46	0.23	0.85	0.64	0.40	0.57	0.45	0.93		0.84
2001	0.47	0.63		0.56	0.49	0.29	0.87	0.67	0.44	0.65	0.46			0.71
2002	0.54		0.57	0.65	0.58	0.29	0.88	0.67	0.47	0.62	0.48	0.24		0.73
2003	0.49	0.63	0.60	0.67	0.56	0.29	0.93	0.68	0.50	0.67	0.48	0.47		0.66
2004	0.47	0.63	0.79	0.77	0.77	0.35	0.92	0.68	0.51	0.70	0.48	0.48		0.76
2005	0.47	0.63	0.83	0.79	0.81	0.34	0.94	0.69	0.53	0.69	0.51	0.45		0.80
2006	0.61	0.62	0.84	0.75	0.68	0.35	0.99	0.71	0.51	0.68	0.53	0.47		0.81
2007	0.68	0.62	0.91	0.79	0.76	0.35	0.95	0.68	0.55	0.70	0.48	0.47		0.79
2008	0.72	0.73	0.94	0.75	0.68	0.40	0.78	0.80	0.57	0.79	0.64	0.54	0.88	0.75
2009	0.69	0.69	0.92	0.79	0.66	0.38	0.89	0.80	0.53	0.76	0.78	0.46	0.84	0.87
2010	0.71	0.65	0.96	0.93	0.65	0.37	0.91	0.80	0.59	0.80	0.68	0.45	0.00	0.76
2011	0.73	0.74	0.88	0.85	0.61	0.42	0.93	0.82	0.58	0.81	0.83	0.45	0.82	0.72
2012	0.75	0.78	0.88	0.79	0.47	0.49	0.96	0.81	0.51	0.81	0.83	0.44	0.81	0.82
Average	0.56	0.66	0.83	0.71	0.60	0.33	0.90	0.72	0.50	0.70	0.58	0.52	0.67	0.77

Year	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
1999	0.13	0.45	0.16		0.54	0.17	0.63	0.36	0.45	0.58	0.52	0.29	0.26
2000	0.32	0.50	0.18		0.51	0.21	0.63	0.40	0.49	0.53	0.56	0.66	0.33
2001	0.36	0.50	0.19		0.58	0.20	0.77	0.47	0.48	0.68	0.60	0.66	0.40
2002	0.32	0.54	0.39		0.62	0.22	0.79	0.44	0.59	0.70	0.63	0.77	0.51
2003	0.35	0.53	0.53		0.63	0.21	0.65	0.47	0.60	0.75	0.65	0.75	0.75
2004	0.36	0.57	0.52	0.63	0.84	0.22	0.67	0.57	0.63	0.83	0.67	0.74	0.77
2005	0.34	0.57	0.54	0.59	0.87	0.23	0.59	0.54	0.80	0.82	0.68	0.74	0.82
2006	0.44	0.58	0.59	0.48	0.87	0.25	0.67	0.61	0.82	0.78	0.71	0.74	0.79
2007	0.44	0.60	0.78	0.43	0.86	0.25	0.71	0.65	0.83	0.82	0.70	0.73	0.78
2008	0.49	0.72	0.62	0.44	0.88	0.28	0.59	0.55	0.82	0.86	0.73	0.80	0.97
2009	0.46	0.68	0.60	0.28	0.80	0.31	0.78	0.66	0.83	0.92	0.84	0.80	0.93
2010	0.59	0.64	0.69	0.34	0.79	0.35	0.79	0.61	0.86	0.92	0.88	0.95	0.92
2011	0.54	0.71	0.57	0.80	0.82	0.30	0.72	0.56	0.81	0.92	0.82	0.84	0.95
2012	0.40	0.61	0.48	0.72	0.78	0.28	0.67	0.59	0.93	0.91	0.76	0.96	0.91
Average	0.40	0.59	0.49	0.52	0.74	0.25	0.69	0.54	0.71	0.79	0.70	0.74	0.72

NOTES: The country codes are AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia).



Table 2: FIRM SIZE DISTRIBUTION IN THE AGGREGATE ECONOMY: 2006

	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
A: Size Distribution Based on all Sectors from Orbis Database																											
Based on Gross output																											
1 to 19 employees	0.16	0.15	0.28	0.15	0.16	0.15	0.35	0.26	0.24	0.15	0.07	0.22	0.23	0.07	0.10	0.21	0.18	0.10	0.36	0.22	0.29	0.07	0.31	0.24	0.26	0.17	0.27
20 to 249 employees	0.32	0.38	0.36	0.39	0.26	0.42	0.47	0.39	0.37	0.33	0.27	0.50	0.38	0.30	0.54	0.44	0.50	0.41	0.44	0.46	0.39	0.40	0.38	0.35	0.32	0.38	0.32
250 + employees	0.52	0.47	0.36	0.46	0.58	0.43	0.18	0.35	0.39	0.52	0.66	0.28	0.39	0.63	0.35	0.35	0.33	0.49	0.20	0.33	0.32	0.53	0.31	0.41	0.42	0.45	0.42
Based on Employment																											
1 to 19 employees	0.15	0.25	0.32	0.13	0.08	0.26	0.32	0.30	0.29	0.21	0.03	0.17	0.23	0.04	0.09	0.17	0.11	0.09	0.32	0.39	0.33	0.06	0.39	0.28	0.32	0.13	0.15
20 to 249 employees	0.34	0.38	0.38	0.42	0.21	0.44	0.46	0.39	0.38	0.35	0.25	0.48	0.35	0.30	0.37	0.44	0.58	0.36	0.42	0.37	0.39	0.36	0.36	0.34	0.31	0.37	0.35
250 + employees	0.51	0.36	0.30	0.45	0.72	0.30	0.22	0.30	0.33	0.45	0.72	0.34	0.43	0.66	0.54	0.39	0.30	0.55	0.25	0.24	0.28	0.58	0.25	0.38	0.36	0.50	0.50
B: Size Distribution in Terms of Gross Output Based on Sectors Present in Eurostat SBS and Orbis Database																											
Orbis																											
1 to 19 employees	0.05	0.06	0.11	0.07	0.08	0.06	0.18	0.22	0.15	0.09	0.05	0.16	.	0.02	0.03	0.15	0.06	0.10	0.22	0.08	0.19	0.03	0.21	0.11	0.24	0.09	0.13
20 to 249 employees	0.28	0.33	0.32	0.34	0.25	0.32	0.64	0.39	0.38	0.26	0.22	0.55	.	0.17	0.51	0.49	0.54	0.74	0.54	0.52	0.43	0.33	0.39	0.31	0.32	0.34	0.29
250 + employees	0.67	0.62	0.57	0.59	0.67	0.62	0.17	0.40	0.47	0.66	0.73	0.29	.	0.81	0.46	0.36	0.41	0.16	0.25	0.40	0.38	0.64	0.40	0.58	0.45	0.57	0.58
Eurostat SBS																											
0 to 19 employees	0.13	0.14	0.11	0.13	0.08	0.28	0.16	0.21	0.11	0.15	0.13	0.33	.	0.12	0.03	0.27	0.07	0.05	0.18	0.13	0.23	0.12	0.23	0.10	0.24	0.17	0.07
20 to 249 employees	0.34	0.28	0.32	0.31	0.24	0.31	0.56	0.38	0.23	0.27	0.27	0.28	.	0.24	0.32	0.37	0.38	0.16	0.50	0.37	0.27	0.28	0.36	0.30	0.31	0.34	0.22
250 + employees	0.53	0.57	0.58	0.56	0.68	0.41	0.27	0.41	0.65	0.58	0.60	0.39	.	0.64	0.66	0.35	0.55	0.79	0.32	0.50	0.49	0.59	0.41	0.60	0.45	0.49	0.70
C: Size Distribution in Terms of Employment Based on Sectors Present in Eurostat SBS and Orbis Database																											
Orbis																											
1 to 19 employees	0.13	0.21	0.16	0.08	0.09	0.21	0.23	0.31	0.23	0.16	0.02	0.13	.	0.02	0.08	0.16	0.06	0.24	0.22	0.29	0.26	0.04	0.33	0.15	0.32	0.09	0.09
20 to 249 employees	0.41	0.42	0.46	0.43	0.34	0.48	0.54	0.47	0.41	0.35	0.35	0.58	.	0.28	0.46	0.53	0.60	0.58	0.55	0.52	0.43	0.42	0.49	0.37	0.31	0.39	0.34
250 + employees	0.46	0.37	0.38	0.49	0.58	0.31	0.23	0.21	0.36	0.48	0.64	0.29	.	0.69	0.46	0.30	0.33	0.17	0.23	0.19	0.31	0.55	0.18	0.48	0.38	0.52	0.57
Eurostat SBS																											
0 to 19 employees	0.24	0.29	0.17	0.26	0.22	0.26	0.24	0.41	0.23	0.30	0.27	0.61	.	0.29	0.08	0.51	0.18	0.19	0.20	0.34	0.30	0.25	0.41	0.14	0.33	0.26	0.13
20 to 249 employees	0.39	0.35	0.45	0.35	0.32	0.38	0.51	0.40	0.33	0.31	0.34	0.24	.	0.34	0.51	0.31	0.51	0.44	0.53	0.38	0.36	0.35	0.42	0.38	0.31	0.36	0.34
250 + employees	0.37	0.35	0.37	0.39	0.46	0.35	0.25	0.18	0.44	0.38	0.39	0.15	.	0.37	0.41	0.18	0.31	0.37	0.27	0.28	0.34	0.40	0.16	0.48	0.37	0.38	0.53

NOTES: The country codes are AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia).

same set of countries. We believe the firm-level datasets they use have unequal coverage leading to different results.

To illustrate the importance of using a representative dataset, we measure industry concentration using different sets of firms. We use the standard indicator in the literature; that is, the market shares of the largest 4 or 8 firms within two-digit industries, by country.<sup>11</sup> Choosing different types of financial statements—called “accounts” in Orbis—will lead to different firms entering into the calculation of the industry concentration.

Usually, the company-headquarters (or the parent) of a group report consolidated financial statements (the headquarter together with all separate legal entities belonging to the group, such as the affiliates, subsidiaries, etc.) and unconsolidated statements for the headquarter only. All affiliates and subsidiaries can also report separate unconsolidated statements. Therefore, for a given company, we might observe consolidated, unconsolidated, or even both types of statements (referred to as the “accounts”), identified by a “consolidation code”.<sup>12</sup> Orbis assigns the company location (country, region, city) by the location of its headquarter. The overwhelming majority of companies in Orbis report unconsolidated accounts. However, a small number of large and very-large companies report consolidated accounts. For example, in 2006, in our sample there are 92,878 companies in the manufacturing sector in Spain with unconsolidated accounts, 274 companies with consolidated accounts, and just 147 companies reporting both. These counts for Italy are 111,200, 803,

---

<sup>11</sup>We follow NACE level 1 and level 2 classifications to aggregate 4-digit industries to 1- and 2-digit industry level data. See Table A.6.2 for NACE Revision 2, Level 2 Classification. Orbis assigns the company to a unique “primary” industry by the largest portion of its revenue; some companies might have multiple “secondary” codes (describing their additional activities). We use the primary code as the “sector” of a given company.

<sup>12</sup>The codes are C2, for the consolidated account of a company-headquarter of a group together with all companies belonging to the group, where the company headquarter also reports the companion unconsolidated account (the headquarter only); C1, for the consolidated account of a group, where the company-headquarter *does not* report the unconsolidated account; U2 for the unconsolidated account of a company with the companion consolidated account; and U1 for the unconsolidated account of a company with no consolidated account. In addition, Orbis contains companies with the accounts of type LF with limited financial information (number of employees, turnover, shareholders’ funds (book equity), capital, or some of these items), and NF with no financial items at all. By construction, the number of C2 accounts equals the number of U2 accounts. In addition, there are entities with “no recent accounts” (NRF) or “no recent limited financials” (NRLF), where “no recent” refers to last 3 years. By default, the Orbis media gives preference to the consolidated accounts, which can be changed via the Search settings. See Figure D.3.1 of Appendix D.3 for an example of how Orbis registers the multiple accounts of different types of Koç Holding Inc., the largest business group in Turkey.

and 346; for Poland, 13,300, 163, and 131.

Before constructing industry concentration measures, we categorize firms into two samples based on account type, available in Orbis. “*Sample 1*” comprises firms reporting *only one* type of account (unconsolidated (U) *or* consolidated (C)) in a given financial year, and “*Sample 2*” comprises firms reporting *both* unconsolidated and consolidated accounts in a given financial year.

We define three measures of industry concentration, using the operating revenue market shares of the top 4 (hereafter, MS4) or top 8 (hereafter, MS8) firms in a given 2-digit industry  $s2$ , country  $k$ , and year  $t$  relative to the population of all firms in the  $s2, k, t$  cell. For all measures, the population of all firms used in the denominator of market shares is exactly the same and includes the output reported by all firms from *sample 1* and *sample 2* after consolidated accounts from *sample 2* are excluded to avoid double counting.

Our market shares MS4 and MS8 would differ only by the numerator, depending on the firms selected and the account types. In the following notation, the first argument in parenthesis defines the account type we use (U or C) and the second argument denotes the sample of firms we aggregate from (*sample 1* or *sample 2*), and we omit the subscript  $s2, k, t$  for brevity:

- MS4(U,1+2) is based on *all* unconsolidated accounts (U), combining firms from *sample 1* and *sample 2*.<sup>13</sup> The goal is to compute the most inclusive measure based on unconsolidated statements.
- MS4(C,1) or MS4(C,1+2) is based on consolidated accounts (C), but using two alternative definitions: i) using firms from *sample 1* that report only consolidated accounts and ii) using *all* consolidated accounts reported by firms from *sample 1* and *sample 2*.<sup>14</sup>

---

<sup>13</sup>There is no double counting of firms because *only unconsolidated* accounts are used from *sample 2*. Numerically, the  $MS4(U,1+2) = \sum_{i \in F} y_i / Y$  where firm set  $F$  includes top 4 firms reporting unconsolidated accounts from *sample 1* and *sample 2* in a given  $s2, k, t$  cell,  $y_i$  is firm-level revenue, and  $Y$  is aggregated revenue of all firms in  $s2, k, t$  from *sample 1* and *sample 2* after consolidated accounts from *sample 2* are excluded. Market share of the top 8 firms is computed similarly.

<sup>14</sup>Compared to the measure MS4(U,1+2), in MS4(C,1) the firm set  $F$  includes top 4 firms reporting consolidated accounts from *sample 1* in a given  $s2, k, t$  cell; and in MS4(C,1+2) the firm set  $F$  includes top 4 firms reporting consolidated accounts from *sample 1* and *sample 2* in a given  $s2, k, t$  cell. In the latter case, there is no double counting of firms because *only consolidated* accounts are used from *sample 2*.

- MS4(U,2) and MS4(C,2) always use the accounts from *sample 2* and differ only by account type.<sup>15</sup> With this measure, we compare the patterns between the account types in a sample covering exactly *the same companies*.

Notice that by computing the concentration shares by country, we treat each economy as a separate market. To present the overall trends in market concentration in Europe, we aggregate our sector-country-year shares by computing the weighted averages applying the GDP weights of the countries in the total sample.

Figure 1 shows how the pattern of concentration in Europe changes depending on the sample of companies and type of statements.<sup>16</sup> In Panel A, we focus on unconsolidated statements and plot market shares MS4(U,1+2) and MS8(U,1+2) of the top 4 and top 8 firms in the largest possible sample of companies reporting unconsolidated accounts, which come from the combined *sample 1* and *sample 2*. In Panel B, the green line presents the pattern when we still use the unconsolidated accounts, but from a narrower *sample 2* and for the top 8 companies (the measure MS8(U,2)). Thus, the green lines in both panels differ only in firm sample. It is clear that market concentration declines in Europe during this period in both the manufacturing and services sectors.

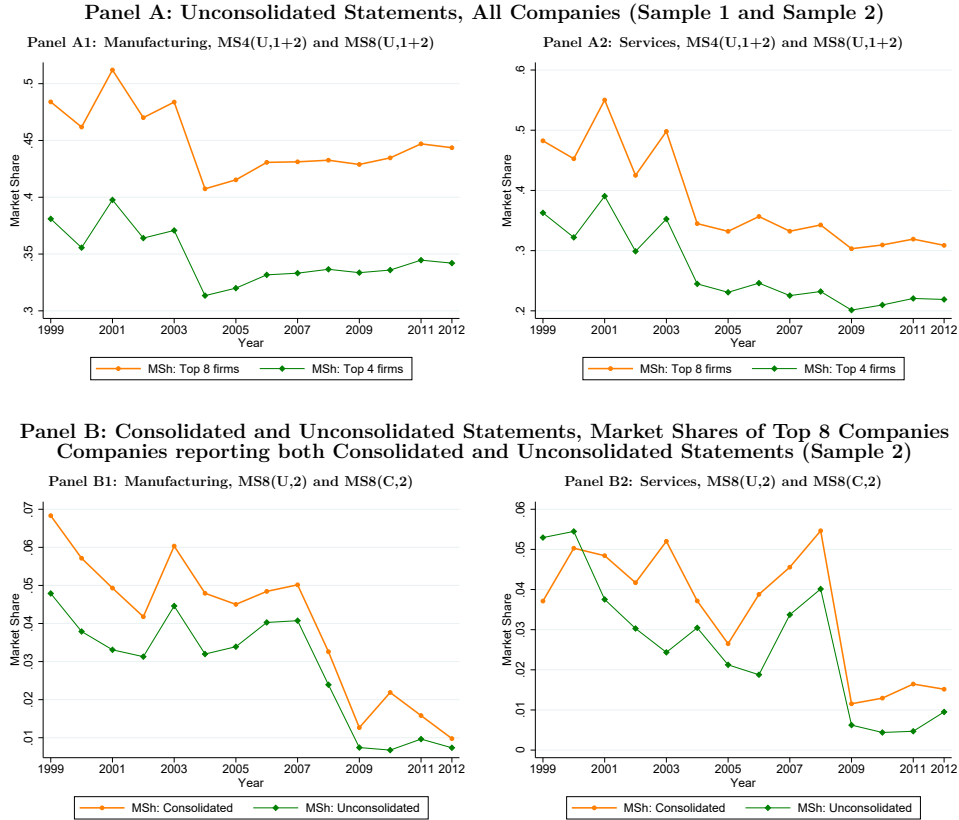
We recognize that a parent company and its subsidiaries might belong to several industries and countries and this distinction might be important for conclusions about industry concentration, precisely because the largest firms in an industry are typically the parents

---

<sup>15</sup>Compared with the measure MS4(U,1+2), the firm set  $F$  includes the top 4 firms from *sample 2* in a given  $s2, k, t$  cell, using their unconsolidated accounts in MS4(U,2) and their consolidated accounts in MS4(C,2). There is no double counting of firms because the underlying measure is constructed separately for a given account type.

<sup>16</sup>Our sample covers Austria, Belgium, Czech Republic, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Italy, Latvia, Luxembourg, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, and Sweden. We do not include Denmark, Ireland, Netherlands, and Lithuania for which our dataset accounts for less than 50 percent of the aggregate. Bulgaria and Croatia are excluded as outliers. The services industry covers all non-financial market services including “accommodation and food services,” “information and communication,” “real estate activities,” “professional, scientific and technical activities,” and “administrative and support services activities.” Our sample is, thus, an extended version of the sample analyzed in [Dotting et al. \(2017\)](#), [Gutiérrez and Philippon \(2017\)](#), and [Gutiérrez and Philippon \(2018\)](#). They use the BvD merged vintage dataset we provided, focusing on a benchmark “EU KLEMS” sample, which consists of Austria, Belgium, Germany, Spain, Finland, France, Italy, and the Netherlands, and for some comparisons Great Britain and Sweden. In addition, they exclude *small* firms and the following sectors: utilities (KLEMS segments D35 to D39), financial firms (segments D64 to D66) and real estate (segment D68), also segments D84 (public administration and defence; compulsory social security) and D99 (activities of extraterritorial organizations and bodies).

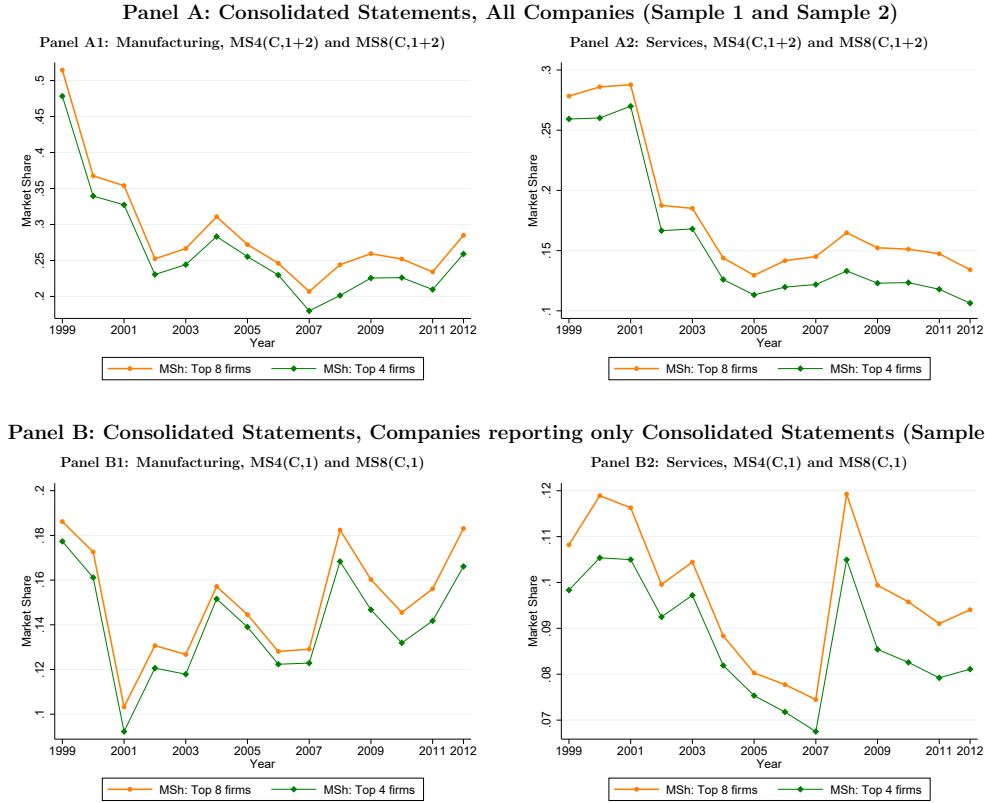
Figure 1: CONCENTRATION IN EUROPE: MARKET SHARE OF TOP FIRMS, UNCONSOLIDATED AND CONSOLIDATED STATEMENTS



NOTES: The figure plots market concentration in the manufacturing industry and the services industry for European countries. The services industry is defined as “non-financial market services,” including “accommodation and food services,” “information and communication,” “real estate activities,” “professional, scientific and technical activities,” and “administrative and support services activities.” We measure country-sector-year market concentration by the market share of largest four or eight firms in terms of firm-level operating revenue from our Orbis dataset in the given country-sector-year, relative to the overall country-sector-year output. In all graphs, we plot “EU-weighted” average shares over the period 1999–2012, using a given country’s GDP weight in the total sample in the corresponding year.

of larger groups. If the patterns of concentration are affected by, for example, mergers and acquisitions when acquisition of a large stand-alone company increases a given company’s consolidated balance sheet dramatically, then the type of account is used in calculating concentration might be important. Thus, in Figure 2 we plot the evolution of market concentration in Europe using consolidated accounts of parents. In Panel A, we plot market

Figure 2: CONCENTRATION IN EUROPE: MARKET SHARE OF TOP FIRMS, CONSOLIDATED STATEMENTS



NOTES: The figure plots market concentration of manufacturing industry, and services industry for the European countries. The services industry is defined as “non-financial market services,” including “accommodation and food services,” “information and communication,” “real estate activities,” “professional, scientific and technical activities,” and “administrative and support services activities.” We measure country-sector-year market concentration by the market share of largest four or eight firms in terms of firm-level operating revenue from our Orbis dataset in the given country-sector-year, relative to the overall country-sector-year output. In all graphs, we plot “EU-weighted” average shares over the period 1999–2012, using a given country’s GDP weight in the total sample in the corresponding year.

shares MS4(C,1+2) and MS8(C,1+2) of the top 4 and top 8 firms reporting consolidated accounts from the largest pool of firms possible. These top firms are defined as having the largest revenue among all consolidated accounts that come from the combined *sample 1* and *sample 2*. Regardless of statement type, we find declining concentration as in [Gutiérrez and Philippon \(2017\)](#) and [Gutiérrez and Philippon \(2018\)](#). In Panel B, we plot market shares

MS4(C,1) (the green line) and MS8(C,1) (the orange line) of the top 4 and top 8 firms that report only consolidated accounts, selected from *sample 1*. In this panel, we find an increasing trend for concentration, especially for manufacturing sector, replicating the results of Bajgar et al. (2019).

All our measures of concentration might be impacted by the fact that we do not cover the universe of firms in any country, although we are close with most countries' coverage being around 70 percent. Nevertheless, in online Appendix D.4, we demonstrate that defining market shares of the top firms vis-a-vis the entire economy does not change our results. To do this, we use the industry-level aggregates reported in the OECD STructural ANalysis (OECD STAN) database (available at <http://stats.oecd.org/Index.aspx?DataSetCode=STAN08BIS>) as the denominators of our market shares.

### 2.3 Fact 3: Increasing industry concentration in Europe is due to the presence of foreign firms

In order to dig deeper and understand what drives the increasing concentration in Panel B of Figure 2, especially for the manufacturing sector, in this section we separate foreign and domestic top firms. The literature has shown that foreign companies/multinationals are large and operate through a network of subsidiaries and affiliates in several countries (e.g., Kalemli-Ozcan et al. (2014)). We split the top 8 firms in each country-sector-year into domestic firms and foreign firms and redefine concentration measure MS8(C,1) accordingly.<sup>17</sup>

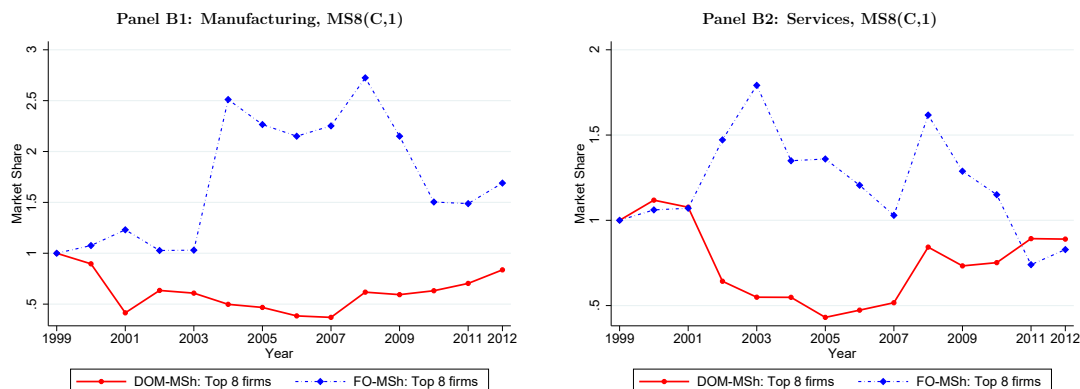
Figure 3 reports these concentration measures for top 8 firms, using firms who report *only* consolidated accounts, drawn from *sample 1*. Unlike Figures 1 and 2, Figure 3 displays the market shares as an index, relative to the value in 1999. It is clear that the patterns of increasing concentration observed in Figure 2 are driven by foreign firms and the pattern is starker for the manufacturing sector than for the services sector. These patterns are

---

<sup>17</sup>That is, for example, in a given country-sector-year, among the largest 8 firms, if 2 firms are domestic and 6 firms are foreign, the summation of the revenue in the numerator of MS8(C,1) is done for 2 domestic firms, and 6 foreign firms, respectively. For any other combinations of foreign-domestic split, MS8(C,1) is computed similarly. Therefore, the vertical summation of the resulting shares equals the values represented by the orange line in Panel B of Figure 2. Any firm whose equity is owned by foreigners in excess of 10 percent is defined as a foreign firm, following the balance-of-payments definition of the IMF.

Figure 3: CONCENTRATION IN EUROPE: MARKET SHARE OF FOREIGN AND DOMESTIC FIRMS WITHIN TOP 8 FIRMS, CONSOLIDATED STATEMENTS

Panel B: Consolidated Statements, Companies reporting only Consolidated Statements (Sample 1)



NOTES: The figure plots market concentration measures for the manufacturing industry and the services industry. The services industry is defined as “non-financial market services,” including “accommodation and food services,” “information and communication,” “real estate activities,” “professional, scientific and technical activities,” and “administrative and support services activities.” We measure country-sector-year market concentration by the market share of largest eight foreign (domestic) firms in terms of firm-level operating revenue from our Orbis dataset in the given country-sector-year, relative to the overall country-sector-year output. The output of the largest firms is determined from the subset of companies reporting *only* consolidated financial statements, drawn from *sample 1*. In all graphs, we plot “EU-weighted” average shares over the period 1999–2012, using a given country’s GDP weight in total sample, in the corresponding year.

not biased by unrepresentative sets of foreign firms because our data is fully representative also in terms of foreign firms. To validate our data along this dimension, 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.<sup>18</sup> Figure B.3.1 of Appendix B.3 graphically presents

<sup>18</sup>AMNE (available on the OECD data portal [https://stats.oecd.org/Index.aspx?DataSetCode=AMNE\\_IN](https://stats.oecd.org/Index.aspx?DataSetCode=AMNE_IN)) covers 28 OECD host countries from 2008 onwards, 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 covers manufacturing only (see [http://stats.oecd.org/Index.aspx?DataSetCode=AFA\\_IN3](http://stats.oecd.org/Index.aspx?DataSetCode=AFA_IN3)). The OECD Foreign Affiliates Statistics (FATS) database gives detailed data on the activities of foreign affiliates in the *services* sector, although, 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



this comparison. The figure demonstrates that our data capture very well the share of multinational activity in total activity reported by the official statistics.<sup>19</sup>

### 3 Conclusion

We construct nationally representative firm-level longitudinal datasets for twenty-seven European countries using firms’ financial accounts from the Orbis global database and provide a road map as a “how-to” guide. We validate the data by comparing aggregates from our constructed data to corresponding official aggregate data. We provide three new stylized facts using our data. First, small and medium size firms (SMEs) account for more than half of the aggregate economic activity in most countries. Second, the degree of industry concentration—defined as the market shares of the largest firms—can be shown to have an increasing and a decreasing trend depending on the type of financial accounts used to calculate the concentration measures. An increasing trend is observed for a selected sample of firms who report only consolidated accounts. Third, the increase in industry concentration observed for consolidated accounts is due to foreign-owned firms. Although foreign-owned firms do not account more than 30 percent of aggregate output, they are important drivers of industry concentration trends.

---

and foreign affiliates abroad controlled by residents of the compiling country. Surveys are conducted, in most cases, by the national statistical office or the central bank of each country. While the key variables in the survey are common across countries, the target sample varies across countries. See Appendix B.2, for a description of the issues considered to maximize comparability across samples.

<sup>19</sup>Appendix B.3 provides these statistics country 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 Appendix B.1.

## References

- Aykut, Dilek, Apurva Sanghi, and Gina Kosmidou**, “What to Do When Foreign Direct Investment Is Not Direct or Foreign: FDI Round Tripping,” Working paper WPS8046, The World Bank 2017.
- Bajgar, Matej, Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Jonathan Timmis**, “Industry Concentration in Europe and North America,” Working paper 18, OECD 2019.
- Crouzet, Nicolas and Janice Eberly**, “Intangibles, Investment, and Efficiency,” *AEA Papers and Proceedings*, 2018, 108, 426–31.
- Dottling, Robin, Germán Gutiérrez, and Thomas Philippon**, “Is There an Investment Gap in Advanced Economies? If so, Why?,” Conference proceedings, ECB Forum on Central Banking: Investment and Growth in Advanced Economies, 26-28 June 2017.
- Force, CompNet Task**, “Micro-based Evidence of EU Competitiveness: The CompNet Database,” NBB Working paper, National Bank of Belgium, Brussels 2014.
- Gutiérrez, Germán and Thomas Philippon**, “Declining Competition and Investment in the US,” Technical Report 2017.
- and —, “How EU Markets Became More Competitive than US Markets: A Study of Institutional Drift,” Working paper 24700, National Bureau of Economic Research 2018.
- Kalemli-Ozcan, Sebnem, Bent Sorensen, and Vadym Volosovych**, “Deep Financial Integration and Volatility,” *Journal of the European Economic Association*, 2014, 12 (6), 1558–1585.

# Online Appendix

to

*How to Construct Nationally Representative Firm Level Datasets from the Orbis Global  
Database: New Facts and Aggregate Implications*

by

Şebnem Kalemli-Özcan,

Bent Sørensen,

Carolina Villegas-Sanchez,

Vadym Volosovych,

Sevcan Yeşiltaş.

## A Construction of the Dataset

### A.1 Organization of Orbis and Amadeus Databases

This appendix explains the practical aspects of constructing the panel data of financial and ownership information that is usable for research purposes based on Orbis and Amadeus datasets provided by BvD Electronic Publishing. In the following pages, we describe what we regard as the “best practice” for obtaining the most comprehensive financial- and ownership datasets over time.

#### A.1.1 Accessing BvD Products

BvD’s two best known products for firm-level data are the global database Orbis and the European database Amadeus. There are other products which are either country specific or region specific (for example for the UK and for the Asia-Pacific region). All these products cover listed and unlisted firms. Amadeus was the original flagship product of BvD with many its features incorporated later on into Orbis database. In what follows, we will write about Orbis and stress the unique features of Amadeus separately.

A researcher can access Orbis and European Amadeus databases in three ways.

1. BvD proprietary browser online ([orbis.bvdinfo.com](http://orbis.bvdinfo.com) and [amadeus.bvdinfo.com](http://amadeus.bvdinfo.com)).
2. BvD historic (CD/DVD-ROM, Blu-Ray) disks or a new related solution Orbis Historical with the linked historical data ([www.bvdinfo.com/en-us/bvd-for-your-business/academic](http://www.bvdinfo.com/en-us/bvd-for-your-business/academic) described under “I need to research historical datasets and changes over time” QnA item). At the moment of writing of this paper, the Orbis Historical solution is still in the developing stage. Therefore, we focus on the steps required to work with the separate historic disks.
3. Through the Wharton Research Data Services (WRDS) from the Wharton School at the University of Pennsylvania (Amadeus only).

Each of the access methods has its benefits and costs for a researcher; the decision about

which one is the most optimal depends on the research budget and the type of the data one expects to work with.

We refer to the information available from BvD disks as “vintages” of the BvD data. We use the same term when we talk about the data retrieved from an online access to BvD or WRDS at a given point of time. In each case, “vintage” will mean the release date of the disk or the time of online access, respectively.

Each product is split by the type of information provided. For example, Orbis contains the “sections” Orbis Financials with firm financial information and Orbis Ownership with ownership information. There are other sections in Orbis with valuable information such as Auditors and Advisors, Board Members, Patents, etc.

The users who obtain Amadeus through WRDS (Method 3) should be aware that the internal organization of the whole database (Amadeus Financials and Amadeus Ownership) consists of three non-overlapping files corresponding to three company size “tiers.” The tiers are defined by BvD in terms of the size of company sales, operating revenue, and employment. The thresholds of these variables for the companies to be classified by BvD as Very Large & Large, Medium, or Small vary by country (e.g., companies in Eastern Europe may have lower sales in order to be categorized as Very Large & Large). This matters when querying and downloading the data with the SAS software available under the WRDS subscription.

### **A.1.2 Financial Module**

The Orbis Financials sub-database includes detailed information about numerous balance sheet items, profit and loss account items, and financial ratios over time as well as static descriptive variables. The descriptive information includes, among other items, official national identification number, address (country, region, city, street), legal form, year of incorporation (entry), status of the company (active/liquidation/merger-acquisition), number of employees, quoted/unquoted indicator, industry and activity codes (4 digit level) and, when available, the description of the nature of the business in the local language and English.<sup>20</sup>

---

<sup>20</sup>The default option is to download textual information, such as company names, in the original format. We encountered problems with the encoding of non-Latin alphabets and non-standard national letters. The user should choose the available option to show the textual information in the international alphabet rather than the original alphabet before download.

Historic (time-series) financial information can be downloaded from the web or from a single disk by selecting several historical years, although there are several issues to be aware of:

- *Download speed and cap issue.* BvD platforms or disks are not designed for the purposes of academic research involving large amounts of data. Extracting large amounts of data from any BvD platform is in general slow and BvD puts a cap on the amount of information researchers can download both from disks and also from its own website. Most of the time this cap does not turn into a termination of the download job, but rather the downloaded files will have missing information. WRDS is the most user-friendly platform because the imbedded WRDS browser allows researchers to run optimized queries and compresses the data at the time of download; WRDS also allows the researcher to retrieve the data by running a SAS for UNIX code directly at their servers. However, WRDS covers only Amadeus, and Orbis and Amadeus do not overlap 100 percent in terms of companies and variables even for a given European country (any country outside Europe will not be in Amadeus).<sup>21</sup>
- *Survivorship bias.* Both Orbis and Amadeus contain historic financial data for a number of years. However, the two databases follow somewhat different rules regarding the inclusion of companies and years. Amadeus provides at most the ten most recent reporting years for the same company while Orbis de facto reports data for the five most recent reporting years at most. Amadeus will delete a company from the database if the company did not report anything in the last 5 years, while Orbis will keep this company as long as the company is active in the business register.
- There is a *reporting lag* of about 2 years, on average, and there are differences in the coverage of particular variables depending on when the BvD product has been released. Hence, for the 2010 vintage, a company may not have the 2010 filings but the 2010 filings will appear in the 2012 vintage. BvD's data collection improves over time and hence this lag might vary by country and by data vintage.

---

<sup>21</sup>For the same company in a given year, it is possible to have more non-missing variables in Orbis than in Amadeus. It is also possible to obtain more firms for the same country and year in Orbis than Amadeus. We have confirmed that both cases are widespread.

- Issues with the *presentation format*. Certain variables, such as employment, will not be on the balance sheet, but rather in memorandum items.
- *Merging issues*. BvD identifies each company by a unique company ID, which is designed to trace the same company across all BvD products. However, a researcher merging the time-series financial information coming from several BvD historic disks, or the online downloads done at various points of time, may encounter occasional *BvD ID changes* over time. The BvD ID number incorporates either the national ID number or the ID provided by their information providers (IP). According to BvD, the ID numbers may change when the national ID numbers change in the official data sources or the BvD IPs decide to switch their ID numbers. The ID changes are related to changes of address, legal form, or M&A activity. In acquisitions, acquiring company will keep its ID and the target’s ID is blocked. BvD mentions that Spanish companies encounter a BvD ID change if they change legal form, while companies incorporated in Germany, Austria, or Italy in some cases see their BvD ID change if the company changes address.<sup>22</sup> Finally, BvD itself can initiate the ID change when an entity is available on more than one product, or is provided by more than one IP, and BvD harmonizes the IDs across databases using a set of priority rules. As long as BvD does not know that a certain company is the same entity, it will have several different BvD ID numbers on Orbis . Because it is hard to keep track of all these idiosyncracies, the researcher should request the “correspondence table” of BvD IDs from their BvD representative. BvD ID changes can also be obtained by subscribing institutions via the dedicated BvD ID Change Lookup tool at [idchanges.bvdinfo.com](http://idchanges.bvdinfo.com).

### A.1.3 Ownership Module

Orbis Ownership sub-database contains information on each company’s equity ownership structure: the names of owners, their respective ownership shares, the level of ownership (direct or ultimate cross-ownership), their countries of origin. For each owner of every target

---

<sup>22</sup>If a company moves from area 1 to area 2 and area 2 has a different office collecting the information, the company will get a new national ID, thus the BvD ID number changes. If in area 2, the same office is responsible for collecting information, the company keeps its national ID.

firm there is one observation (we refer to such a record as an “ownership link”). There are two major issues of concern for the construction of time-series ownership information.

- The *vintage issue* for ownership. BvD proprietary browser online and WRDS contain only the *latest* available ownership information. If one were to access ownership information through the browser, through a specific vintage disk of Amadeus or Orbis or through WRDS, ownership information will be static (“as of date”). The only option to reconstruct the historic (time-series) ownership information is by purchasing the historic Orbis disks from BvD. It is possible to retrieve the historic ownership information through the BvD browser access (Method 1) or historic disks (Method 2) for individual companies by revealing the company’s Standard report and then selecting “the Shareholders history - current, previous and archived data” option in the *Ownership data* menu. But this method is not practical if one is interested in large datasets. In addition, the extent of the survivorship bias through this method is not clear to us.
- *Merging issues*. The same issue related to BvD ID changes emerges if one downloads ownership data from several vintages of Orbis. The issue is more acute because one needs to rely on annual vintages to track annual ownership changes.

There is a separate BvD product that tracks all mergers and acquisitions, that is *changes* in ownership, at the transaction level over time (Zephyr). In principle, one can supplement the ownership stakes from Orbis Ownership with the transaction data from Zephyr, by adjusting the equity stakes reported in Orbis Ownership prior to transaction. The data we present in this paper incorporates such an adjustment.

Next, we propose a detailed approach to constructing the most comprehensive financial and ownership information for a representative set of firms over time.

## A.2 Downloading from Orbis and Amadeus Databases

While Amadeus and Orbis have an impressive number of unique firm IDs, as many researchers have discovered, a large number of those IDs contain only information on company name and a few other variables. When a researcher requests some key variables, such as total assets,



sales or employment, these turn out to be missing. As we detailed out above there are several reasons for this. It is necessary to download data in a non-standard way to overcome these problems. As we have explained in the previous chapter, there are three different ways to access BvD data:

1. Through BvD's proprietary web platform available by the direct subscription.
2. Through BvD's historical vintages, available on historic CD-ROM disks (or harddrives/blue-ray disks).
3. Through WRDS archives.

The standard and the most commonly used method is Method 3. To avoid the pitfalls mentioned above one must follow Method 2, however. Let us explain the advantages of Method 2 over other methods and how one can maximize the coverage and representation of small firms while recovering many financial variables by using Method 2.

To maximize coverage (for European countries), a researcher must use both Orbis and Amadeus and several vintages from both databases.<sup>23</sup> The reason is that these databases follow different rules regarding the inclusion of companies and years.<sup>24</sup> Amadeus provides at most 10 recent years of data for the same company while Orbis, de facto, only reports data for up to 5 recent years, despite the possibility of choosing 10 years of data going back in the Orbis software. The justification given to us by BvD was that the global Orbis database contained much more information and the information included in a given vintage had to be limited because of the media capacity. Moreover, Amadeus drops firms from the database if they did not report anything during the last 5 years while Orbis keeps the information for these companies as long as companies are still in the business register. This problem makes it clear why Method 2 is superior to other methods. A company might file information with BvD for the last time in year 2007. However, in the business registry, this company is still active. Due to non-reporting in the last 5 years, in Amadeus-2013 vintage this company will

---

<sup>23</sup>For countries outside Europe, the only option is Orbis. Our instructions below related to Orbis will apply to other countries, such as the United States, in terms of maximizing coverage and representation.

<sup>24</sup>One also needs to keep in mind that Amadeus, being a regional database, includes some Europe-specific variables that the Orbis Global Standard Format may not contain. Plus some variables may be coded slightly differently in the two databases; for example, the type of owner is textual in Amadeus while in Orbis this variable contains standardized single-letter codes.

not be included, but the same company’s information for the period 2002–2007 will still be reported in Orbis-2013 disk. In addition, because there is a *reporting lag* of financial data of usually 2 years (it varies by country), the coverage of, for example, years 2007 and 2008 from the 2009 Orbis disk (or an online download done in the year 2009) will be very poor. For this purpose, again Method 2 will be better because recent database vintages will complement earlier ones and, hence, one can get more firms for the years 2007 and 2008 from the 2010 vintage (or the 2010 WRDS download). It is also the case that information is updated over time and some variables that were not available in early disks is made available in later vintages.

There are differences in coverage of certain variables across Orbis and Amadeus going back in time. The reason for this might be a combination of the issues mentioned above, or the fact that all the access methods will *cap the amount of information* one can download in one run (the number of firms and the number of variables). This cap unfortunately translates into missing observations in the resulting download instead of termination of the download job. This issue can also be dealt with, if we use Method 2. To illustrate the problem, consider a researcher who in April 2015 wants to obtain data on Spanish firms for the year 2006. The researcher goes to WRDS (Method 3) and downloads the data by choosing the year 2006. Table A.1.1 presents the distribution by size category in 2006 of firms in Spain in our data constructed based on the methodology we prescribe in this paper (RAW) and the same size distribution based on the data downloaded from WRDS on April 2015 for the year 2006 for Spanish companies. Panel A uses employment to measure firm size distribution, whereas Panel B uses wage bill. In Panel A, in column (1) we see that firms with less than 20 employees account for 25 percent of total employment, whereas firms who have between 20–249 employees account for almost 50 percent and firms who have more than 250 employees account for 26 percent of employment. Panel B delivers a similar size distribution. The point we want to make is that if we use our data as shown in column (1), put together from several vintages using Method 2 for download, or we use direct download from WRDS for the year 2006, where we access WRDS on April 2015 as shown in column (2), we obtain a similar size distribution.

What is the problem then, if our data and a single shot download from WRDS delivers the same data? The problem starts when the researcher wants to have full information on a

certain set of variables. For example, if we want to calculate total factor productivity, then we need to have the variables output, employment, capital stock and materials reported and, hence, we need to go down to a subset of firms that report all these variables. When we do that column (3), which is our data, performs very well and delivers a similar size distribution. However, column (4), which is the direct download from WRDS as of April 2015, performs clearly worse because there are practically no small firms reporting information on materials. This is an artifact of the one-time download from WRDS. As we show in Table 2, the size distribution based on our data for several countries (including the ones shown in Table A.1.1) matches the official size distribution provided by Eurostat based on national censuses.

Table A.1.1: Company coverage comparison in the Spanish manufacturing sector in 2006 obtained from our data (RAW) vs. Amadeus online from WRDS

Sample	RAW	WRDS	RAW-TFP	WRDS-TFP
Panel A: Employment				
1-19	24.7%	24.0%	24.2%	0.6%
20-249	49.2%	49.1%	49.7%	50.1%
250+	26.1%	26.9%	26.1%	49.3%
Panel B: Wage Bill				
1-19	19.6%	18.0%	19.2%	0.6%
20-249	46.6%	44.7%	47.0%	44.4%
250+	33.8%	33.5%	33.8%	53.4%

NOTES: RAW refers to the sample of firms after basic cleaning in our data. WRDS refers to the sample of firms as obtained from the WRDS after basic cleaning. TFP refers to the sample of firms after basic cleaning and with the required information to compute total factor productivity (TFP), i.e., non-missing values for employment, output, capital stock, and materials. WRDS-TFP refers to the sample of firms in WRDS after basic cleaning, with available information to compute TFP.)

In order to maximize the coverage of firms and variables by country over time, our download strategy (Method 2) **for financials** makes use of several vintages of BvD products: Orbis disk 2005, Orbis disk 2009, Orbis disk 2013, Amadeus online 2010 (from WRDS; accessed in May), and Amadeus disk 2014. We chose these vintages to ensure a time overlap to get around the reporting rules in Amadeus and Orbis.

**For ownership** our download strategy is more involved because, as we explained, any

of the three access methods—browser access (Method 1 or 3), or the current vintage at hand (Method 2)—would give access to the ownership information “as of date.” We use the Method 2 download strategy, but construct the historic (time-series) ownership information by purchasing the historic Orbis disks from BvD *for every year*. Because we prefer to record the ownership information as of the end of each calendar year, we choose the Orbis disk issued as closely as possible to the end of the desired year. For example, to obtain the ownership as of end of 2010, we use the Orbis disk issued in January 2011, and so on (disks are issued monthly but firms report yearly, though ownership can change within a year). To take full advantage of slight differences in ownership data in Amadeus and Orbis, we combine bi-annual vintages of Amadeus Ownership with annual vintages of Orbis Ownership. These differences will be discussed in Section [B.1](#).

## A.3 Financial Reporting in BvD Databases

### A.3.1 Time Stamp

Before downloading, one has to choose how to time stamp the year of financial data. One option is the conventional “absolute years,” where the year of the financial accounts explicitly refers to calendar years, 2006, 2007, and so on. An alternative is the so called “relative years,” where the most recent year of non-missing financials (as available to the BvD) is referenced as the “Latest Year” and the earlier observations are referenced as the Latest Year  $-1$ , Latest Year  $-2$ , and so on.

We advise to download financials via the “relative years” option. While it a priori seems counter intuitive, the relative year option is superior to the absolute year option, because of the reporting lag and the survivorship bias-issues detailed above. The distinction in the date option is extremely important for the companies which report irregularly, with gaps in their time series. For such companies, 5 relative years might cover a longer calendar period than 5 years requested explicitly (recall that Orbis de facto reports only the most recent 5 years of a given company, with a reporting lag of 1–2 years). For example, consider a company reporting data for year 2007, and then again for each year between 2009 and 2012. Then, asking for the 5 latest absolute years 2008–2012 would leave us with just 4

observations (2009–2012), while asking for 5 relative years would retain the 2007 value (the 2007 is referenced as “Latest Year –4” in this case). To obtain the correct assignment of the relative years in terms of calendar years, we use the special BvD index variable (Closing date, Latest Year; Closing date, Year –1 and so on).

### **A.3.2 Consolidation of Subsidiaries**

For a given company, the BvD databases report one or more financial statements, whose type is specified by the “Consolidation Code.” Most of the large companies with good coverage report either consolidated accounts (the statement of a parent company integrating the statements of its controlled subsidiaries) or unconsolidated accounts (the statement not integrating the statements of the controlled entities). Some companies report both kinds of accounts. Finally, there are entities with limited financial data, no recent financials (where the last available accounts are more than 48 months old), and the combination of the two. In most of these cases, only the number of employees and the operating revenue are available. The type of account reported is related to country filing requirements for particular size or legal type of companies, as detailed in Table A.6.1 (e.g., the non-independent branches (establishments) are often included in Orbis for the United States). To speed-up the process, we download the data separately for non-limited financial accounts (which includes most companies) and limited financials accounts. We download both consolidated and unconsolidated accounts and, so far, use unconsolidated accounts in all of our applications. Any user can choose which account to use by looking at the “U” or “C” letters at the end of the firm ID (or by using the BvD consolidation code) for unconsolidated and consolidated statements, respectively. Consolidated accounts will involve double counting when both consolidated account of the parent (with all its subsidiaries) and the unconsolidated account of the parent (without subsidiaries) are reported.

### **A.3.3 Units and Currency of Financials**

Balance sheet financial variables are book values. Care should be exercised when choosing how the financial data will be downloaded both in terms of the monetary units and currency. By default, the formatted export from BvD disks will be in the units and currency in which

a particular company originally filed its financials. This means that a given company may report in thousands in some years and in millions in other years. To avoid spurious jumps in the data, the best practice is to choose the units (units, thousands, millions) in the Formatted Export Wizard explicitly, rather than use the default. In case the default was chosen, the variable `UNITS` lists the units a given observation is recorded in (the values are textual in Orbis and numeric powers of 10 in Amadeus). In the latter case, the harmonization is achieved by the transformation  $x = x * 10^{\text{UNITS}}$ . But note that the `UNITS` variable might have errors in certain disks, as we have discovered and, therefore, the best practice is to force the download to be in specific units (such as thousands) and not use the default.<sup>25</sup>

We choose the data in the original currency filed by the company, which might be the national currency of the country or sometimes a foreign currency. The currency of the given account is available in the variable “Account published in currency.” It is tempting to use the available BvD variable “Exchange Rate” in order to convert the data from all companies into a single currency. We advise against using this variable, because different products use it for different purposes. In the Orbis or Amadeus disks, once we choose to download the data in “local currency” the values of the variable “Exchange Rate” are always set to 1. However, in the Amadeus downloads from WRDS, this variable contains the actual exchange rate of the currency of account against the U.S. dollar. Hence, in general, the variable “Exchange Rate” does not contain the rate versus, say, the U.S. dollar (as was the case with WRDS Amadeus), but has the exchange rate of the currency of an account in relation to the currency chosen by the person downloading the data.

## A.4 Data Selection and Exporting

Before discussing the merging and cleaning steps, we touch upon some other subtle issues in the setup of the download process. The process starts from selecting the sample of companies in a given BvD disk, typically by country.<sup>26</sup> To overcome the download limits, large countries

---

<sup>25</sup>By errors, we mean the cases when the value of the `UNITS` switches from, say, thousands to millions, but the corresponding financial variables do not show the 1000x decrease in the order of magnitude. We describe a filter we developed to check for these issues in Section [A.5.2](#).

<sup>26</sup>One must choose industrial companies from the start since there are also banks and insurance companies in Orbis. “Type” variable helps this selection. In addition global format is what is available for private firms whereas detailed format is available for listed firms.

may be downloaded by regions. An alternative is to use sectoral splits in a given country. The software then displays the set of the selection criteria, combined by the “AND” expression, and the resulting number of unique companies satisfying all the criteria. One can replace the “AND” expression to adjust the selection criteria and/or display the list of companies. Regardless of the variable used for splitting the large country data for separate downloads, it is necessary to watch out for the cases when the variable used for splitting is not available for a subset of companies. For that, one needs to select the criterion of the entire country (such as “all companies in the U.S.”) and add the criterion where *all* regions or sectors are explicitly listed, but precede the latter condition with the “NOT” expression (AND NOT “companies in sectors A, B, C...”).<sup>27</sup>

To download from the *older disks* (issued prior to January 2012 for Orbis or December 2010 for Amadeus) one has to navigate to the Formatted Export menu (**File>Export>Formatted Export** To on the upper-left) in order to adjust the default list of variables; the user does not need to display the list of companies after this. In the Formatted Export menu one can also choose the time period of the data (absolute/relative years), currency, units of financials, and then the export layout. See Section E.1 for illustration of this interface. An alternative, “List Export” method is accessible from the list of companies. To adjust the default list of variables and make other choices, one has to navigate to **Options>List Format>New Format** menu. We do not discuss this method, because we did not see a clear advantage over the Formatted Export method in older BvD disks. If anything, a potential disadvantage is the increase in waiting time for the list of companies to be displayed before the download can start.

In *newer disks* (issued after January 2012 for Orbis or December 2010 for Amadeus), which emulate the BvD online interface (Method 1), more download options are available after the resulting number of unique companies satisfying the criteria is revealed. After selecting the Formatted Export menu on the right, one now needs to further choose from the Database export, Statistical export (not useful), or Custom export. The closest analog to the Formatted Export interface in the older disks is the Custom Export where one can adjust the variables, data time stamp (absolute/relative years), units, currency, and layout. Section E.2

---

<sup>27</sup>This additional download can be ignored if the number of companies lost due to the missing splitting variable is negligible.

shows the details of its interface. However, for unknown reasons the list of variables one can choose from does not include all the variables available in Orbis.<sup>28</sup> In particular, none of the ownership data is available. A potential remedy is the Database Export (another sub-heading of Formatted Export), specifically designed for downloading large amounts of data according to BvD. For the interface of this method, see Section E.3. Database Export lets one choose all the variables in Orbis, currency, and units. The download works much faster than other methods, which is an advantage for large downloads such as ownership data. In our experiments, we saw roughly a twofold time gain compared to other methods. However, a serious drawback of the Database Export is the inability to download financials with the option of “relative” years (it only downloads in terms of “absolute years”). For this reason, we have to choose the Custom Export option as the only way to force the relative years download. For ownership downloads the Database Export is acceptable because, as we explained in Section A.2, the time dimension is irrelevant given the availability of ownership data as a snapshot in a given release. A final option is the List Export, illustrated in Section E.4. One can chose absolute and/or relative years and all of the variables under the List Export option. Hence, this method works for both financial and ownership data downloads. The only drawback of List Export is the extra time needed to first reveal the list of companies before the download can commence.

## A.5 Building Financial Panel Data

### A.5.1 Vintage Raw Data

The following are the key steps we take in constructing the firm-level financial database. We execute these steps for each vintage of the BvD database we use.

1. Extract the data from Orbis with the 5 latest “relative years” in ASCII comma-separated value format,<sup>29</sup> transform it to Stata using Stattransfer, name and label the variables using the database codes and names. The data comes in Stata wide format

---

<sup>28</sup>Because all the variables were available in older disks under Formatted Export, we believe the reason they are not under Custom Export (which is a sub-heading of Formatted Export) is due to space limitations in the newer disks due to the increase in the amount of information given more years.

<sup>29</sup>For Amadeus data, we select 10 relative years for the reasons explained in Section A.2.



with the rows consisting of *unique* records (lines) per company consisting of BvD ID and similar non-time varying data, the last available year per firm `LASTYEAR`; and the blocks of time-varying variables in the form `X` in the Latest Year, `X` in Latest Year  $-1$ , and so on, until `X` in the Latest Year  $-4$ . We rename the variables using internal BvD codes to become, correspondingly, `X1`, ..., `X5`.

2. Delete the observations with just a name of company and no other information and the observations with missing BvD ID or BvD Account number (the main account identifier).
3. Notice that at this stage the data does not have a harmonized time variable showing the precise calendar year of each data point. In order to reshape the data from Stata wide format to the convenient long format, we use the row number as the cross-section identifier and specify in Stata that the number following the stub `X` of the time-varying variables `X1`, ..., `X5` has to be treated as the “fake” time identifier `YEAR`. The resulting `YEAR` variable is a mere marker of the observation per firm because it does not account for the possible gaps in reporting that we have discussed in Section A.3.1. There are also a lot of redundant empty observations because the `-reshape-` command creates a full panel and the observations in the periods when a given firm does not report the data will be empty.
4. Use the variable “Account Closing Date” `CLOSEDATE` to correctly assign the calendar year. We re-construct the `YEAR` variable based on the following convention. If the closing date is after or on June 1st, the current year is assigned (if `CLOSEDATE` is 4th of August, 2003, the year is 2003). Otherwise, the previous year is assigned (if `CLOSEDATE` is 25th of May, 2003, the year is 2002).<sup>30,31</sup>

---

<sup>30</sup>The raw data may contain duplicates in terms of ID-YEAR. These duplicates arise for two reasons: i) The presence of both quarterly and annual reports. ii) Firms switching from presenting their end of accounting year balance sheet information in one month to some other month (from December to May, for example). We would like to keep the most recent reports for these companies. One way to handle duplicates is to retain the data for the closing date closest conceptually to the end of year. An additional step to handle duplicates is to use a flow variable with good coverage such as Operating Revenue to identify potential quarterly reports (quarterly sales should be much lower than yearly sales). Then we drop duplicates whose revenue are less than the maximum per firm-year. For example, in 2005 vintage, there are around 34 thousand duplicates like that out of over 18 million observations. We eliminate a small number of remaining duplicates for which we cannot know whether these remaining reports refer to annual or monthly data.

<sup>31</sup>One might think that BvD correctly marks the lags in relative years taking into account the possible

5. Create our main company identifier `ID_NUMBER`, which is a copy of the BvD ID number. The difference between the BvD account number and the BvD ID number is the single letter, U or C, in the end of the BvD account number, reflecting what account type the record represents (consolidated and unconsolidated). If the BvD account number is missing, we concatenate the BvD ID number with the first letter of non-missing BvD Consolidation Code, following the BvD convention for the BvD account numbers (U, C and nothing for limited financials accounts). We create the country code based on the first two letters of the `ID_NUMBER` which by BvD convention starts by a two-letter country code (BE for Belgium, US for the U.S., GB for the UK, and so on).
6. All the financial variables are transformed from strings into numeric data type. In the vintages where we downloaded the data in the “original units,” we transform the variable `UNITS` from textual form to the integer power of ten to express all financials in the same units. We developed a filter to check for mistakes in the `UNITS` variables as explained below.
7. Clean the raw vintage data.
  - Drop duplicates in terms of `ID_NUMBER` and `YEAR`.<sup>32</sup>
  - Drop observations with no financial information. We verified that these company-year points do not have an account closing date and are mostly missing observations generated during the reshaping stage.
  - Drop observations for which country code created based on the BvD ID numbers does not correspond to BvD’s country ISO code.<sup>33</sup>
  - Drop observations with missing currency.

---

time gaps in financial data. As we discovered, this is not the case. Hence it is essential that one downloads the `CLOSEDATE` variable, together with BvD ID number, BvD account number, and the last available year variable.

<sup>32</sup>These will have the same financial data but different industry codes. In Orbis the first observation per duplicate represents the main industry, hence, we keep the main industry per company for companies that had multiple industries reported.

<sup>33</sup>As we mentioned, BvD IDs start from the 2-letter code corresponding to the company’s country. BvD country ISO code is the same for all the companies in a given country. Hence, if a company has a different country code, we interpret this as a mistake. This can also be a tax front. Note that these are not the multinationals since the country code attached to BvD ID numbers reflect where the company operates. If the company is part of a multinational, this will be reflected in the ownership variables not in the ID numbers.

- Drop observations with missing Account Closing Date.

### A.5.2 Merging Vintages

At this stage each observation in each vintage is uniquely identified by the variables `ID_NUMBER` and `YEAR`. The following steps are performed at each consecutive merge, with vintage specifics highlighted.

1. As explained in Section A.1.2, in some instances the unique company identifiers change over time which would affect the success of the merge. Before the merge, we replace the old IDs for the countries which had their ID changed over time with the most recent ones. We retain the old “legacy IDs” on the data.

A slight complication emerges for IDs in former Yugoslavia, coming from earlier Orbis disks. The same companies “belong” to Yugoslavia in the early 2000s (country code and the first two letters of IDs are YU), then to Serbia-Montenegro (CS) in disks from the mid-2000s, then to, separately, Serbia (RS) and Montenegro (ME), or even Kosovo (KV). In the IDs of these companies only the letter part changes over years—the numeric part does not. Where it is possible, we assign YU and CS to RS and ME using the numeric part of IDs; where we do not know, we assume they are in RS. We do this after the replacement of IDs using the correspondence table, because some firms with ID from CS are present in that table and are already assigned to RS or ME.

2. Before the merge, we harmonize the names of the variables across vintages and products.
3. Before the merge, we check the consistency of the Units of Financials variable (`UNITS`). Recall that at this stage the financial data is reported in various units of local currency (units, thousands, etc.) for different companies and even for the same company over time. This is due to BvD non-harmonization across products. Downloading in “default” option for the units requires the researchers to use the `UNITS` variable to harmonize the units. Another way is to force the download to be in a particular unit such as units or thousands. We employ the following strategy to deal with this issue if the researcher downloads the data with the “default” option. For each company we

check if the moment of switch in units coincides with a “reasonable” move of total assets (can also do other financial variables); if not—we drop the entire firm. For the lower threshold of assets growth, we choose -99% because the 1000x decrease (due to a change in the UNITS variable) of otherwise unchanged assets is -99.9% growth. For the upper bound of assets growth, we choose 19,800% because the 1000x increase of otherwise unchanged assets is 99900% growth. If we allow the company to have a large (70 percent) drop in assets in the year when the units switch 1000x ( $x \cdot 0.3 \cdot 1000$ ), this is close to 19900% growth.<sup>34</sup>

Then we calculate the actual growth of assets and verify if this growth lies outside of these thresholds *in the years when UNITS change*. We drop the entire company where the moment of switch in UNITS does not coincide with the reasonable move in assets.<sup>35</sup> One can keep the companies which are marked by BvD as “inactive” because the assets of those firms can genuinely go down to (almost) zero.<sup>36</sup>

4. Before the merge, we express all financial variables in harmonized units by the transformation  $x = x \cdot 10^{\text{UNITS}}$ , except in the 2013 Orbis vintage that is downloaded with the explicit “in units of local currency” option. The UNITS variable is then discarded.
5. Before the merge, we create the textual identifier for the vintage to keep track of the vintage which contributes to a given observation after all merging steps are done.
6. An additional methodological complication arises when we combine the 2013 Orbis vintage with earlier vintages. This is related to the change in sectoral classification in 2008 from NACE Rev 1.1 to NACE Rev 2 by the Eurostat.<sup>37</sup> In 2005 and 2009 Orbis vintages, the sector is identified by the “NACE Rev. 1.1, Core code (4 digits)” (NACECD); in 2010 WRDS Amadeus, 2013 Orbis, and 2014 Amadeus vintages it is already “Rev.

---

<sup>34</sup>Recall that all the balance sheet values are book values for the non-listed firms.

<sup>35</sup>As a result, we eliminate about 3% of observations in the 2005 and 2009 Orbis vintages; less than 1% in the 2010 WRDS Amadeus vintage; and less than 0.5% in the 2014 Amadeus vintage.

<sup>36</sup>The variable Status takes the values Dissolved, Dissolved (merger or take-over), In liquidation, Inactive (no precision), Dissolved (liquidation), Dissolved (merger or take-over), Inactive (no precision), Bankruptcy, Dissolved (bankruptcy), Dissolved (demerger) or some peculiar active types Active (default of payments), Active (dormant), Unknown.

<sup>37</sup>Table A.6.2 lists sectors classified by NACE Rev 2, Level 2. For space considerations we do not report the 4-digit industry classification.

2 Core code (4 digits)” (NACEREV2CCODE). Both codes are unique per company.<sup>38</sup> We prefer to use the more recent NACE Rev. 2 classification in our data.

We went through a detailed process of matching the industry classifications pre- and post-2008. We start from the official Eurostat correspondence table between NACE Rev 1.1 and NACE Rev 2. To supplement the official correspondence table in a way that we have a one-to-one match for every sector, we proceed as follows. Most often, multiple NACE Rev 2 codes correspond to a given NACE Rev 1.1 code. In the official correspondence tables, when multiple NACE Rev 2 codes are matched to a unique NACE Rev 1.1, they are sorted in the ascending order of the numeric NACE Rev 2 code. The first code is the *most closely related sector* to the one in NACE Rev 1.1 classification. We retain the first NACE Rev 2 code provided in the official table and discard the rest.<sup>39</sup> This approach attains a good match for manufacturing sectors (codes 10 and higher in NACE Rev 1.1) but not as good a match for agricultural sectors (codes below 10 in NACE Rev 1.1). We manually match codes by reading the long descriptions of the codes. We do the same if there are sectors that are completely missing in the official correspondence tables. Our own correspondence table is available upon request.

Before merging with the 2013 Orbis vintage we merge the earlier vintage data with our sector correspondence table. We keep the original sector classification from each vintage just in case.

7. Care should be exercised when combining the disk data from Orbis with Amadeus downloads from the WRDS (in our case, we used the 2010 download of Amadeus from the WRDS), mostly with respect to the harmonization of variable names. Here are some peculiarities which the users are advised to verify. i) As we explain above, the “Exchange Rate” variable of WRDS should not be used. ii) In May 2010 WRDS Amadeus , the label of the core sector variable NACECD states NACE Revision 1.1. but we dis-

---

<sup>38</sup>There are numerous “secondary” codes in BvD. All of the above is for the “main” code. We also keep the alternative unique industry codes such as SIC and NAICS.

<sup>39</sup>For example, the NACE Rev 1.1. code 10.20: Mining and agglomeration of lignite is matched to three NACE Rev 2 codes: 05.20: Mining of lignite; 09.90: Support activities for other mining and quarrying; 19.20: Manufacture of refined petroleum products. We retain the first line from the correspondence table and matches “10.20: Mining and agglomeration of lignite” to “05.20: Mining of lignite.”

covered from observing the values that it is in fact Revision 2. We verified this with WRDS, who in turn confirmed this by contacting BvD. For this reason, no sector correspondence table needs to be applied to this vintage. iii) In WRDS, the listed company identifier is a binary variable (“Quoted company” LISTED) while in Orbis it is textual having the values like “Delisted,” “Listed,” “Unlisted.” We add the textual variable to WRDS Amadeus vintage before merging with the other data.

8. We use the values from the later vintages to supplement missing values. A non-missing value, however, will never be replaced with a missing. (In Stata language, we merge with update and replace options). Depending on the order of merge and the computing power (RAM) availability the users may merge all the vintages at once or, if the data gets big to fit the RAM, merge countries one-by-one and then append (stack) the country data.
9. Check for duplicates by ID and YEAR and, in case of duplicates, retain only observations coming from the most recent vintage.

### A.5.3 Cleaning Merged Data

After we merge the financial data from individual vintages, we perform some further data harmonization and very mild cleaning for obvious data mistakes.

1. Companies in several countries report financials in multiple currencies. We always retain the accounts in major currencies, such as, U.S. dollar, Euro, UK Pound, but delete the observations with missing or unreasonable currencies which probably are mistakes (for example South African Rand or Canadian dollar for European companies).
2. We express the financial variables in real dollars 2005 base. To convert from the units of the nominal currency of accounts we i) convert the currency of accounts to the official currency of the country; ii) deflate the series by the national GDP deflator with the 2005 base from the World Bank; and iii) divide by the exchange rate of the official currency to the U.S. dollar in the year 2005. A number of complications arise at this stage.

Because companies in several countries report in multiple currencies in order to add the official exchange rate, we do not use the country code but the currency code.<sup>40</sup> The problem with the World Bank data (or the IMF’s IFS data) is that the source does not report the ISO currency code even if the country changes the currency; we just observe a jump in the exchange rate. We obtain the Compustat Global exchange rates, which use currency ISO symbols (USD, GBP, etc.) as the main identifier of the existing and legacy currencies. We also supplement and harmonize the Compustat currency series rates with currencies actually observed in our data in particular country-years. In case the legacy currencies are missing in Compustat in some years (e.g., in the former Soviet Union, in Yugoslavia, etc.), we refer to the national central banks’ data.

For the conversion step i), we need to make sure that the currency of financials for all firm-years is the “official” local currency as of today and not other currencies. This is important because we will use the deflator in the official local currency. In particular, we need to decide what official currency to use with the recent Eurozone members in conjunction with the GDP deflator data. In our sample, Estonia, Slovenia, Slovakia, and Malta are such countries; Latvia who adopted the euro on 1 January 2014 is still marked as having currency as Latvian Lats in our data. Consider Slovenia as an example. The country adopted the euro on 1 January 2007. In Slovenian data we observe the companies which report the data in single currency (Euro or Slovenian Tolar) in all years, pre- and post-2007. The note to the WB GDP deflator data for Slovenia says “A simple multiplier is used to convert the national currencies of EMU members to euros. The following irrevocable euro conversion rate entered into force on January 1, 2007: 1 euro = 239.64 Slovenian Tolar.” This implies that the deflator is effectively in Tolar (SIT) until 2007 because it is a ratio of year 2006 to year 2005 (the base year in WB data) and both of these years are multiplied by the same number 239.64 (the fixed parity rate). Hence, the official currency of Slovenia is considered to be SIT before 2007, and all financial data is recalculated to SIT before 2007. From 2007 the deflator reflects the dynamics of local prices in Euros, and we express all the

---

<sup>40</sup>As we explained in Section A.3, the BvD variable “Exchange Rate” is useless for conversion of the data downloaded in the “original currency.” Users may choose to download the raw data in some other currency, such as U.S. dollars, in which case the Exchange Rate might report the exchange rate to the currency of accounts. We prefer to convert the data ourselves and not rely on the internal BvD data convertor.

financials of Slovenia in euro.

3. Drop company-years with missing information on total assets and operating revenue and sales and employment (simultaneously).
4. Drop the entire company (all years) if total assets is negative in any year.
5. Drop the entire company if employment (in persons) is negative in any year and companies with employment larger than that of Walmart (2 million) in any year.
6. Drop the entire company if sales are negative in any year. Of note, we do not perform this filter in terms of Operating Revenue because this P&L account item is equal to sales + Other operating revenues + Stock variations. While sales cannot be negative, revenue can be negative if a company has a sizable financial loss (say, loss due to hedging, etc.). For countries, like Denmark, whose firms do not report sales but only operating revenue, we cannot use this filter.
7. Drop the entire company when reporting in any year a value of employment per million of total assets larger than the 99.9 percentile of the distribution.
8. Drop the entire company when reporting in any year a value of employment per million of sales larger than the 99.9 percentile of the distribution.
9. Drop the entire company when reporting in any year a value of sales to total assets larger than the 99.9 percentile of the distribution.
10. Drop the entire company if Tangible Fixed Assets (such as buildings, machinery, etc.) is negative in any year.
11. For a given company ID year, we replace missing strings which are unlikely to change over time with values for this company for other years. We complement information on country, company name, city, region, postal code, legal form, and date of incorporation with lagged/lead values in the years where such info is present. This is reasonable because if a company changes the legal form it obtains a new BvD ID and will be treated as a new entity. If information is missing in all years, they remain missing.



## A.6 Filing Requirements and Sector Correspondence

Table A.6.1: BvD COMPANY FILING REQUIREMENTS AND DATA PROVIDERS FOR SELECTED COUNTRIES

COUNTRY CODE	WHICH COMPANIES HAVE TO FILE ACCOUNTS?	HOW MANY COMPANIES DOES THAT REPRESENT?	DATA PROVIDER
AT	AG, starting 1994 also GmbH and very large companies. Based on their size, companies may file shortened balance sheet and no PL account.	50,000	Creditreform
BE	Depends on the legal form: <ul style="list-style-type: none"> <li>Companies that must file their accounts are: SA; SPRL; SCRL (Société coopérative à responsabilité limitée); SE (Société européenne); GEIE (Groupement européen d'intérêt économique); GIE (Groupement d'intérêt économique); Foreign companies located in Belgium.</li> <li>Companies that have to file their accounts under certain conditions are: SCS (Société en commandite simple) if the company is large and one of the associates is an individual; SCRI (société coopérative à responsabilité illimitée) if the company is large and one of the associates is an individual; SNC (société en nom collectif) if the company is large and one of the associates is an individual; ASBL and Foundations if they are large or very large; Other (there are some other specific cases).</li> </ul>	420,000	National Bank of Belgium, Coface Services Belgium
BG	All companies, which match 2 of the following 3 criteria: at least 50 persons staff, total assets at least eur 500.000, turnover at least eur 1.000.000	Less 10% of all active companies	Creditreform
HR	Private and public limited liability companies, general and limited partnerships, cooperatives have to file accounts to the State Authorities (State Register of accounts, established 2003).	Approximately 100,000 legal subjects filed their accounts for 2012, although there are more than 300,000 registered subjects	Creditreform
CY	All Cypriot Companies, whether local or international, must maintain accurate books of accounts, which should reflect the true and correct position of their conduct, as well as give adequate explanation of their operations. Audited financial statements and an Income Tax Return are required for all companies, even companies with no taxable income and/or dormant companies. Registered Branches (in Cyprus) of foreign companies are not legally bound to compile full separate branch accounts however when taxed on the island, are obliged to do so for income tax purposes. The following types of companies are obliged to file their financial statements: <ol style="list-style-type: none"> <li><i>Limited Liability Companies.</i> They are obliged to submit a copy of their annual report, including their audited, financial statements for the year</li> <li><i>Public Companies.</i> They are obliged to submit a copy of their annual report, including their audited, financial statements for the year. In addition, public companies that are listed in the Cyprus Stock Exchange are obliged to publish their quarterly financial statements as well.</li> </ol> Partnerships are exempt from any requirement to prepare audited accounts, but they are legally bound to keep proper books of account which must be available for scrutiny by individual partners.	This represents all the active companies operating in Cyprus (approximately 90.000 companies). However, this does not include the International Business Companies (IBC's) which are registered in Cyprus for taxation purposes and are actually operating abroad as: 1. There is no clear indication of which companies are actually IBC companies. 2. The vast majority of these companies avoid submitting their annual financial statements	Infocredit Group
CZ	All companies filled in Business Register. There is duty to fill full financial statements for companies with obligatory audit. Other companies in BR have duty to fill shortened FS. Obligatory audit refer joint stock companies (A.S.) and Limited liab. companies (S.R.O.) od cooperatives with (1) assets over 40 mil. CZK, (2) turnover 80 mil. CZK, (3), over 50 empl. For A.S. if is fulfilled any condition, for other companies fulfilled two of three conditions.	330,000 (many companies could be inactive, only registered with no financial statement). There is currently 330.000 companies in BR, 10% in liquidation or bankruptcy, and about 25% (80.000) companies registered but without turnover, not registered as VAT payers, etc. so suspicious.	Soliditet- main source is Business register where companies publish FS. They permanently go through the list of all companies in BR and seek newly published FS. This source is updated daily, but some companies put documents into BR with big delays.

Notes: Filing requirements were taken from the Orbis Online Manual on February 3d, 2014.

Table A.6.1 (Cont'd.): FILING REQUIREMENTS AND DATA PROVIDERS

COUNTRY CODE	WHICH COMPANIES HAVE TO FILE ACCOUNTS?	HOW MANY COMPANIES DOES THAT REPRESENT?	DATA PROVIDER
EE	Private limited companies, joint-stock companies, non-profitmaking associations, cooperative societies, general partnerships, limited partnerships, foundations.	C.a. 124,000 – 125,000	Kreditdiinfo AS
FI	All joint-stock companies and all co-operatives; Limited partnerships, partnerships and private firms, which meet two of the following three conditions: <ul style="list-style-type: none"> <li>• turnover over 7.30 million EUR;</li> <li>• balance sheet total over 3.65 million EUR;</li> <li>• number of personnel over 50.</li> </ul>	The exact number is not known, but the estimate is approx. 120,000 companies	Suomen Asiakastieto Oy
FR	All of the following: <ul style="list-style-type: none"> <li>• les sociétés à responsabilité limitée (SARL et EURL) ;</li> <li>• les sociétés de personnes (sociétés en nom collectif et sociétés en commandite simple), sous certaines conditions : les sociétés en nom collectif (SNC) dont au moins l'un des associés est une personne physique ne sont pas dans l'obligation de déposer leurs comptes annuels (pour plus de précisions, se référer à l'article L. 232-21 du Code de Commerce) ;</li> <li>• les sociétés par actions (sociétés anonymes, sociétés par actions simplifiées et sociétés en commandite par actions) ;</li> <li>• les sociétés commerciales dont le siège est situé à l'étranger qui ont ouvert un ou plusieurs établissements en France ;</li> <li>• les sociétés d'exercice libéral (SELARL, SELAFA, SELCA, SELAS) ;</li> <li>• les sociétés coopératives et unions sous certaines conditions (pour plus de précisions, se référer à l'article R. 524-22-1 du Code Rural).</li> </ul>	1,400,000	Ellisphere
DE	Corporate enterprises (AG, GmbH) and cooperatives (e.G).	Breakdown: <ul style="list-style-type: none"> <li>• <i>small cooperate enterprises</i>: approx. 980.000. Definition ( 267 HGB): staff: ≤ 50 individuals turnover: ≤ 9.680 TEUR total assets: ≤ 4.840 TEUR at least two criteria must apply They have to announce only the balance sheet information and the notes on the accounts.</li> <li>• <i>medium sized cooperate enterprises</i>: approx. 80.000. Definition ( 267 HGB): staff: between 50 and 250 individuals turnover: between 9.680 TEUR and 38.500 TEUR total assets: between 4.840 TEUR and 19.250 TEUR at least two criteria must apply They have to announce the balance sheet information as well as the statement of income and the notes on the accounts.</li> <li>• <i>big cooperate enterprises</i>: approx. 33.000. Definition ( 267 HGB): staff: more then 250 individuals turnover: more then 38.500 TEUR total assets: more then 19.250 TEUR at least two criteria must apply They have to announce the balance sheet information as well as the statement of income and the notes on the accounts.</li> <li>• <i>cooperatives</i>: approx. 7.500. They have to announce the balance sheet information as well as the statement of income and the notes on the accounts.</li> </ul>	Creditreform and Creditreform Rating AG

Table A.6.1 (Cont'd.): FILING REQUIREMENTS AND DATA PROVIDERS

COUNTRY CODE	WHICH COMPANIES HAVE TO FILE ACCOUNTS?	HOW MANY COMPANIES DOES THAT REPRESENT?	DATA PROVIDER
GB	Limited, PLC, LLP, LP.	1,000,000	Jordans Limited
GR	Societe Anonyme and Limited Liability Companies.	The Societe Anonyme and Limited Liability Companies that publish Balance Sheets represent approximately the 5% of the total active Business Universe in Greece.	ICAP
HU	All companies have to file accounts, except private enterprises. The companies have to send the accounts to the Ministry of Justice and to the Registry Court. The one-person firms and the limited deposit companies do not have to send it to the Ministry of Justice.	About 40%.	Creditreform
IS	Sameignarfélag (fulfil two out of the three following prerequisites : total assets > 230 ML ISK, operating revenue > 460 ML ISK, average number of employees >50, or if the mother company has to file accounts), Samvinnufélag, Samlagsfélag (if mother company has to file accounts), Einkahlutafélag, Hlutafélag.	Approx. 20,000 companies were to file their accounts.	Icecredit Info.
IE	Limited.	100,000	Jordan Limited
IT	Includes: <ul style="list-style-type: none"> <li>• S.p.A. (Società per Azioni),</li> <li>• S.r.l. (Società a responsabilità limitata),</li> <li>• Sapa (Società in accomandita per azioni),</li> <li>• Società Cooperative,</li> <li>• Società Consortili,</li> <li>• G.e.i.e, Società di persone (only consolidated accounts),</li> <li>• Consorzi con qualifica di Confidi.</li> <li>• Società a responsabilità a socio unico e società per azioni a socio unico.</li> </ul>	Approximately 900,000	
LV	All companies, except sole proprietor enterprises, peasant farms and fishers farm whose annual turnover does not exceed LVL 200.000 (EUR 284.6 thousand).	Approximately 100,000.	Creditreform
LT	Includes: <ul style="list-style-type: none"> <li>• Limited liability companies;</li> <li>• Joint stock companies;</li> <li>• State enterprises;</li> <li>• Municipal enterprises;</li> <li>• Agricultural companies;</li> <li>• Cooperative companies.</li> </ul>	79,823 (excluding bankrupted, liquidating and inactive companies).	Creditreform
LU	Public (S.A.), limited (S.A.R.L.)	Not possible to find out : there are in total about 25,000 companies and 15,000 Holdings, total 40,000.	Creditreform

Table A.6.1 (Cont'd.): FILING REQUIREMENTS AND DATA PROVIDERS

COUNTRY CODE	WHICH COMPANIES HAVE TO FILE ACCOUNTS?	HOW MANY COMPANIES DOES THAT REPRESENT?	DATA PROVIDER
NL	All limited companies (B.V.s and N.V.s) and some sole traders and cooperations.	680,000	LexisNexis Benelux, Graydon and Chambers of Commerce
NO	Limited.	Approximately 120,000.	Creditreform
PL	<i>Filing of the accounts:</i> All companies registered at the National Court Registry (KRS): joint-stock companies (S.A.), limited liability companies (Sp. z o.o.), cooperatives, state enterprises, etc., except for general partnership (sp.j.), professional partnership (sp.p.) that do not reach the annual turnover of 800,000 EUR. <i>Publishing of the accounts:</i> Joint-stock companies, banks, insurance companies, investment funds, plus others (limited liability companies, cooperatives, large private companies, etc.) complying with 2 of the following criteria: i) average annual employment > 50; ii) total assets at the end of a financial year > 2.5 million euro; iii) annual net profit > 5 million euro.	About 200,000 companies. Among these companies, approx. 10,000 companies are also obliged to publish accounts.	InfoCredit, collects information from National Court Registry, journals for entities and cooperatives, Judicial and Business Journal ("Monitor Sdowy i Gospodarczy," directly from the companies, as well as other alternative sources, if necessary). Coface MOPE.
PT	The Portuguese law compels all the companies to deposit the Balance Sheet. Therefore, about 57,500 balance sheets are public information.	These 57,500 balance sheets correspond to a small percentage of the totality of Portuguese companies, if we verify that the totality number of the universe has about 800,000 companies. That means that practically the companies do not deliver financial information.	
RO	Joint stock companies, partnerships limited by shares, limited liability companies, state owned concerns, co-operative companies.	500,000	Chamber of Commerce and Industry of Romania
RU	All juridical persons have to represent their accounts (individual entrepreneurs (manufacturers) and farms are not juridical persons)	Approximately 40% of all active companies file their accounts. So, if there are 1,500,000 registered active companies in Russia, the accounts are available for approximately 800,000 companies. Most of these are included in Russia.	
SK	All companies filed in Business Register.	About 70-80,000 (many companies could be inactive, only registered with no financial statement). Many companies do not file their accounts, there are no penalties for it.	Soliditet, s.r.o.
SI	All companies and sole proprietors	Around 160,000 companies.	Coface Slovenia
ES	S.A., S.L.	776,000	INFORMA
SE	Limited companies.	About 348,000.	UC AB
CH	There are no legal requirements to file their accounts in Switzerland (like in the UK). Public quoted companies make their financial statements available and all are collected, analysed and provided to the Orbis database.	N/A	Worldbox AG
UA	All local companies and trade or representative offices of foreign companies have to file accounts (except companies mentioned in 3.) <i>Notes:</i> In the database of the Central Statistical Department there are about 330,000 accounts of the biggest Ukrainian companies. The remaining companies are split in 800 regional Departments of Statistics: there are currently no means by which we can get them for the time being.	There are more than 1,000,000 companies and trade representative offices registered.	Creditreform

Table A.6.2: NACE Revision 2, Level 2 Classification.

Code	Name of the Level 2 NACE sector
01	Crop and animal production, hunting and related service activities
02	Forestry and logging
03	Fishing and aquaculture
05	Mining of coal and lignite
06	Extraction of crude petroleum and natural gas
07	Mining of metal ores
08	Other mining and quarrying
09	Mining support service activities
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and of products of wood and cork, except furniture, etc.
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
22	Manufacture of rubber and plastic products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metals
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment n.e.c.
29	Manufacture of motor vehicles, trailers and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment
35	Electricity, gas, steam and air conditioning supply
36	Water collection, treatment and supply
37	Sewerage
38	Waste collection, treatment and disposal activities; materials recovery
39	Remediation activities and other waste management services
41	Construction of buildings
42	Civil engineering
43	Specialised construction activities
45	Wholesale and retail trade and repair of motor vehicles and motorcycles
46	Wholesale trade, except of motor vehicles and motorcycles
47	Retail trade, except of motor vehicles and motorcycles
49	Land transport and transport via pipelines
50	Water transport
51	Air transport
52	Warehousing and support activities for transportation
53	Postal and courier activities
55	Accommodation
56	Food and beverage service activities
58	Publishing activities
59	Motion picture, video and television programme production, sound recording and music publishing
60	Programming and broadcasting activities
61	Telecommunications
62	Computer programming, consultancy and related activities
63	Information service activities
64	Financial service activities, except insurance and pension funding
65	Insurance, reinsurance and pension funding, except compulsory social security
66	Activities auxiliary to financial services and insurance activities
68	Real estate activities
69	Legal and accounting activities
70	Activities of head offices; management consultancy activities
71	Architectural and engineering activities; technical testing and analysis
72	Scientific research and development
73	Advertising and market research
74	Other professional, scientific and technical activities
75	Veterinary activities
77	Rental and leasing activities
78	Employment activities
79	Travel agency, tour operator and other reservation service and related activities
80	Security and investigation activities
81	Services to buildings and landscape activities
82	Office administrative, office support and other business support activities
84	Public administration and defence; compulsory social security
85	Education
86	Human health activities
87	Residential care activities
88	Social work activities without accommodation
90	Creative, arts and entertainment activities
91	Libraries, archives, museums and other cultural activities
92	Gambling and betting activities
93	Sports activities and amusement and recreation activities
94	Activities of membership organizations
95	Repair of computers and personal and household goods
96	Other personal service activities
97	Activities of households as employers of domestic personnel
98	Undifferentiated goods- and services-producing activities of private households for own use
99	Activities of extraterritorial organizations and bodies

## B Foreign Ownership

### B.1 Building Foreign Investment Panel Data

#### B.1.1 Vintage Raw Data

The following are the key steps we take in constructing the firm-level ownership database. In some respect, the process of preparing the raw vintage data is more straightforward than that for the financials because each vintage corresponds to a single time observation. However, compared to the financial data, the ownership data has more than two identifiers (the company ID and year). In the case of ownership, the additional dimension comes from the fact that each company could have multiple owners or subsidiaries.

The raw ASCII data has rows consisting of non-time varying data and the blocks of variables corresponding to the groups in Orbis Ownership database. They are company's shareholders, domestic ultimate owners (UOs), global UOs, and subsidiaries. In order to have the data for a given country in one file, we download data using the Formatted Export method in older disks and the List Export method in newer disks, as discussed in Section A.2. Of note, we download some variables describing the company itself such as name, BvD ID, size category, type of company, location information, core industry, and three key financials (employment, total assets, and operating revenue) in the original currency of accounts and in the last available year, together with Units and Currency Code in case we would need these to identify the company better. The key identifier is still the company BvD ID.

Assume a Company A has 2 shareholders, 1 global UO, and 3 subsidiaries. The data for this company will have 3 observations (rows) in a given vintage, corresponding to the largest number of observations across the groups in Orbis Ownership database:

NAME*	BvD ID*	Core Industry*	Shareholder Name	Shareholder % Stake	GUO Direct	GUO Name	GUO % Stake	Subsidiary Name	Subsidiary % Stake
Company A	ZZ1234U	6123	Company B	70		Company B	100	Company C	100
Company A	ZZ1234U	6123	Mr. Smith	30				Company D	90
Company A	ZZ1234U	6123						Company E	WO

The variables marked with (\*) are unique per company-vintage. Because we export the data asking the disk to repeat each single item all these variables will be repeated for each

company record in the raw downloads. The order of the non-unique variables corresponds to the internal order in the Orbis Ownership or Amadeus Ownership databases.

We execute the following steps for each vintage of the Orbis Ownership or Amadeus Ownership database we use.

1. We extract the data in ASCII comma-separated value format, transform it to Stata using Stattransfer, name and label the variables following the database names.
2. Delete observations with only a company name and no other information.
3. Generate the YEAR variable for the year in which the ownership information is recorded in a given vintage of BvD product following our assumption that YEAR takes the value of one less the year when the Orbis disk was issued. Recall that we choose the vintages that are closest, but subsequent to that year. For example, the ownership data coming from the Orbis disk no. 27 issued in January 2011 is assigned to YEAR 2010, and so on.
4. As in case of financial data, we create our main identifier ID\_NUMBER, which is a copy of the BvD account number. It is a copy of the BvD ID number if the BvD account number is missing.
5. To keep the file size manageable we split the data generating a separate file for a given country in a given year and the group of variables in Orbis Ownership database. As a result we have the following files per country-year.
  - In the file with the information about the company itself, we keep all the identifiers mentioned above, removing the duplicates in terms of all remaining variables and then removing duplicates by ID and Employment.
  - In the file with information on company shareholders, we keep ID, YEAR and all variables from this group, dropping observations with missing information on owners and then removing the duplicates in terms of all remaining variables.
  - We do the same in the file with separate information for company's immediate shareholders, for its domestic ultimate owners, for its global ultimate owners, and for its subsidiaries.

### B.1.2 Merging and Cleaning Ownership Data

In this section, we illustrate how we create the country panels of company shareholder data; that is, the information on the direct ownership of the equity rights for the universe of companies covered in the Orbis Ownership data set. The preparation of other ownership variables involves similar steps. We first describe how we prepare the company panels of “direct ownership links;” that is, the data where the unit of observation is the company-shareholder pairs in a given year. After that, we discuss how we build the company-year level data where the links information is aggregated to the company level each year.

### B.1.3 Data with Company Shareholder Links

Each of the following steps is performed for individual country ownership files coming either from the Orbis Ownership or the Amadeus Ownership database.

1. Combine (stack) all annual ownership data files for a given country. Recall that we have bi-annual vintages of Amadeus Ownership since 2000 and annual vintages of Orbis Ownership since 2005.
2. Create the country code based on the first two letters of the `ID_NUMBER`, which by BvD convention should start from two-letter country code (BE for Belgium, US for the United States, GB for the UK and so on).
3. Convert the character variables Percentage Owned Direct (`ODIRECT`) and Percentage Owned Total (`OTOTAL`) into numeric format, replacing some special character values they may take in the raw data. In particular, we replace percentage with a leading `<`, `>`, `±` with the percentage after the symbol; eliminate possible `%` sign; replace special codes “WO” (wholly owned) with 100%, “MO” (majority owned) with 50.01% (because by the GAAP practice the majority ownership involves 50% plus one share but the smallest stakes reported by BvD are 0.01%), “CQP1” (50% plus 1 share) with 50.01%, “NG” (negligible) with 0.01% (again, the smallest observed stake according to BvD), “-” (not significant) or “n.a.” (not available) with missing; “BR” (branch, Orbis Ownership only) with 100%; “JO” (jointly owned, Amadeus Ownership only)



with 50% (our exploration of such cases shows that there is always exactly two owners in case of the JO code). We keep the character versions of these variables.

4. Implement company ID changes following the procedure described in Section A.5.2 for financial data, saving the legacy IDs on the dataset.
5. To determine the total foreign ownership at the company level, we proceed as follows. Whenever the variable Shareholder Country ISO code (**OCOUNTRY** in Amadeus or **SHARCOUN** in Orbis) is different from the company own Country code we consider the link foreign. By default, we assume that the shareholder with missing country code or with **SHARCOUN** taking the values of “-” and “n.a.” is located in the same country as the given company as it is done in the literature.
6. For Amadeus Ownership, we further improve the above simple rule by manual assignments by country, based on the variable Shareholder Name (**ONAME**). That variable contains some indication of whether the owner is foreign.<sup>41</sup> We determine the unique values of **ONAME** and then manually replace the ownership links which have the missing **OCOUNTRY**. One can further try to compare the company country code to the first two letters of the available Shareholder BvDEP ID number (**OID** in Amadeus or **SHARSIR** in Orbis), but not every shareholder has the BvD ID available. In addition, one can investigate the values of the variable Shareholder Type (the textual variable **OTYPE** in Amadeus or the standardized letter code **SHARTYPE** in Orbis), assuming that certain owner types are domestic (such as company employees or management, or the entries like “private individual(s)” or “unnamed private shareholders”, and so on) unless it is clear from the name they are foreign (for example, “foreign investors”).
7. Combine shareholder information from our Amadeus Ownership and Orbis Ownership, making sure we do not have duplicates in overlapping years. In overlapping years, we establish which database has more recorded shareholders with non-missing direct ownership stakes for a given company-year (non-missing variable Shareholder - Percentage of ownership, direct, **SHARDPER** and use the record with more data.<sup>42</sup> In case of an ex-

---

<sup>41</sup>We make this assignment mostly based on the parts of the names reflecting the legal type of the shareholder. We compare that to what is typical for the company’s own country.

<sup>42</sup>We encountered the case when the same observation (ownership links) got the value “-” in Amadeus

act match, we use Orbis but retain the variable Shareholder NACE Rev. 2 Core Code (ONACE) and other useful information available only in Amadeus. Actual elimination of the duplicates is done later, because we use some information across two products in a given year to supplement the data.

8. Supplement the missing textual information (Shareholder Type in textual form from Amadeus and 1-letter code from Orbis; Shareholder NACE Rev. 2 Core Code) using the unique shareholder IDs and shareholder names. Supplement missing Shareholder Type in textual form from Amadeus using non-missing 1-letter code from Orbis and vice versa.
9. Because we first establish the foreign ownership link status at the individual vintage level, we can take advantage of the full panel and revise the foreign link status based on *other* years when this info is available for a given company ID. As above, using the unique shareholder IDs and shareholder names, we replace links ever found to be as foreign in at least one year as such in all the years.
10. After all the information across products has been used, we delete the duplicated observations (all the links for a given company-year) from Orbis and Amadeus keeping the ones with the best coverage. We save the data files of direct ownership links for individual countries.

#### B.1.4 Direct Ownership Types

Once we have the shareholder links data for country-years, we can aggregate it in the variety of ways because each link record has a number of variables describing not only the equity stake held by a given (direct) investor but also investor’s location, type, or industry.

For illustration, we will discuss our work to aggregate ownership stakes by foreign/domestic status, further split by the type of the shareholder involved. In particular, this data can be

---

Ownership and the value of exactly 0% in Orbis Ownership. We think that in Amadeus Ownership, the code “-” actually means “negligible” while it means “missing” in Orbis Ownership. Because we could not confirm or refute this distinction, we recoded “-” as missing. Because the 0% stakes do not bear any useful information, we treat the observations with Shareholder Percentage=0% as missing when counting the shareholders with available ownership stake.

merged with the financials data panel, described in part A.5, also uniquely identified by the company ID (IDNUMBER) and YEAR. Each of the following steps is performed over individual country data files.

1. We start from the ownership links data obtained as described in section B.1.3 and delete the link records which have no usable information.
2. Identify foreign and domestic links, specific to the owners of a particular type. The exercise is similar to what we did to find foreign/domestic links, but here we use the information on whether a link is foreign or domestic and, in addition, take advantage of the information in the variable Shareholder Type (the textual variable OTYPE in Amadeus or the standardized letter code SHARTYPE in Orbis). Recall that in the codes preparing raw links data we supplemented the values of both variables with one another and across all the years. In particular we define the following indicators:
  - *Foreign (Domestic) Owner-Industrial Type*, =1 if Foreign (Domestic) owner has the (textual) type Industrial company, Corporate, Self-owned, Branch, or types reflecting the individuals working for the company (such as, employees, personnel, managers, directors, self-ownership) because these owners are likely to bring similar types of “expertise” as industrial owners for the majority of the companies in our financials database.<sup>43</sup>
  - *Foreign (Domestic) Owner-Financial Type*, =1 if Foreign (Domestic) owner has the type Bank, Financial company, Insurance company, Other financial institution, Mutual & Pension Fund/Nominee/Trust/Trustee, Foundation/Research Institute, Private Equity firms, Venture capital, Hedge funds.
  - *Foreign (Domestic) Owner-Government Type*, =1 if Foreign (Domestic) owner has type Governments, State, Public authority.
  - *Foreign (Domestic) Owner-Individuals Type*, =1 Foreign (Domestic) owner has the “individuals” type with known names. Besides single private individuals or families, this category includes shareholders designated by more than one *named*

---

<sup>43</sup>Orbis has some companies in the financial intermediation and insurance sector. Our assumption is less valid for such entities.

individual or families (the entries like “Mr Gregory Edward Bailey & Mrs Margaret Ethel Bailey” or “Mme Bringaud et son fils”. The idea behind this is that they would probably exert their voting power alone or together.

- *Foreign (Domestic) Owner-Other Types*, =1 if Foreign (Domestic) owner has the *aggregated* types including unnamed individuals, the entries indicating that there are more than one private shareholders, collectively designated (for example, “Individual(s) or family(ies)”), unknown types of owners (“Unnamed private shareholders, aggregated”, Miscellaneous; Undefined company, Unknown, “n.a.”, NA), or simply missing owner type.
- *Owner is Public*, =1 if the company is owned by numerous shareholders, collectively designated as “public.” The owner type “Public” is possible only for publicly quoted companies.<sup>44</sup>

3. In addition we specify two sub-types of the owners of financial type:

- *Foreign (Domestic) Owner-Active Financial Type*, =1 if Foreign (Domestic) owner has the type Financial company, Insurance company, Other financial institution, Mutual & Pension Fund/Nominee/Trust/Trustee, Foundation/Stichting, Private Equity firms, Venture capital, Hedge funds.<sup>45</sup>
- *Foreign (Domestic) Owner-Passive Financial Type*, =1 if Foreign (Domestic) owner has the type Bank.

4. Once we identify the stakes held by the direct investors of certain type with the help of those dummies, we can generate the company-level foreign and domestic ownership variables out of the links data. We “collapse” data by summing up the stakes of the same owner type, separately domestic and foreign, by year and company ID. We also generate the simple count of number of owners, foreign owners as well as dummies

---

<sup>44</sup>Notice, that for public companies BvD may report some owners of more specific types (banks, individuals, industrial companies, etc.) with their corresponding stakes, as long as those stakes are known. We do not assign the owners designated collectively as “Public” to either foreign or domestic type because we do not know how the shareholder base is split.

<sup>45</sup>Unfortunately, we cannot separate such arguably passive types of institutional investors as pension funds from the combined type “Mutual & Pension Fund/Nominee/Trust/Trustee” or corresponding code SHARTYPE=“E”.

identifying all owner types at the company-year level. For convenience we also create the variable “Check 100 Ownership” which is the sum of all known ownership stakes. If this variable is less than 100, we have unassigned ownership percentages in certain company-years which we assign as domestic. After summation by collapse, the ownership stake percentages larger than 100% are possibly due to rounding and replacing of some special codes such as “NG” with 0.01%, or “WO” with 100%, or simply due to ownership data mistakes. Hence, we perform some cleaning steps.

5. We round all the direct ownership percentages to the second digit after the decimal (to repeat, the smallest stake observed by Orbis is 0.01%).
6. Remove duplicates in terms of `ID_NUMBER` and `YEAR` by retaining that of the two duplicates which has larger number of ownership observations, hoping that there is a bigger chance it provides useful data for percentages, etc.
7. After this, we delete the observations with the ownership percentages larger than 103% and then replace the values “slightly more” than 100%, that is in  $(100, 103]$  range, with exactly 100%. At this stage, we have constructed an unbalanced country panel of company-level direct ownership data.
8. The ownership panel is merged with financials panel by `ID_NUMBER` and `YEAR`.
9. In the combined dataset, we fill in the missing ownership time series, regardless of the time coverage of financials. In particular, we supplement ownership data in missing years using previous or consecutive non-missing values (using Stata `-carryforward-` command) assuming the following: 1. We use the existing earliest value of ownership to carry forward until i) the new non-missing value is reached or ii) the end of the time-series is reached for that company. 2. For the missing initial years of ownership, we assume that it is the same as in the first observation of non-missing data.<sup>46</sup>
10. After we filled in ownership data using lags and leads, we assign the companies with financial data but no ownership data to the category of companies with Domestic Industrial ownership.

---

<sup>46</sup>We can keep track of the filled-in ownership data by using the “Check 100 Ownership” variable which will be missing for filled-in observations.

## B.2 OECD Official Data Sources

We compare our ownership panel to alternative source on inward and outward activities of multinationals. The OECD provides data on the activities of foreign affiliates of multinationals in OECD countries in the AMNE (Activity of Multinational Enterprises) database available on the OECD data portal [https://stats.oecd.org/Index.aspx?DataSetCode=AMNE\\_IN](https://stats.oecd.org/Index.aspx?DataSetCode=AMNE_IN). The key variables presented are production, employment, value added, research and development, labor compensation, and exports. The data is broken down by country of origin of the ultimate owner (inward investment), location (outward investment), and main sector of economic activity of the multinational company following the ISIC revision 4 sector classification. AMNE covers 28 OECD host countries from 2008 onwards, although the coverage varies by country and over time.

The historical data is available from two prior databases that use the ISIC revision 3 classification: AFA (Activities of Foreign Affiliates) and FATS (Foreign Affiliates Statistics). The AFA database presents detailed data on the performance of foreign affiliates in the *manufacturing* industry for 28 OECD countries. The FATS database gives detailed data on the activities of foreign affiliates in the *services* sector for 25 OECD countries. These databases can be accessed at [http://stats.oecd.org/Index.aspx?DataSetCode=AFA\\_IN3](http://stats.oecd.org/Index.aspx?DataSetCode=AFA_IN3) and [https://stats.oecd.org/Index.aspx?DataSetCode=FATS\\_IN3\\_SERV](https://stats.oecd.org/Index.aspx?DataSetCode=FATS_IN3_SERV).

The OECD databases consist of data reported to the OECD by Eurostat based on 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 statistical office or the central bank of each country. While the key variables in the survey are common across countries, the target sample varies across countries. In order to compare the BvD ownership data to the official OECD data, the following issues have to be considered:

- Both AMNE and Orbis report data at the firm level, while AFA-FATS, covering the pre-2008 period, report the “number of enterprises or establishments.” Therefore, in certain countries, the number of foreign affiliates is not comparable with Orbis data because establishments (plants) are not included in the BvD database. All other key variables refer to a *firm* in all data sets.

- The notion of *Foreign Affiliate* is based on the concept of the “controlling interest.” According to AFA-FATS, a single institutional unit (another corporation, a household, or a government body) secures control over a corporation by owning more than half the voting shares or otherwise controlling more than half the shareholder voting power. However, the definition of controlling interest varies across countries. In most countries, controlling interest is based on direct majority ownership (50%) while others (Hungary and the United States) also consider minority control (between 10% and 50%). Moreover, some countries also include *indirectly* owned foreign affiliates. In Table B.2.1, we provide a summary of the characteristics of the AMNE database by country, whether indirect foreign control is considered, multiple owners and the main data source.
- In AMNE database, starting in 2008, the total economy is defined as sectors B–N at the ISIC revision 4. Prior to 2008, the total economy includes additional sectors.
- As shown in Table B.2.1, in the OECD databases, some countries do not sample all firms and the inclusion of firms in the surveys depends on certain economic thresholds. We always use all information available in Orbis, regardless of the firm size.
- In all cases, OECD aggregates the entire output of the entities designated as “foreign” and expresses them in national currency (Euro for Eurozone countries) or, additionally in the AFA database, as the ratio of the total manufacturing output in a given country.

Table B.2.1: OECD AMNE database sources

Country Code	Indirect Ownership	Multiple Ownership	Main Sample Characteristics
AT	from 2007		The two major sources are the structural business statistics collected by Statistics Austria and the direct investment survey conducted by the OeNB (including a question on the ultimate parent company). The OeNB submits a list of all resident foreign-controlled enterprises to Statistics Austria. Next, Statistics Austria aligns this list of enterprises with the entries in the statistical business register, with a view to creating a dataset that matches the scope of the relevant annual structural business statistics.
CZ	no	yes	Data are extracted from the annual structural survey (for non-financial enterprises) and from the annual survey in the financial sector. No special survey on enterprises with foreign participation is conducted, they are identified in the Business Register. The information on the share of foreign participation is updated by the Czech National Bank.
DK	yes	yes	From 2000 data, a fourth data source was added: the register of foreign-owned companies held by Købmandstandens Oplysningsbureau (KOB). This new data source allows for the identification of small foreign-owned companies.
EE	yes		
HU	yes	yes	Survey. Companies (NACE Rev. 1. A-O) reported by groups of firms: companies with 10-49% foreign direct investment, companies with 50-50 % foreign direct investment, and companies with more than 50% foreign direct investment according to the FDI register.
FI	yes	no	The FDI survey data from the Bank of Finland provides information on direct foreign owners. This information is obtained by a universe inquiry at a five years interval. It is supplemented every year with a limited survey. Other sources are also used to update this information (annual reports of enterprises, information on corporate acquisitions). The Enterprise Group Register is used to identify indirectly foreign-owned enterprises. It provides information on ownership relations between enterprises belonging to a group. The size threshold is approximately 60 persons employed in a group.
FR	yes		Up to 2001, source data are derived from the annual business survey (Enquête Annuelle d'Entreprise- EAE) complemented with results from the Institut National de la Statistique et des Études Économiques (INSEE) survey on financial ties (LiFi). From 1999, information from the Diane database (coedition Bureau Van Dijk Electronic Editions / Coface SCRL) has been added. The data for food industries come from the annual business survey from the SCEES, Ministry of Agriculture and these data have been included from 1999 onwards.
DE	from 2002		From 2002 to 2006, the submission of reports is required of every German enterprise with a balance sheet total of more than EUR 3 million; from 1999 until 2001, enterprises in Germany with a balance sheet total of more than EUR 500 000 were covered if a non-resident (or several economically linked non-residents) held 50% or more of the shares or voting rights of the German enterprise; reports were also required of German enterprises with a balance sheet total of more than EUR 5 million in which a non-resident (or several economically linked non-residents) held at least 10%, but less than 50% of the shares or voting rights in the German enterprise concerned. Prior to 1999, the enterprises covered were those with foreign participating interests of more than 20% and with a balance sheet totals exceeding EUR 500 000.

*Continued on next page*



**Table B.2.1 – continued from previous page**

Country Code	Indirect Ownership	Multiple Ownership	Main Sample Characteristics
GB			Information directly from the UK statistical office, the OECD does not provide info. The surveys are run from a register, which is compiled primarily from administrative information such as VAT from HM Customs and Excise and PAYE from the Inland Revenue. The register holds a number of variables including information on the country of ownership for each group. It also holds information on which UK groups have foreign affiliates. The main source of information on these foreign links for the latest annual survey was a Dun and Bradstreet publication. This was supplemented with information from ONS surveys into acquisitions and mergers of companies. These surveys are conducted on a continuous basis, collecting information when a UK company acquires or disposes of a foreign company and similarly when a foreign company acquires or disposes of a UK company. Work is currently being undertaken to review the register sources used for the survey to ensure completeness. In particular, Dun and Bradstreet's "WorldBase" information has been used to give a better estimate of the population of companies with foreign links. For both populations, in order to maximise the survey coverage of foreign direct investment assets, all groups in the top strata (containing the largest businesses) are sent questionnaires. However, in strata containing smaller businesses, only a proportion are selected. Additionally, the sample of smaller businesses is rotated to minimise burden on the respondents.
ES	yes		The average response rate to the survey exceeds 50% in terms of firms and 65% in terms of persons employed. After manual controls for non-respondents, the response rate raises up to 80% in terms of persons employed.
IT	yes		The data come from the Enterprise Department of the Central Statistical Office of Poland (CSO). They are collected via the annual statistical survey on entities with foreign capital. When there is a case of non-response of a huge company (in terms of size class or capital value), data are imputed by using data from other surveys and ownership structure is taken from known sources (previous years' questionnaires, media etc.).
PL	yes	considered	Structural Business Statistics is the main data source (no separate data collection for statistics on foreign-controlled enterprises). In order to define which companies are controlled from abroad, administrative data are used, but it is not possible to know the ultimate unit of institutional control. Thus, INE takes some steps to identify the foreign unit that actually exercises control over the affiliates in Portugal, by phone contacts with the resident enterprise.
PT	yes		The target population consists of all active foreign-controlled enterprises in Sweden and all foreign-controlled enterprises of major economic importance. The statistics cover all enterprises identified where more than 50% of the voting rights are controlled by foreign investors. The reporting unit is, in most cases, the enterprise. Statistics Sweden's Structural Business Statistics: all non-financial enterprises are surveyed annually. These statistics are based on data from annual reports, tax returns and questionnaires on revenues and costs, etc. Collection of these data is mandatory for all Swedish enterprises. Growth Analysis's survey on ownership: from reference year 1996, data on foreign-owned affiliates are based on annual questionnaires to all parent companies, subsidiaries as well as to all branches located in Sweden. The Growth Analysis register of international enterprises/groups and the business database and foreign trade statistics of Statistics Sweden are combined and merged annually.
SE			Foreign direct investment data of the Bank of Slovenia are used to identify foreign affiliates. The Structural Business Statistics database is the source for foreign affiliates' variables.
SI	yes		Data come from the annual structural business survey.
SK	yes	considered	

*Continued on next page*

**Table B.2.1 – continued from previous page**

Country Code	Indirect Ownership	Multiple Ownership	Main Sample Characteristics
NL	no		The ultimate controlling unit is determined on an annual basis by combining enterprise information from various sources, notably two surveys conducted by Statistics Netherlands, the Financial Statistics of Large Enterprise Groups (SFGO) and the Community Innovation Survey (CIS). However, from 2006 onwards, the UCI list of enterprises in the General Business Register (GBR), which is based on SFGO and CIS information, has been completed by the addition of information from an external source (Dun & Bradstreet database), providing a better insight into the total share of foreign enterprises in the Netherlands. In addition, a redesign of the Social Statistical Database in that year has resulted in a better match of the key indicators on employment, via the unique enterprise identifier (BEID), with the concurrent economic indicators. These improvements mean that the locus of control can now be established for over 90 percent of the total population of enterprises (with registered jobs in the Social Statistical Database) in the Netherlands. A weighting procedure was developed for the remaining share of enterprises that could not be matched to the GBR.
NO	yes		Statistics Norway's Structural Business Statistics is the main data source (no separate data collection for statistics on foreign-controlled enterprises). In order to define which companies are controlled from abroad, the register of foreign assets and liabilities in Norway (the SIFON register) and the balance of payments reporting are used. The Directorate of Taxes' register of shareholders has been the main source for updating the SIFON register. The press and Internet etc. are also used to map new foreign-controlled enterprises in Norway, and to map the ultimate country of ownership.

*Notes:* Source: OECD AMNE-database documentation online and national statistical office when not available. AT – Austria, CZ – the Czech Republic, DE – Germany, EE – Estonia, ES – Spain, FI – Finland, FR – France, GB – the United Kingdom, GR – Greece, HU – Hungary, IT – Italy, NL – Netherlands, NO – Norway, PL – Poland, PT – Portugal, SI – Slovenia, SK – Slovakia, SE – Sweden.

### B.3 Coverage of Foreign Ownership

Bearing all these caveats in mind, we proceed with the comparison of the data we compile from Orbis-Amadeus to the data collected by the OECD in the AMNE, FAS, and FATS databases. In our data comparison, we measure output by *total turnover* and we limit ourselves to *manufacturing* because this 1-digit sector has been covered for the longest period of time in the most OECD countries.

For the OECD data, we take the multinational turnover data from the AFA and AMNE databases, expressed in a single currency using the end-of-period exchange rates from Bloomberg, divided by the total manufacturing turnover taken from the OECD STAN database.

As said before, to stay as consistent as possible with the OECD data we identify the companies in our Orbis-Amadeus database as foreign if 10 or more percent of their equity is owned directly or, in case all direct owners are domestic over all years, ultimately by one or several foreign entities. We compute the country's foreign output share in our data as the ratio of total output aggregated over all identified foreign firms to total output of all firms in a given country.

Table B.3.2 reports these ratios by country and year. We select all the OECD member-states and Switzerland, but the Swiss data is not available in AMNE-AFA databases. We present the countries by groups, including Eurozone and non-Eurozone member states as of 2012 (Estonia became a member in 2011). Averages are simple means of ratios in columns over all years. As seen, some countries have short time series in the OECD database, either because of data privacy issues (such as Belgium or Ireland) or due to their later inclusion in the OECD (such as Hungary, Latvia, Slovenia, or Slovakia).

To facilitate the comparison of averages, Table B.3.3 reports country average shares, the average by two groups, and the overall average in a balanced sample of countries that report the data to the OECD in every year over the 2003–2012 period.

Both tables show that foreign activity makes up at most 40 percent of the total activity on average. The averages mask a lot of cross-country heterogeneity, however. In many larger European economies, the average share of foreign sales is much smaller than the group averages, being around 20 percent in Germany, France, or Italy, or 30–40 percent

in Spain, the UK, and Sweden. Interestingly, some advanced economies with well-known MNCs, such as Denmark or Finland, have average shares of around 20 percent. In other countries the average share is larger than the group averages; OFCs such as Luxembourg, and emerging markets Czech Republic, Estonia, Slovak Republic, Hungary, and Poland all have multinational sales amounting to more than half of total manufacturing sales. Table B.3.2 also makes it clear that OFCs, such as Ireland, Luxembourg, or the Netherlands display the largest differences between our data and the OECD data. This is not surprising, because our data exclude this type of financial FDI.

Table B.3.1 shows the foreign affiliate turnover as a share of total turnover in total economy (all sectors) for the firms covered by our data, both for non-Eurozone and Eurozone countries. Each cell is the ratio of the value of total output produced by our firms, which have at least 10 percent foreign ownership, relative to the value of total output produced by all firms with non-missing data in Orbis in the corresponding year and country. Some countries report data as early as the late 1980s, but the coverage in earlier years is poor. To keep the sample consistent with the following exercise, comparing coverage of Orbis with the OECD data, we limit ourselves to the 1999–2012 period.

Table B.3.1 shows the variation in multinational activity over time and across countries. In larger advanced economies, such as Germany, France, and Italy, the turnover share of foreign owned firms is in the range of 15-25 percent (and less in some years). Some countries have low shares below 10, as in Malta, Bosnia, and Belarus, while other have larger shares close to or exceeding 40 percent, as in Belgium, Estonia, Luxembourg, the Netherlands, and Slovakia. Generally, smaller economies and emerging markets in Eastern Europe have larger multinational presence at around 30 percent or more. Some of these, such as Luxembourg, Switzerland, Cyprus, and the Netherlands, are offshore financial centers (OFC), known for large round-tripping of foreign investment or regulatory/tax regimes favoring multinational affiliation. According to the recognized definition, OFC is a center which provides some or all of the following services: low or zero taxation; moderate or light financial regulation; banking secrecy and anonymity (Aykut et al., 2017).

There is sizable time variation in foreign activity for some economies. In Iceland, turnover designated as “foreign” jumped to 71 percent of total turnover in 2005 only to collapse to

5 percent in 2009, after the Global Financial Crisis (GFC) of 2008-09. Such volatility is not common, although Cyprus displays a similar pattern. In general, however, more stable economies with less questionable data quality have fairly stable shares of foreign activity over time, resembling the time averages. In almost all economies, the multinational sales share declined following the GFC from the peak in the mid-2000s. All in all, the patterns of the foreign turnover across countries reassure the reliability of our data.

Table B.3.1: Foreign Affiliate Turnover as Share of Total Turnover (Total Economy) over 1999–2012, %

Panel A: Eurozone member countries																										
	AT	BE	CY	DE	EE	ES	FI	FR	GR	IE	IT	LU	MT	NL	PT	SI	SK									
1999	31	34	46	11	30	26	14	23	17	38	17	71	1	38	21	8	33									
2000	26	34	27	10	33	28	15	20	32	18	76	6	49	24	15	42										
2001	34	34	0	12	33	27	15	21	20	26	18	75	6	48	24	15	40									
2002	28	40	91	12	33	25	20	20	22	32	8	73	5	49	22	14	44									
2003	26	40	88	9	34	25	20	18	22	26	16	62	6	50	21	17	51									
2004	28	43	12	9	38	25	28	20	25	26	15	67	8	52	30	21	46									
2005	34	46	9	9	37	25	30	25	24	39	15	51	7	45	29	18	43									
2006	32	46	11	15	39	26	30	25	24	32	16	56	2	43	29	16	44									
2007	32	45	5	19	40	26	30	25	24	34	15	63	2	49	30	16	32									
2008	31	43	4	21	37	29	29	24	29	28	16	58	1	48	30	15	44									
2009	24	45	0	20	38	29	27	22	26	29	16	38	0	38	25	18	36									
2010	20	44	0	20	39	29	16	22	27	32	15	22	0	40	26	17	34									
2011	24	43	0	20	39	29	17	21	29	31	16	22	0	41	29	16	37									
2012	22	44	0	20	37	28	15	20	30	34	14	26	0	39	31	16	41									
Avg.	28	42	21	15	36	27	22	22	24	31	15	54	3	45	26	16	40									

Panel B: Non-Eurozone member countries																										
	AL	BA	BG	BY	CH	CZ	DK	GB	HR	HU	IS	LT	LV	MD	ME	MK	NO	PL	RO	RS	RU	SE	TR	UA		
1999	79	5	18	1	22	28	20	26	17	18	4	35	19	13		29	23	31	28	15	30	21	1	8		
2000	18	6	19	2	22	31	23	26	22	24	4	39	21	18	23	28	24	33	28	16	11	20	38	13		
2001	22	3	22	2	21	34	22	26	22	6	4	39	24	30	10	30	24	31	27	14	9	21	26	0		
2002	25	4	31	1	20	31	22	26	23	22	3	29	25	32	4	26	21	34	30	17	7	23	33	9		
2003	25	4	28	0	20	32	21	25	22	72	7	26	27	26	15	25	22	34	35	21	10	23	28	9		
2004	33	4	22	1	35	35	21	26	24	26	8	28	23	29	9	27	28	38	36	23	9	29	17	9		
2005	51	5	19	1	35	38	23	26	23	30	71	28	26	34	23	43	32	40	38	28	10	28	11	13		
2006	45	4	26	1	36	39	22	25	26	31	13	29	24	40	22	29	30	40	42	26	10	28	12	16		
2007	31	3	29	8	40	39	23	24	26	29	24	23	27	28	26	23	29	40	37	22	11	27	22	16		
2008	12	3	29	2	39	37	22	29	25	33	18	25	28	47	25	16	28	40	36	18	14	29	23	18		
2009	10	2	27	0	29	35	22	25	25	5	28	24	34	21	6	22	38	34	16	14	28	20	20			
2010	22	3	28	0	22	32	17	27	26	29	5	32	26	34	22	12	26	34	34	24	13	24	18	22		
2011	18	3	29	0	23	32	16	28	23	29	6	33	26	37	19	9	25	36	34	26	12	23	14	23		
2012	19	3	30	0	15	32	16	25	23	27	9	35	26	11	21	6	15	35	34	24	14	23	16	20		
Avg.	29	4	25	1	27	34	21	26	23	29	13	31	25	30	18	22	25	36	34	21	12	25	20	14		

Notes: The table reports sales of foreign affiliates as percentage of total sales in total economy (all sectors), as reported by Orbis database. 'Avg.' is the average share over all available years. The country codes correspond to the following Eurozone countries: AT (Austria), BE (Belgium), DE (Germany), EE (Estonia), ES (Spain), FI (Finland), FR (France), IE (Ireland), IT (Italy), LU (Luxembourg), NL (the Netherlands), PT (Portugal), SK (Slovakia), SI (Slovenia), and the following non-Eurozone countries AL (Albania), BA (Bosnia and Herzegovina), BY (Belarus), CH (Switzerland), CZ (Czech Republic), DK (Denmark), GB (United Kingdom), HR (Croatia), HU (Hungary), IS (Iceland), LT (Lithuania), LV (Latvia), MD (Moldova), ME (Montenegro), MK (Macedonia), NO (Norway), PL (Poland), RO (Romania), RS (Serbia), RU (Russian Federation), SE (Sweden), TR (Turkey), UA (Ukraine).

Table B.3.2: FOREIGN AFFILIATE TURNOVER AS A SHARE OF TOTAL TURNOVER (MANUFACTURING), %

Panel A: Eurozone member countries

	AT		BE		DE		EE		ES		FI		FR		IE		IT		LU		NL		PT		SI		SK	
	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD
1999	31	.	54	.	18	10	45	.	36	30	16	16	35	34	49	.	20	.	83	45	36	31	27	16	13	.	35	.
2000	28	.	56	.	16	9	47	.	38	30	14	14	35	35	35	.	23	.	91	44	40	24	25	17	22	.	54	.
2001	25	.	54	.	22	8	48	.	36	29	15	16	34	36	36	.	23	22	85	45	44	35	25	16	21	.	56	.
2002	32	.	57	.	21	25	47	.	36	27	14	16	34	34	44	.	20	19	77	47	47	40	23	16	20	.	53	.
2003	32	30	60	.	19	27	49	49	36	27	15	16	34	35	37	.	21	20	81	50	42	40	24	24	23	24	65	63
2004	28	32	57	.	25	27	57	52	39	26	36	16	25	32	32	.	21	20	89	49	43	43	31	24	21	25	65	69
2005	37	35	61	.	23	29	55	51	36	26	39	16	26	34	43	.	20	19	82	48	42	41	25	25	23	27	60	70
2006	36	37	61	.	19	30	59	53	38	30	40	16	26	33	29	.	22	19	86	51	41	43	26	22	22	27	66	71
2007	33	39	63	.	18	29	58	52	36	30	40	18	25	30	48	.	21	19	87	54	39	45	27	28	22	27	68	78
2008	29	38	62	.	20	30	54	54	41	31	38	19	26	33	39	77	22	19	84	56	28	46	27	25	21	27	64	77
2009	28	36	66	.	23	26	50	53	41	30	40	19	25	30	47	79	22	19	26	59	27	46	24	25	21	29	64	78
2010	24	41	62	62	21	27	56	59	41	37	19	18	25	30	43	79	22	19	18	62	26	47	24	26	20	29	65	76
2011	25	40	63	63	23	29	58	61	40	39	20	19	25	30	40	78	23	19	23	62	26	48	23	27	19	30	66	78
2012	25	39	65	63	25	30	57	60	40	40	20	18	26	29	37	79	21	21	23	64	28	50	24	26	18	32	68	81
Avg.	29	37	60	63	21	24	53	54	38	31	26	17	29	32	40	78	21	20	67	53	36	41	25	23	20	28	60	74

Panel B: Non-Eurozone member countries

	CH		CZ		DK		GB		HU		LV		NO		PL		SE	
	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD	Orbis	OECD
1999	21	.	42	27	14	12	31	33	24	.	16	.	38	28	45	34	20	29
2000	23	.	41	39	15	18	31	34	33	.	15	.	43	27	46	35	19	33
2001	23	.	42	43	15	17	31	37	28	.	16	.	44	30	43	35	23	39
2002	23	.	44	46	17	18	29	39	23	.	16	.	25	29	44	39	24	40
2003	23	.	43	48	20	19	28	40	28	55	19	.	31	26	46	42	27	41
2004	55	.	46	53	21	21	32	40	33	60	18	.	47	25	50	45	42	40
2005	57	.	52	53	25	23	30	43	40	59	22	.	49	27	53	45	42	39
2006	58	.	54	56	24	24	27	44	46	62	20	.	48	29	55	46	41	42
2007	61	.	52	64	24	26	26	45	40	66	29	.	47	28	55	46	41	40
2008	45	.	51	62	25	26	42	46	40	65	31	30	47	27	57	45	43	40
2009	23	.	50	64	36	26	36	47	41	66	32	32	46	32	57	48	46	42
2010	20	.	50	65	29	30	37	50	44	68	25	32	32	32	48	48	33	44
2011	20	.	50	66	28	31	39	54	41	69	25	31	41	32	47	47	34	44
2012	17	.	52	66	28	31	38	54	39	69	24	31	32	33	45	45	34	44
Avg.	34	.	48	54	23	23	33	43	36	64	22	31	41	29	50	43	34	40

NOTES: The table report sales of foreign affiliates in manufacturing as percentage of total manufacturing sales as reported in the Orbis database (under "Orbis") and the OECD's AFA and AMNE databases (under "OECD"). "Avg." is the average share over all available years. The country codes correspond to the following Eurozone countries: AT (Austria), BE (Belgium), DE (Germany), EE (Estonia), ES (Spain), FI (Finland), FR (France), IE (Ireland), IT (Italy), LU (Luxembourg), NL (the Netherlands), PT (Portugal), SK (Slovakia), SI (Slovenia), and the following non-Eurozone countries CH (Switzerland), CZ (Czech Republic), DK (Denmark), GB (United Kingdom), HU (Hungary), LV (Latvia), NO (Norway), PL (Poland), and SE (Sweden).

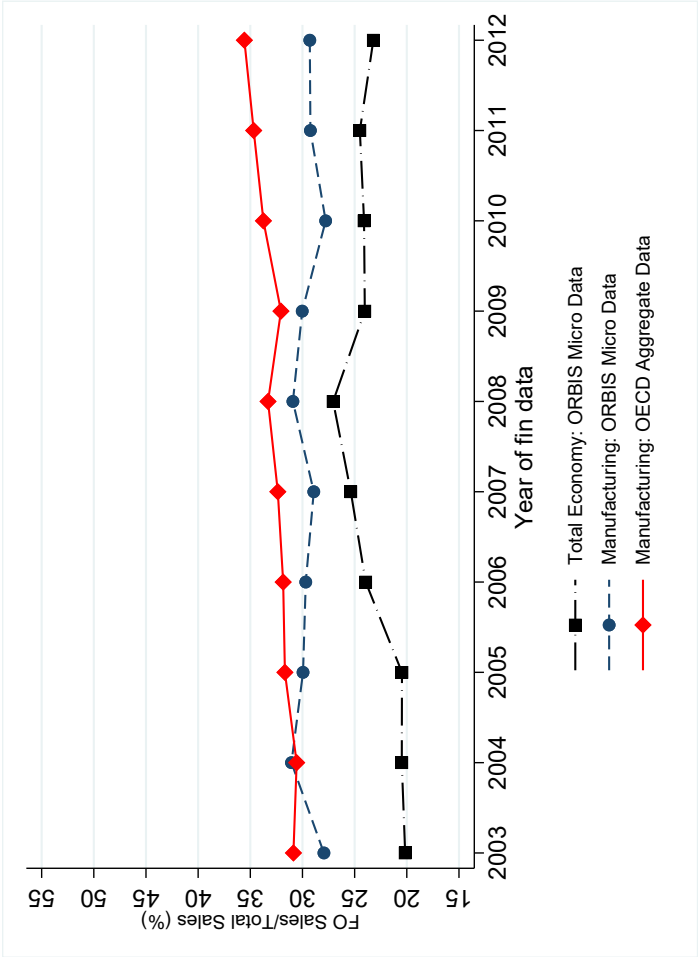
Table B.3.3: AVERAGE FOREIGN AFFILIATE TURNOVER AS A SHARE OF TOTAL TURNOVER (MANUFACTURING), %, 2003-2012

Country	Country Code	Orbis	OECD
Austria	AT	30	37
Germany	DE	22	28
Estonia	EE	55	54
Spain	ES	39	32
Finland	FI	31	18
France	FR	26	31
Italy	IT	21	19
Luxembourg	LU	60	55
Netherlands	NL	34	45
Portugal	PT	25	26
Slovenia	SI	21	28
Slovak Republic	SK	65	74
Average Eurozone		36	37
Czech Republic	CZ	50	60
Denmark	DK	26	26
Great Britain	GB	33	46
Hungary	HU	39	64
Norway	NO	42	29
Poland	PL	51	46
Sweden	SE	38	42
Average Other		40	45
Average OECD		38	41

NOTES: The table reports sales of foreign affiliates in manufacturing as percentage of total manufacturing sales as reported in the Orbis database (under *Orbis*) and the OECD's AFA and AMNE database (under *OECD*). The sample is balanced, with data for all countries available in all years 2003-2012. The numbers for each country represent the average share over all available years, while the group averages are computed over all countries and years.



Figure B.3.1: FOREIGN FIRMS SALES SHARE IN TOTAL SALES IN ORBIS AND IN OECD



NOTES: The figure reports sales of foreign affiliates in the total economy as percentage of total sales as reported by Orbis database (long-dashed line with squares), sales of foreign affiliates in manufacturing as percentage of total manufacturing sales as reported by Orbis database (dashed line with circles) and the OECD's AFA and AMNE database (solid line with diamonds). In every year, we first aggregate the sales of multinationals in monetary terms across all sample countries and divide by the total sales in these countries. With Orbis data we compute the multinational activity from the turnover of firms with 10 or more percent foreign ownership stake and divide by the total turnover across all Orbis firms (in the total economy and in the manufacturing sector, correspondingly) in our sample countries. Foreign multinational activity from the OECD data is the sum of the multinational turnover in manufacturing reported by the AFA and AMNE databases of the OECD divided by the total manufacturing turnover in these countries from the OECD STAN database. The sample is balanced with the data for all countries available in all years 2003-2012. The countries include the following Eurozone countries: Austria, Germany, Estonia, Finland, France, Italy, Luxembourg, the Netherlands, Portugal, Slovak Republic, Slovenia, and Spain and the following non-Eurozone countries Czech Republic, Denmark, the UK, Hungary, Norway, Poland, and Sweden. Belgium, Ireland, Switzerland and Latvia are eliminated due to short or missing OECD data (see Table B.3.2)

## C Aggregate Comparison

### C.1 Eurostat Data for Aggregate Economic Activity

Eurostat provides data on main economic indicators as well as some additional variables for all EU member states, Norway, Switzerland, and some candidate and potential candidate countries. The data is collected either by national statistical agencies via surveys or by drawing on business registers and/or other administrative sources. Eurostat has multiple sub-databases categorized according to different themes. To serve for our purpose, we focus on the following two: Structural Business Statistics (SBS) and Business Demography (BD). Both databases cover variables related to business demography, while their scope differs in some aspects. So, before presenting the details on our comparison exercises, we firstly describe basic characteristics of these two databases and then explain how we utilize them in turn to compare the Eurostat data with our data obtained from Orbis-Amadeus.

Starting in 1995, the **SBS** data provides information that describes the structure, conduct, and performance of economic activities at a very fine level of detail (several hundred economic sectors).<sup>47</sup> The SBS coverage was limited to Sections C to K of NACE Rev. 1.1 until 2007. Starting from the reference year 2008, the data is available for sectors B to N and sub-sector S95 of NACE Rev. 2 classification. Some of these sectors, such as NACE Rev. 2 sector K and NACE Rev. 1.1. sector J are only partially covered with data for insurance services, credit institutions, and pension funds but not all financial intermediation activities.

The main variables in the SBS data are business demographic variables (e.g., number of enterprises), “output related” variables (e.g., turnover, value added), “input related” variables such as labor input (e.g. employment, hours worked), goods and services input (e.g., total purchases), and capital input (e.g., material investments). This information is available for different size categories. Among all, we use “TOTAL,” “0-19 employees,” “20-249 employees,” and “250+ employees” for our purpose.

In 1997, the **BD** data started to collect variables mostly related to the demography of the business population.<sup>48</sup> The BD data is not as detailed in terms of variables as SBS data,

---

<sup>47</sup>See [http://ec.europa.eu/eurostat/cache/metadata/en/sbs\\_esms.htm](http://ec.europa.eu/eurostat/cache/metadata/en/sbs_esms.htm) for a detailed description.

<sup>48</sup>See [http://ec.europa.eu/eurostat/cache/metadata/EN/bd\\_esms.htm](http://ec.europa.eu/eurostat/cache/metadata/EN/bd_esms.htm) for a detailed description.

which enables it to cover a larger number of sectors. Starting with the reference year 2008, the BD data covers NACE Rev. 2 sections B to N (excluding activities of holding companies, K64.2). Data for sections P, Q, R and S are provided on a voluntary basis. NACE Rev. 1.1 was used up to the reference year 2007 covering the data for sections C to K (excluding activities of holding companies-K74.15). Sectors M, N, and O were reported on a voluntary basis and, therefore, not available for all countries.

The main indicators of the BD data category include population of active enterprises, number of enterprise births, number of enterprise survivals up to five year, number of enterprise deaths, related variables on employment, derived indicators such as birth rates, death rates, survival rates and employment shares, and an additional set of indicators on high-growth enterprises and “gazelles” (high-growth enterprises that are up to five years old). This information is available for firms of different size and legal form. The three main categories by size are All, Zero, 10+.

As it is immediately seen, there are several differences between the BD data and the SBS data, important for our purposes. The BD data counts only active enterprises. Unlike the BD data, the SBS data counts both active and inactive companies. Moreover, the SBS data provides the information for multiple firm size groups based on employment, whereas the firm size categories are limited in the BD data. Another advantage of the SBS data over the BD data is that the SBS data provides information on monetary values such as turnover, wage bill, investment, etc., and all these variables are available for multiple firm size categorizes. The BD data covers more sectors but contains less variables. In addition, the BD data provides information separately for “Zero” firm size category that counts self-employed workers. Having these differences in mind, if we do comparison based on number of enterprises and employment for “Total” or “Zero” categories, we always use the BD data. In case, we do comparison based on gross output, we always use the SBS data. For all other cases where we do comparison for SMEs, we always use the SBS data. We cannot use the BD data for comparisons based on employment for SMEs because, as mentioned above, the BD data lacks employment information for SMEs. In what follows, we present the details of the comparison of our data to statistics computed using Eurostat data.

## C.2 Comparison of Orbis-Amadeus to Eurostat Data: Implementation Details

There are structural differences between Orbis-Amadeus and Eurostat. Given this, we follow some steps in order to make Eurostat data sets and our BvD data comparable before conducting comparison exercises. The details on these steps are given in below:

1. We identify Eurostat sectors based on NACE Rev. 2 classification because our Orbis-Amadeus data uses this classification (see step 6 in Section A.5.2). In order to do this, we go through the process of matching the industry classifications pre- and post-2008 within each Eurostat data category. We first create a variable **NACE1** in both Rev. 1.1. and Rev. 2 files. We then fill in the values of this variable by values of NACE Rev. 2 codes based on our augmented BvD sector variable (**NACEREV2CCCODE**) using the official NACE Rev. 2 Level 1 classification. For example, we replace **NACE1** with C if it is coded as D in NACE Rev. 1.1. file. Then, we merge these two files by **NACE1** for each country-year-company category (company category is based either on size and/or legal form). Thus, this common **NACE1** variable in Orbis-Amadeus and Eurostat enables us to compare them based on any sector defined at NACE Rev. 2 Level 1 classification.
2. We determine overlapping sectors across Orbis-Amadeus and Eurostat SBS data sets based on the **NACE1** variable. While Orbis-Amadeus data, by construction, covers all sectors for a given country-year, Eurostat data provides information for the business economy with the exception of some sectors. Table C.2.2 lists the sectors with available information in Eurostat data sets. However, this list is differential based on the variable of interest. For example, the variable Turnover provided by Eurostat SBS data is not available for all sectors given in Table C.2.2. The availability of this variable differs across country-sector-year triplets, and this should be taken into account especially in total economy comparison exercises.

To exemplify, Eurostat data sets provide information for the construction sector (NACE Rev. 2 sector F) between 1995–2012. However, according to Table C.2.3, Eurostat SBS data lacks information on gross-output for this sector in Belgium for the years 2008, 2009, 2011 and 2012. If we had disregarded such detail and had computed total

gross output over BvD Belgium firms for all overlapping sectors by just considering the correspondence given in Table C.2.2, we would have inflated the BvD aggregate for 2008, 2009, 2011 and 2012. To put it differently, since BvD aggregates are used as numerators in the ratios we use to construct percentages, overlooking such details would produce artificial higher percentages. In order to avoid such mistakes, we instead follow the sector correspondence given in Table C.2.3, and construct the percentages accordingly. To be more exact, in case we compute the percentage based on gross output for Belgium in 2009, we aggregate gross-output over the sectors which have non-missing gross output in *both* Eurostat SBS and BvD data sets. We label those sectors with 1 in Table C.2.3.

3. As stated above, Eurostat SBS data provides information on economic activity (i.e., gross-output and employment) for different size categories. Among all, we use “TOTAL,” “0-19 employees,” “20-249 employees,” and “250+ employees” to compare size distribution for total economy and manufacturing sector. BvD provides firm-level information on economic activity for *all sectors* of a given European country-size category in any year of interest, whereas Eurostat SBS data provides information on the corresponding measure in a given size category for *selected sectors*, which are labelled with 1 in Table C.2.4. So, to make sure we do not create measurement error we construct a “hypothetical aggregate” economy based on aggregating the sectors where official data by size class is provided by the Eurostat. Doing this, we compare exact sectors, then compute the percentages in Table 2, accurately.
4. Eurostat BD data provides information on the main indicators for the categories of different firm size and legal form. Among the legal form categories, we are forced to use “Total” which refers to all firms with different legal forms assuming that BvD collects data from firms of all legal firms. We do not want to deal with legal firm issue since it is hard to identify the legal form information for each country in Orbis-Amadeus data given differential filing policy across countries (See Table A.6.1). Among the size categories, we use “Total” and “Zero” to construct a new size category i.e. “AllminusZero” which refers to all firms excluding the *self employed workers*. This enables us to compare manufacturing employment of Orbis-Amadeus to that of Eurostat BD in the most

appropriate manner since BvD excludes self-employed workers by construction.

Eurostat SBS data provides information only for different size categories. However, it doesn't provide information under "Zero" size category, which forced us to use "Total" size category in all comparison exercises where Eurostat SBS data was used.

5. In our comparison exercises, we use two main indicators to measure economic activity: employment and gross-output. Eurostat data provides multiple variables related to economic activity and the related variables are different in terms of definition. So, in order to precisely compare the coverage based on economic activity, we pick Eurostat variables that have close definitions with those of BvD data. To illustrate, in our comparison exercises based on gross-output we use `OPRE` from BvD and `V12110-Turnover` from Eurostat SBS data. We express these financial variables in real dollars 2005 base using original values in Eurostat SBS data (see Step 2 in Chapter A.5.3). Additional correspondences as well as the variable definitions are given in Table C.2.1.
6. In comparison exercises based on number of enterprises, we use either Eurostat SBS or Eurostat BD data. In case we used the BD data, we drop *inactive* companies checking the variable `STATUS` in our data because the BD data counts only active companies.<sup>49</sup>
7. We construct two different samples using our BvD data and repeat our comparison exercised for each sample: Total Sample and TFP Sample. The Total Sample consists of firms that report data with positive values of the corresponding measure (i.e. employment (`EMPL`) and gross output (`OPRE`)), whereas the TFP Sample consists of firms that report positive values on employment (`EMPL`) or wage bill (`STAF`), and tangible fixed assets (`TFAS`), gross output (`OPRE`), materials (`MATE`).
8. For a given company, Orbis-Amadeus provides financial statements regarding different consolidation codes i.e. C1, C2, U1 and U2.<sup>50</sup> Given this fact, we first drop C2 accounts

---

<sup>49</sup>If a given BvD company is read as "Inactive," "Dissolved," "In liquidation," and "Bankruptcy," we count it as inactive.

<sup>50</sup>C1: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.), where the company headquarter has no unconsolidated account, C2: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.) where the company headquarter also presents an unconsolidated account, U1: account of a company with no consolidated account, and U2: account of a company with a consolidated account.

to avoid double accounting in our comparison exercises. For some specific cases, we further drop C1 accounts. To illustrate, in cases where we use the Total Sample, we drop C1 accounts for all countries except Spain and Italy. In the cases where we use the TFP Sample, we drop C1 accounts for all countries except Spain, Italy, Cyprus, Denmark, the UK, Greece, Ireland, and Lithuania.

9. As a final step, we want to make sure percentages in our comparison exercises are not inflated by outliers. So, we check distributions of the underlying economic activity measure within a given country-sector-year triplet and winsorize data if neccessary.<sup>51</sup>

We put Table C.2.1 as a guide for readers to help them follow the tables related to our comparison exercises.

---

<sup>51</sup>For space considerations, the list of the winsorized country-sector-year triplets as well as details on the amount of winsorizing (which varies between 0.01% and 0.5%) are not reported here. All these details are available upon request.

Table C.2.1: DETAILS OF COMPARISON EXERCISES

Table 1		
Data	Orbis-Amadeus	Eurostat SBS
Year	1999-2012	1999-2012
Variable	OPRE	V12110-Turnover
Size	All	TOTAL - Total
Sector	See Table C.2.3	See Table C.2.3
Panel B of Table 2		
Data	Orbis-Amadeus	Eurostat SBS
Year	2006	2006
Variable	OPRE	V12110-Turnover
Size	1-19, 20-249, 250 + employees	0-19, 20-249, 250 + employees
Sector	See Table C.2.4	See Table C.2.4
Panel B of Table 2		
Data	Orbis-Amadeus	Eurostat SBS
Year	2006	2006
Variable	EMPL	V16110
Size	1-19, 20-249, 250 + employees	0-19, 20-249, 250 + employees
Sector	See Table C.2.4	See Table C.2.4

NOTES: Table C.2.1 presents the details of the comparison exercises. The sectors are compared based on NACE Rev. 2 Level 1 Classification (**NACE1**). In Tables D.2.1–D.2.2, total economy in the reference country-year corresponds to all common available sectors with Eurostat BD and SBS data, respectively. The definitions of the variables are as follows: **OPRE**: Total operating revenues (Net sales + Other operating revenues+ Stock variations). The figures do not include VAT. Local differences may occur regarding excises taxes and similar obligatory payments for specific market of tobacco and alcoholic beverage industries; **EMPL**: Total number of employees included in the company’s payroll, **V16910**: Number of persons employed in the population of active enterprises in t, **V13310**: Personnel costs are defined as the total remuneration, in cash or in kind, payable by an employer to an employee (regular and temporary employees, as well as home-workers) in return for work done by the latter during the reference period, **V12110**: Turnover comprises the totals invoiced by the observation unit during the reference period, and this corresponds to the total value of market sales of goods and services to third parties, **V16110**: Number of persons employed. In BvD data, for the construction of the TFP sample, we also used the variables **STAF**, **TFAS**, and **MATE**. **STAF**: All the employees costs of the company (including pension costs), **TFAS**: Book value of tangible fixed assets i.e. plant, equipment and machinery, and **MATE**: Material Costs.



Table C.2.1 (Cont'd.): DETAILS OF COMPARISON EXERCISES

Tables D.1.1 & D.1.2		
Data	Orbis-Amadeus	Eurostat SBS
Year	1999-2012	1999-2012
Variable	OPRE	V12110-Turnover
Size	All	TOTAL - Total
Sector	C	C
Tables D.1.3 & D.1.2		
Data	Orbis-Amadeus	Eurostat SBS
Year	1999-2012	1999-2012
Variable	EMPL	V16110
Size	All	TOTAL - Total
Sector	C	C
Panel A in Table D.1.5		
Data	Orbis-Amadeus	Eurostat SBS
Year	2006	2006
Variable	OPRE	V12110-Turnover
Size	1-19, 20-249, 250 + employees	0-19, 20-249, 250 + employees
Sector	C	C
Panel B in Table D.1.5		
Data	Orbis-Amadeus	Eurostat SBS
Year	2006	2006
Variable	EMPL	V16110
Size	1-19, 20-249, 250 + employees	0-19, 20-249, 250 + employees
Sector	C	C
Table D.2.1		
Data	Orbis-Amadeus	Eurostat BD
Year	2007, 2008, 2009	2007, 2008, 2009
Variable	ID_NUMBER	V11910-Population of active enterprises in t
Size	All	Country specific
Sector	Total economy	Total economy
Table D.2.2		
Data	Orbis-Amadeus	Eurostat SBS
Year	2007, 2008, 2009	2007, 2008, 2009
Variable	ID_NUMBER	V11110-Number of enterprises
Size	All, 1-19, 20-249, 250 + employees	TOTAL-Total, 0-19, 20-249, 250 + employees
Sector	Total economy	Total economy

Table C.2.2: SECTOR COVERAGE IN EUROSTAT DATA

CODE	DEFINITION	SBS			BD	
		1995-2007	2008-2012	1997-2007	2008-2012	
A	Agriculture, forestry and fishing	0	0	0	0	0
B	Mining and quarrying	1	1	1	1	1
C	Manufacturing	1	1	1	1	1
D	Electricity, gas, steam and air-conditioning supply	1	1	1	1	1
E	Water supply, sewerage, waste management and remediation	1	1	1	1	1
F	Construction	1	1	1	1	1
G	Wholesale and retail trade, repair of motor vehicles and motorcycles	1	1	1	1	1
H	Transportation and storage	1	1	1	1	1
I	Accommodation and food service activities	1	1	1	1	1
J	Publishing, audiovisual and broadcasting activities /telecommunications/IT and other information services	1	1	1	1	1
K	Financial and insurance activities	1	1	1	1	1
L	Real estate activities	1	1	1	1	1
M	A Legal, accounting, management, architecture, engineering, technical testing and analysis activities	1	1	1	1	1
N	Administrative and support service activities	1	1	1	1	1
O	Public administration and defence, compulsory social security	0	0	0	0	0
P	Education	0	0	0	1	1
Q	Human health services and residential care and social work activities	0	0	0	1	1
R	Arts, entertainment and recreation	0	0	0	1	1
S	Other services	1	1	1	1	1
T	Activities of households as employers; undifferentiated goods and services producing activities of households for own use	0	0	0	0	0
U	Activities of extra-territorial organisations and bodies	0	0	0	0	0

NOTES: Table C.2.2 presents sector coverage for the SBS data and the BD data, respectively. Sectors are defined according to NACE Rev. 2 Level 1 classification. We label sector-period pairs with 1 if they report non-missing information in any data category, 0 otherwise. In the SBS data, sectors K and S are partially covered that the first provides information only for insurance services, credit institutions, and pension funds; the latter provides information only for repair of computers and personal and household goods (S95) starting from 2008. In the BD data, the information from sectors P, Q, R, and S are provided on a voluntary basis, and Sector K64.2 (activities of holding companies) is excluded.

Table C.2.3: SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS: 1999–2012  
 BASED ON GROSS OUTPUT

AT	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	1	0

BE	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	0	0	0	0	0	0	0	1	0	0
C	1	1	1	0	1	1	1	1	1	1	1	1	1	1
D	1	1	1	0	1	1	1	1	1	0	0	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	0	1	1	1	1	1	0	0	1	0	0
G	1	1	1	0	1	1	1	1	1	0	0	1	0	0
H	1	1	1	0	1	1	1	1	1	1	1	1	1	0
I	1	1	1	0	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	0	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	0	0	0	0

BG	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	1	1	1	1	1	1	0	0	0	0	0
C	0	0	0	1	1	1	1	1	1	1	1	1	1	1
D	0	0	0	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	0	0	0	0	0	0	1
F	0	0	0	1	1	1	1	1	1	0	0	0	0	0
G	0	0	0	1	1	1	1	0	1	0	0	1	0	0
H	0	0	0	1	1	1	1	1	1	1	1	1	1	0
I	0	0	0	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	1	1	1	1	1	1	1	0	0	0	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	0	0	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

NOTES: Table C.2.3 presents the coverage in the Eurostat SBS data by economic activity for given country-years. The variable of interest is gross-output and the information is based on NACE Rev. 2 Level 1 categories. This table is used as reference for total economy comparisons based on gross-output. In the table, the corresponding country-year is labelled as 1 if information on gross-output is available, 0 otherwise. The country codes reported are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia).

Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

CZ	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	1	1	1	1	0	1	1
F	1	1	1	1	1	1	1	1	1	1	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	0	0	0
I	1	1	1	1	1	1	1	1	1	1	1	0	0	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	0	1	1	1	1	1	1	1	1	1	0	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	0	0	0	0
DE	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	0	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DK	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	1	1	1	0
EE	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	0	0	0	0	0	0	1	1	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	0	0	0	0	0	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	1	1	1	0
N	0	0	0	0	0	0	0	0	0	0	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

ES														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	0	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FI														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	1	1	1	1	1	1
F	1	1	1	1	1	1	0	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	1	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	0	0
M	0	0	0	0	0	0	1	0	0	1	1	1	1	0
N	0	0	0	0	0	0	1	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	1	1	0	0

FR														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	1
G	1	1	1	0	1	1	1	1	1	0	0	1	0	1
H	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1
J	0	0	0	0	0	0	0	0	0	1	1	1	1	1
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M	0	0	0	0	0	0	0	0	0	1	1	1	1	1
N	0	0	0	0	0	0	0	0	0	1	1	1	1	1
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

GB														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	1	1	1	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

NOTES: See the notes above.

Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

GR	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	1	1	1	1	1	1	1	1	1	1
C	1	1	0	0	1	1	1	1	1	1	1	1	1	1
D	0	0	0	0	1	1	1	1	1	0	1	1	1	1
E	0	0	0	0	0	0	0	0	0	0	1	1	1	1
F	0	0	0	1	1	1	1	1	1	0	0	0	0	0
G	0	0	0	0	1	1	1	1	1	0	0	0	0	0
H	0	0	0	0	1	1	1	1	1	1	1	1	0	0
I	0	0	0	0	1	1	1	1	1	1	1	1	0	0
J	0	0	0	0	0	0	0	0	0	1	1	1	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	1	1	1	1	1	1	1	1	0	0
M	0	0	0	0	0	0	0	0	0	1	1	1	0	0
N	0	0	0	0	0	0	0	0	0	1	1	1	0	0
S	0	0	0	0	0	0	0	0	0	1	0	1	0	0

HR	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	0	0	0	0	0	0	1	1	1	1
C	0	0	0	0	0	0	0	0	0	1	1	1	1	1
D	0	0	0	0	0	0	0	0	0	0	0	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	0	0	0	0	0	0	0	0	0	0	0	1	0	0
G	0	0	0	0	0	0	0	0	0	0	0	1	0	0
H	0	0	0	0	0	0	0	0	0	1	1	1	1	0
I	0	0	0	0	0	0	0	0	0	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

HU	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	0	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	0	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	1	0	0	0

IE	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	0	1	1	0	0	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	0	0	0	0	0	0	1	1	0	0	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	0	0	0	0	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	0	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

IT														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	1
G	1	1	1	1	1	1	1	1	1	0	0	1	0	1
H	1	1	1	1	1	1	1	1	1	1	1	1	1	1
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1
J	0	0	0	0	0	0	0	0	0	1	1	1	1	1
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0	0	0	1	1
M	0	0	0	0	0	0	0	0	0	1	1	1	1	1
N	0	0	0	0	0	0	0	0	0	1	1	1	1	1
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

LT														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	0	0	0	0	0	0	0	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	0	0	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	0	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	0	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	1	0

LU														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	1	1	1	1	1	1	1	1	1	0
C	0	0	0	0	1	1	1	1	1	1	1	1	1	0
D	0	0	0	0	1	1	1	1	1	1	1	1	1	0
E	0	0	0	0	0	0	0	0	0	1	1	1	1	0
F	0	0	0	0	1	1	1	1	1	0	0	1	0	0
G	0	0	0	0	1	1	1	1	1	0	0	1	1	0
H	0	0	0	0	0	1	1	1	1	1	1	1	1	0
I	0	0	0	0	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	1	1	1	1	1	0	0	0	1	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

LV														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	0	0	0	0	0	0	0	0	0	1	1	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	0	1	1	1	1	1
E	0	0	0	0	0	0	0	0	1	1	0	0	1	0
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

NL														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	1	1	1	0	0	0	0	0	1	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	0	0	1	1	1	1	0	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	0	0	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

NO														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	0	0	0	0	1	1	1	1	1	1	1	1	1	1
C	1	1	0	1	1	1	1	1	1	1	1	1	1	1
D	0	0	0	0	0	0	0	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	0	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	0	1	0	0	0	1	1	1	1	1	1	1	1	0
I	0	1	0	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	1	0	0	1	1	1	1	0	1	0	0	0	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

PL														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	1	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	1	0	0	0

PT														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	1	1	1	1	1	1	1
F	0	0	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	1	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Table C.2.3 (Cont'd.): SECTOR COVERAGE OF EUROSTAT COUNTRIES ACROSS YEARS:  
1999–2012

BASED ON GROSS OUTPUT

RO														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	1	1	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	0	1	1	1	1	1	1	1	1	0	0	1	0	0
H	0	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0	0	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SE														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	1	1	1	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	1	1	1	1	1	1	1	0
N	0	0	0	0	0	0	1	1	1	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SI														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	0	0	1	1	1	1	0	1	0	1	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	0	0	0	0	1	1	1	1	1
E	0	0	0	0	0	0	0	1	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	1	1	1	1	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	1	0	0	0
SK														
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
B	1	1	1	1	1	1	0	1	0	0	0	0	0	0
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	0	0	0	0	0	1	0	1
E	0	0	0	0	0	0	0	0	0	0	1	1	1	1
F	1	1	1	1	1	0	0	0	0	0	0	1	0	0
G	1	1	1	1	1	1	1	1	1	0	0	1	0	0
H	1	1	1	1	1	1	1	1	1	1	1	1	1	0
I	1	1	1	1	1	1	1	1	1	1	1	1	1	0
J	0	0	0	0	0	0	0	0	0	1	1	1	1	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0	0	1	0	0
M	0	0	0	0	0	0	0	0	0	1	1	1	1	0
N	0	0	0	0	0	0	0	0	0	1	1	1	1	0
S	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table C.2.4: SECTOR COVERAGE OF EUROSTAT COUNTRIES, 2006

BASED ON ECONOMIC ACTIVITY

PANEL A: GROSS OUTPUT																												
Sector		AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HU	HR	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
B	1	0	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
E	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0
F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

PANEL B: EMPLOYMENT																												
Sector		AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HU	HR	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
B	1	0	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
E	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

NOTES: Table C.2.4 presents sector coverage of the Eurostat countries based on the SBS data. The SBS data provides information on economic activity by size class for only a subset of sectors in the economy. In this table, we label country-sector pairs with 1 if they provide non-missing information by size class in 2006 for the corresponding measure (gross output or employment), 0 otherwise. Sectors are defined according to NACE Rev. 2 Level 1 classification. We use this table as reference for total economy comparisons, presented in Table 2. The country codes reported are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia).

## D Additional Material

### D.1 Detailed Statistics for the Manufacturing Sector

Tables [D.1.1](#) & [D.1.4](#) show how much of the official gross output and employment data from Eurostat we cover in our data for the manufacturing sector in the EU countries. We show these tables in two different samples: the Total Sample and the TFP Sample. The Total Sample consists of firms that report data for the variables employment (**EMPL**) *or* wage bill (**STAF**) and gross-output (**OPRE**), and TFP sample consists of firms that in addition report data on tangible fixed assets (**TFAS**) and materials (**MATE**) so that researchers can calculate total factor productivity for these firms.

The table clearly shows that the coverage improves over time for all countries until 2005 and is stable thereafter. In the case of Germany, on average, we observe worse coverage than in other countries which is explained by the under-representation of small firms in Germany. Only recently, EU harmonization laws made reporting by small firms compulsory also in Germany. There are other countries (Finland, Ireland, the Netherlands, if we were to use employment, and also Latvia, Luxembourg, for the data needed for TFP calculation), with problems similar to Germany, but overall we cover 70-80 percent of the real economy in manufacturing in the bulk of European countries. Denmark and the UK perform worse in the TFP sample, because firms do not report materials use in these countries. Missing percentages appear in some country-years because there is no data available for the corresponding measure in Eurostat for the manufacturing sector.

Table [D.1.5](#) reports the firm size distribution in the manufacturing sector. In the table, each panel presents the share of economic activity (gross-output and employment) accounted for by firms belonging in three size categories in a randomly picked year 2006. In each panel, the first three rows report the measures from Orbis-Amadeus and the next three are the corresponding numbers from Eurostat SBS data. Each column is a different country. Row entries denote the fraction of total economic activity accounted for by firms belonging to each size class. As for the total economy in terms of number of firms, we match well the official statistics in terms of the size distribution of economic activity undertaken in manufacturing sector.

Table D.1.1: COVERAGE OF THE MANUFACTURING SECTOR BASED ON GROSS OUTPUT, TOTAL SAMPLE

Year	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU
1999	0.2	0.75		0.54	0.21	0.04	0.83	0.75	0.3	0.64	0.6	0.95		0.84
2000	0.29	0.8		0.63	0.36	0.29	0.9	0.77	0.34	0.76	0.6	0.93		0.88
2001	0.47	0.78		0.73	0.5	0.31	0.92	0.78	0.36	0.79	0.68			0.91
2002	0.61		0.57	0.7	0.51	0.34	0.93	0.8	0.37	0.82	0.66			0.97
2003	0.59	0.81	0.58	0.77	0.57	0.34	0.93	0.79	0.39	0.79	0.66	0.92		0.84
2004	0.47	0.8	0.72	0.84	0.64	0.41	0.97	0.79	0.41	0.83	0.62	0.73		0.91
2005	0.45	0.8	0.83	0.86	0.9	0.43	0.95	0.78	0.41	0.82	0.7	0.69		0.88
2006	0.67	0.78	0.89	0.81	0.73	0.44	0.94	0.83	0.4	0.84	0.72	0.66		0.91
2007	0.72	0.79	0.92	0.88	0.77	0.43	0.96	0.81	0.45	0.87	0.69	0.68		0.88
2008	0.77	0.78	0.99	0.8	0.64	0.43	0.95	0.85	0.49	0.9	0.67	0.64	0.92	0.81
2009	0.72	0.74	0.96	0.85	0.6	0.36	0.92	0.87	0.46	0.89	0.81	0.51	0.99	0.9
2010	0.78	0.78	0.97	0.94	0.6	0.38	0.92	0.9	0.47	0.92	0.84	0.47	0.95	0.84
2011	0.8	0.77	0.89	0.9	0.57	0.43	0.95	0.85	0.5	0.96	0.89	0.45	0.99	0.77
2012	0.76	0.77	0.85	0.87	0.48	0.43	0.96	0.83	0.51	0.95	0.83	0.4	0.98	0.86
Average	0.59	0.78	0.83	0.79	0.58	0.36	0.93	0.81	0.42	0.84	0.71	0.67	0.97	0.87

Year	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
1999	0.12	0.61	0.13		0.56	0.19	0.6	0.51	0.59	0.66	0.6	0.3	0.32
2000	0.27	0.66	0.15		0.56	0.19	0.6	0.56	0.65	0.61	0.67	0.77	0.42
2001	0.38	0.65	0.17		0.63	0.2		0.62	0.63	0.76	0.73	0.83	0.54
2002	0.36	0.71	0.35		0.62	0.23	0.75	0.57	0.71	0.74	0.76	0.9	0.62
2003	0.33	0.7	0.43	0.46	0.61	0.26	0.68	0.59	0.73	0.83	0.74	0.92	0.83
2004	0.36	0.73	0.4	0.47	0.73	0.27	0.72	0.71	0.75	0.96	0.77	0.89	0.88
2005	0.29	0.77	0.41	0.43	0.79	0.24	0.69	0.67	0.91	0.95	0.75	0.91	0.78
2006	0.43	0.79	0.44	0.3	0.8	0.31	0.75	0.7	0.93	0.84	0.79	0.91	0.78
2007	0.42	0.79	0.57	0.33	0.81	0.32	0.76	0.74	0.92	0.91	0.78	0.91	0.76
2008	0.5	0.9	0.51	0.21	0.85	0.29	0.69	0.6	0.94	0.9	0.74	0.92	0.99
2009	0.47	0.86	0.57	0.24	0.81	0.3	0.88	0.74	0.92	0.97	0.87	0.91	0.95
2010	0.58	0.87	0.52	0.92	0.83	0.31	0.83	0.69	0.93	0.93	0.88	0.99	0.94
2011	0.57	0.89	0.5	0.9	0.87	0.3	0.83	0.6	0.93	0.91	0.85	0.83	0.98
2012	0.4	0.86	0.5		0.83	0.31	0.88	0.62	0.92	0.93	0.78	0.97	0.91
Average	0.39	0.77	0.40	0.47	0.74	0.27	0.74	0.64	0.82	0.85	0.77	0.85	0.76

NOTES: Table D.1.1 presents ratios calculated based on different measures of economic activity based on gross-output for the EU countries. The country codes within these classifications are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Percentages correspond to the ratio of aggregation of the corresponding measure over all firms in Orbis-Amadeus sample to the corresponding number in Eurostat sample. The percentages based on gross-output are obtained by comparison of our data and Eurostat SBS data, which includes self-employed workers. Those percentages are produced for “Total Sample,” which consists of firms that report data with positive values of gross-output.

Table D.1.2: COVERAGE OF THE MANUFACTURING SECTOR BASED ON GROSS OUTPUT, TFP SAMPLE

Year	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	HR	HU
1999		0.71		0.07	0.01	0.01	0.81	0.75	0.25	0.6			0.72
2000	0.01	0.76		0.19	0.04	0.01	0.88	0.76	0.27	0.72			0.72
2001	0.11	0.74		0.45	0.17	0.01	0.89	0.77	0.29	0.71			0.78
2002	0.27		0.51	0.54	0.22	0.01	0.9	0.79	0.3	0.78			0.83
2003	0.3	0.78	0.52	0.73	0.27		0.6	0.78	0.31	0.75			0.7
2004	0.33	0.77	0.45	0.82	0.28	0.04	0.56	0.78	0.32	0.78			0.82
2005	0.33	0.77	0.58	0.81	0.42	0.05	0.56	0.77	0.32	0.76			0.82
2006	0.55	0.75	0.75	0.79	0.39	0.05	0.56	0.77	0.31	0.8			0.85
2007	0.5	0.76	0.85	0.87	0.42	0.05	0.56	0.77	0.32	0.81			0.84
2008	0.55	0.76	0.9	0.79	0.34	0.05	0.55	0.72	0.34	0.85	0.15	0.9	0.74
2009	0.47	0.73	0.87	0.84	0.33	0.05	0.58	0.75	0.31	0.76	0.22	0.87	0.8
2010	0.54	0.76	0.88	0.87	0.37	0.05	0.59	0.74	0.31	0.83	0.22	0.83	0.76
2011	0.55	0.76	0.87	0.84	0.35	0.05	0.64	0.75	0.34	0.84	0.21	0.86	0.69
2012	0.38	0.76	0.83	0.83	0.31	0.04	0.65	0.72	0.35	0.82	0.2	0.86	0.76
Average	0.38	0.76	0.73	0.67	0.28	0.04	0.67	0.76	0.31	0.77	0.2	0.86	0.77

Year	IE	IT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
1999		0.6			0.07	0.47	0.39	0.51	0.66	0.21		0.02
2000	0.03	0.66		0.03	0	0.47	0.43	0.46	0.6	0.22		0.06
2001	0.03	0.65		0.04	0.08		0.46	0.46	0.75	0.25	0.08	0.23
2002	0.05	0.7		0.05	0.1	0.6	0.35	0.56	0.74	0.27	0.79	0.55
2003	0.01	0.69	0.17	0.06	0.11	0.56	0.36	0.69	0.82	0.26	0.78	0.83
2004	0.01	0.73	0.1	0.05	0.13	0.68	0.43	0.71	0.95	0.29	0.85	0.86
2005		0.76	0.25	0.04	0.09	0.66	0.42	0.87	0.94	0.28	0.89	0.76
2006		0.78	0.24	0.04	0.1	0.71	0.44	0.91	0.83	0.31	0.9	0.73
2007		0.78	0.3	0.05	0.11	0.72	0.47	0.9	0.9	0.31	0.9	0.71
2008		0.88	0.19	0.05	0.09	0.65	0.39	0.93	0.89	0.28	0.92	0.78
2009		0.84	0.22	0.04	0.02	0.84	0.5	0.91	0.96	0.34	0.9	0.75
2010	0.01	0.84	0.87	0.06	0.05	0.79	0.46	0.92	0.92	0.34	0.97	0.75
2011	0.06	0.87	0.8	0.06	0.01	0.79	0.4	0.92	0.9	0.32	0.8	0.93
2012	0.06	0.84		0.05		0.83	0.42	0.91	0.92	0.27	0.94	0.88
Average	0.03	0.76	0.35	0.05	0.07	0.68	0.42	0.76	0.84	0.28	0.81	0.63

NOTES: Table D.1.2 presents ratios calculated based on different measures of economic activity based on gross-output for the EU countries. The country codes within these classifications are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Percentages correspond to the ratio of aggregation of the corresponding measure over all firms in Orbis-Amadeus sample to the corresponding number in Eurostat sample. The percentages based on gross-output are obtained by comparison of our data and Eurostat SBS data, which includes self-employed workers. Those percentages are produced for “TFP Sample,” which consists of firms that report positive values on employment or wage bill, tangible fixed assets, gross-output, and materials.

Table D.1.3: COVERAGE OF THE MANUFACTURING SECTOR BASED ON EMPLOYMENT, TOTAL SAMPLE

Year	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU
1999							0.54	0.33	0.51	0.56				0.03
2000				0.62			0.81	0.57	0.41	0.57	0.58			0.05
2001				0.72			0.85	0.61	0.43	0.59	0.58			0.22
2002				0.77			0.88	0.67	0.46	0.59				0.09
2003				0.71			0.83	0.68	0.46	0.68	0.56			0.17
2004	0.7	0.9	0.77	0.53	0.63		0.84	0.67	0.48	0.67	0.54	0.49		0.32
2005	0.71	0.87	0.73	0.58	0.66		0.86	0.69	0.49	0.63	0.55	0.49		0.42
2006	0.73	0.79	0.92	0.75	0.65	0.68	0.7	0.5	0.62	0.56	0.43			0.82
2007	0.76	0.8	0.95	0.77	0.74	0.68	0.69	0.53	0.67	0.58	0.43			0.66
2008	0.77	0.82	0.84	0.65	0.74		0.78	0.48	0.73	0.63				0.89
2009	0.65	0.78	0.83	0.82	0.73	0.58	0.74	0.79	0.44	0.45	0.58			0.89
2010	0.72	0.79	0.83	0.94	0.66	0.56	0.86	0.84	0.45	0.49	0.61			0.89
2011	0.72	0.81	0.84	0.95	0.66	0.56	0.85	0.87	0.49	0.44	0.63			0.89
2012	0.69	0.82	0.94	0.86	0.61	0.54	0.82	0.82	0.35	0.63			0.83	0.92
Average	0.72	0.80	0.88	0.77	0.66	0.61	0.83	0.71	0.46	0.57	0.58	0.46	0.83	0.49

Year	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
1999		0.44		0.31		0.17			0.07		0.5		
2000		0.52	0.11	0.36	0.43	0.25			0.08	0.76	0.57	0.57	0.23
2001		0.53	0.14	0.33	0.48	0.19			0.07	0.71	0.58	0.53	0.27
2002		0.6	0.29		0.51	0.32			0.08	0.74	0.59	0.74	0.33
2003		0.59		0.01	0.52	0.52	0.69		0.08	0.8	0.62	0.8	0.4
2004		0.53	0.41	0.01	0.76	0.44	0.68	0.5	0.08	0.85	0.64	0.8	0.48
2005		0.43	0.42	0.2	0.79	0.4	0.77	0.51	0.07	0.92	0.65	0.8	0.69
2006		0.52	0.48	0.02	0.82	0.45	0.77	0.51	0.81	0.76	0.66	0.84	0.76
2007		0.53	0.71	0.34	0.84	0.56	0.72	0.49	0.83	0.97	0.69	0.83	0.74
2008	0.53	0.63	0.69	0.5	0.87	0.52	0.75	0.52	0.89	0.91	0.75	0.84	0.81
2009	0.43	0.58	0.75	0.6	0.87	0.51	0.73		0.88	0.78	0.7	0.8	0.84
2010	0.41	0.58	0.79	0.68	0.89	0.54	0.72	0.82	0.89	0.93	0.74	0.91	0.85
2011	0.39	0.76	0.62	0.67	0.9	0.77	0.78	0.88	0.89	0.94	0.78	0.93	0.8
2012	0.23	0.75		0.59	0.9	0.69	0.82	0.72	0.9	0.96	0.77	0.92	0.14
Average	0.40	0.57	0.49	0.36	0.74	0.45	0.74	0.62	0.47	0.85	0.66	0.79	0.57

NOTES: Table D.1.3 presents ratios calculated based on different measures of economic activity based on employment for the EU countries. The country codes within these classifications are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), SK (Slovakia). Percentages correspond to the ratio of aggregation of the corresponding measure over all firms in Orbis-Amadeussample to the corresponding number in Eurostat sample. The percentages based on employment are obtained by the comparison of our data and Eurostat BD data, which excludes self-employed workers. Those percentages are produced for "Total Sample," which consists of firms that report data with positive values of employment.

Table D.1.4: COVERAGE OF THE MANUFACTURING SECTOR BASED ON EMPLOYMENT, TFP SAMPLE

Year	AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	HR	HU
1999							0.54	0.29	0.5				
2000				0.19			0.79	0.56	0.36	0.55			0.02
2001				0.42			0.83	0.6	0.38	0.57			0.03
2002				0.58			0.85	0.65	0.4	0.65			0.09
2003				0.65			0.58	0.66	0.4	0.65			0.08
2004	0.21	0.4	0.73	0.13	0.05		0.59	0.65	0.41	0.64			0.15
2005	0.25	0.39	0.69	0.19	0.06	0.61	0.67	0.42	0.61				0.28
2006	0.34	0.62	0.72	0.24	0.06		0.68	0.43	0.6				0.38
2007	0.36	0.63	0.68	0.75	0.25	0.06	0.67	0.44	0.64				0.77
2008	0.37	0.64	0.66	0.64	0.24		0.67	0.39	0.69	0.16			0.6
2009	0.34	0.61	0.61	0.8	0.24	0.07	0.46	0.71	0.36	0.43	0.16		0.65
2010	0.41	0.62	0.61	0.83	0.24	0.07	0.55	0.7	0.36	0.47	0.16		0.66
2011	0.38	0.63	0.75	0.85	0.25	0.08	0.55	0.72	0.4	0.42	0.16		0.66
2012	0.27	0.65	0.86	0.75	0.21	0.08	0.53	0.67	0.34	0.18	0.81	0.65	
Average	0.33	0.63	0.61	0.66	0.22	0.07	0.63	0.65	0.39	0.55	0.16	0.81	0.39

Year	IE	IT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
1999		0.44	0.14		0.03			0.07		0.23		
2000		0.51	0.18	0.01	0.04			0.08	0.76	0.25		0.04
2001		0.53	0.17	0.04	0.04			0.07	0.7	0.26		0.09
2002		0.6		0.04	0.04			0.07	0.73	0.27	0.68	0.26
2003		0.54		0.05	0.05	0.52		0.07	0.79	0.28	0.72	0.39
2004		0.52		0.04	0.07	0.63	0.27	0.08	0.84	0.31	0.74	0.44
2005		0.43		0.04	0.06	0.72	0.28	0.07	0.9	0.32	0.77	0.66
2006		0.52		0.04	0.05	0.73	0.3	0.79	0.72	0.33	0.81	0.7
2007		0.53	0.29	0.03	0.06	0.68	0.29	0.8	0.94	0.34	0.81	0.68
2008		0.6	0.41	0.03	0.05	0.71	0.32	0.86	0.89	0.38	0.8	0.72
2009		0.56	0.51	0.03	0.01	0.69	0.3	0.85	0.76	0.35	0.78	0.78
2010	0.01	0.57	0.59	0.04	0.01	0.68	0.19	0.86	0.91	0.37	0.88	0.73
2011	0.14	0.73	0.56	0.04		0.74	0.2	0.86	0.92	0.39	0.9	0.67
2012	0.14	0.73	0.48	0.04		0.78	0.11	0.86	0.93	0.31	0.88	
Average	0.10	0.56	0.37	0.04	0.04	0.69	0.25	0.46	0.83	0.31	0.80	0.51

NOTES: Table D.1.4 presents ratios calculated based on different measures of economic activity based on employment for the EU countries. The country codes within these classifications are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Percentages correspond to the ratio of aggregation of the corresponding measure over all firms in Orbis-Amadeus sample to the corresponding number in Eurostat sample. The percentages based on employment are obtained by the comparison of our data and Eurostat BD data, which excludes self-employed workers. Those percentages are produced for “TFP Sample,” which consists of firms that report positive values on employment or wage bill, tangible fixed assets, gross-output, and materials.

Table D.1.5: SIZE DISTRIBUTION IN THE MANUFACTURING SECTOR: TOTAL SAMPLE, 2006

PANEL A: GROSS OUTPUT		AT	BE	BG	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU	IE	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK
<b>Orbis-Amadeus</b>																												
1 to 19 employees	0.03	0.05	0.08	0.04	0.06	0.03	0.15	0.13	0.08	0.05	0.03	0.14	0.12	0.01	0.02	0.12	0.04	0.03	0.14	0.04	0.11	0.02	0.13	0.09	0.09	0.07	0.09	
20 to 249 employees	0.23	0.30	0.31	0.32	0.23	0.29	0.67	0.40	0.38	0.23	0.23	0.55	0.36	0.14	0.49	0.49	0.51	0.97	0.62	0.45	0.40	0.32	0.44	0.32	0.26	0.32	0.33	
250 + employees	0.74	0.66	0.61	0.64	0.70	0.68	0.18	0.47	0.54	0.72	0.74	0.31	0.52	0.85	0.49	0.40	0.45	0.00	0.25	0.51	0.49	0.67	0.43	0.59	0.65	0.61	0.58	
<b>Eurostat (SBS)</b>																												
0 to 19 employees	0.07	0.08	0.08	0.10	0.06	0.10	0.12	0.13	0.06	0.09	0.08	0.26		0.07	0.03	0.20	0.05	0.02	0.12	0.09	0.13	0.09	0.14	0.08	0.09	0.12	0.05	
20 to 249 employees	0.32	0.27	0.31	0.31	0.22	0.34	0.60	0.38	0.21	0.27	0.29	0.26		0.21	0.25	0.41	0.34	0.10	0.54	0.33	0.36	0.28	0.42	0.31	0.28	0.32	0.23	
250 + employees	0.60	0.65	0.61	0.59	0.72	0.55	0.28	0.49	0.73	0.63	0.63	0.48		0.72	0.72	0.38	0.61	0.88	0.33	0.59	0.51	0.62	0.43	0.60	0.63	0.56	0.72	
PANEL B: EMPLOYMENT																												
<b>Orbis-Amadeus</b>																												
1 to 19 employees	0.08	0.13	0.16	0.05	0.05	0.14	0.15	0.25	0.16	0.10	0.01	0.13	0.12	0.02	0.06	0.13	0.05	0.14	0.19	0.21	0.18	0.03	0.26	0.12	0.17	0.08	0.06	
20 to 249 employees	0.36	0.41	0.48	0.41	0.32	0.48	0.58	0.49	0.43	0.34	0.35	0.57	0.39	0.26	0.49	0.55	0.57	0.86	0.56	0.54	0.47	0.40	0.54	0.38	0.34	0.38	0.33	
250 + employees	0.56	0.45	0.36	0.53	0.63	0.38	0.27	0.26	0.41	0.56	0.64	0.30	0.50	0.72	0.46	0.32	0.38	0.00	0.25	0.24	0.35	0.57	0.20	0.50	0.48	0.55	0.61	
0 to 19 employees	0.17	0.18	0.16	0.19	0.15	0.15	0.17	0.31	0.14	0.19	0.19	0.50		0.20	0.10	0.40	0.15	0.09	0.18	0.25	0.20	0.21	0.32	0.12	0.18	0.18	0.10	
20 to 249 employees	0.38	0.37	0.49	0.37	0.32	0.40	0.53	0.43	0.34	0.34	0.38	0.28		0.36	0.43	0.38	0.49	0.30	0.52	0.42	0.42	0.38	0.49	0.40	0.33	0.36	0.34	
250 + employees	0.46	0.45	0.35	0.44	0.53	0.46	0.30	0.26	0.51	0.47	0.43	0.22		0.44	0.47	0.22	0.35	0.61	0.30	0.33	0.38	0.41	0.19	0.48	0.49	0.45	0.55	

NOTES: Table [D.1.5](#) presents the share of economic activity (gross output and employment). We pick 2006. In each panel, the first three rows report the measures from Orbis-Amadeus for “Total Sample,” which consists of firms that report data with positive values of the corresponding measure, the next three rows are same numbers from Eurostat’s SBS data. Row entries denote the fraction of total economic activity accounted for by firms belonging to each size class. Each column is a different country. The country codes within these classifications are as follows: AT (Austria), BE (Belgium), BG (Bulgaria), CZ (Czech Republic), DE (Germany), DK (Denmark), EE (Estonia), ES (Spain), FI (Finland), FR (France), GB (United Kingdom), GR (Greece), HR (Croatia), HU (Hungary), IE (Ireland), IT (Italy), LT (Lithuania), LU (Luxembourg), LV (Latvia), NL (Netherlands), NO (Norway), PL (Poland), PT (Portugal), RO (Romania), SE (Sweden), SI (Slovenia), and SK (Slovakia). Row entries in Eurostat-SBS are empty due to missing output/employment data for size classifications of our interest.



## D.2 Alternative Data Sets and Their Coverage

An alternative data set, CompNet, has recently been developed by the European Central Bank (ECB), covering selected European countries. CompNet is a sector-level database constructed by the ECB from similar sources as Orbis, working with the national central banks. Although it is based on firm-level data taken from business registries, CompNet reports aggregates by sector and country and does not provide the underlying firm-level data set.

Because there are no the firm-level data in CompNet, we cannot sum up the output of firms ourselves. Consequently, the comparison below is based on the number of firms and relative to the “best” (official) Eurostat data.

Tables D.2.1–D.2.2 show these statistics. Table D.2.1 shows the firm coverage of CompNet database and our database constructed from Orbis data relative to Eurostat, with respect to the number of firms in each country-year. We define the total economy as the overlapping sectors of our data with the Eurostat data. Each cell corresponds to the number of firms in the total economy from the relevant data source, relative to the number of firms given by Eurostat. Although this type of comparison is less informative, because many firms in Eurostat have zero employment (self-employed), we show it in order to be able to compare our data to the alternative sector-level database CompNet. Neither our database, nor CompNet, includes self-employed and hence a comparison based solely on the number of firms might be misleading. In spite of this caveat, our data captures the number of firms in the total economy well.

Table D.2.2 may be a more relevant comparison if one worries about “representativeness.” The table shows the size and sector distribution of firms in CompNet and BvD, compared to the Eurostat data.<sup>52</sup> Each cell corresponds to the share of the indicated category’s number of firms in the total economy from the relevant data source for the given country-year (%). For example, in Belgium, 88.9 percent of firms have less than 10 employees in our data and CompNet, whereas the official number from Eurostat is 96 percent. According to this table, our number of firms in terms of employment are very close to what is reported by Eurostat. In terms of sectors; based on our data for Belgium, 13.3 percent of the firms in

---

<sup>52</sup>CompNet and Eurostat numbers come from Tables 6 and 7 of [Force \(2014\)](#).

the total economy operates in the manufacturing sector similarly to the fraction found using the Eurostat data.<sup>53</sup>

Table D.2.1: COVERAGE RELATIVE TO EUROSTAT FOR NUMBER OF FIRMS IN TOTAL ECONOMY IN ORBIS AND COMPNET

COUNTRY	YEAR	COMPNET	BvD
Belgium	2008	26.5	65.2
Estonia	2007	65.9	97.1
France	2009	30.6	85.2
Germany	2008	3.1	63.6
Hungary	2007	3.6	37.9
Italy	2008	2.2	58.8
Poland	2007	1.2	12.3
Slovakia	2008	12.8	40.3
Slovenia	2007	28.4	19.5
Spain	2008	23.6	41.7

NOTES: Each cell corresponds to the number of firms in total economy from the relevant data source relative to the number of firms in Eurostat for the given country-year (%).

<sup>53</sup>We pick these countries for comparison because only these countries' statistics are reported in Tables 6 and 7 of [Force \(2014\)](#). Notice that Tables 6 and 7 of [Force \(2014\)](#) also report coverage numbers on Amadeus/Orbis, but their numbers are clearly different from our numbers reported here, indicating worse coverage. We suspect that this is an artifact of the way the Amadeus/Orbis data was put together by the authors of the [Force \(2014\)](#). As we detail in the current paper, for the best coverage a certain procedure has to be followed.



### D.3 Treatment of Various Account Types

In our merged data set, multiple accounts of different types within a given financial year are available *for the same company* (the same BvD ID Number). We tag those accounts as “duplicate accounts.” The number of firm-year observations with duplicate accounts is 607,839, constituting 0.24% of total observations in our merged data set. Such duplicate accounts arise for two reasons:

1. A company reports two accounts with the same BvD ID Number and different consolidation codes and the different values of financial and real variables for each consolidation code. The example is the company Ford Otosan Inc. shown in Figure D.3.1. This situation emerges because companies may switch the type of statements they report over time and when we combine different vintages of Orbis (and Amadeus) we might have such different accounts co-exist even in the same year.
2. A company reports two accounts with the same BvD ID Number and different consolidation codes but the same values of financial and real variables for either consolidation code.

Whenever we need a unique company-year observation we make the following choices. For Case 1, we give preference to unconsolidated accounts to avoid double-counting.

To guide us in resolving duplicates in Case 2, Figure D.3.2 presents distribution of duplicates by account type, each bar corresponding to the share of a given combination of two different accounts appearing in the same financial year. The highest fraction of duplicate accounts in the same year is the combination of LF&U (62.52%) followed by C1&U1 (32.24%). Figure D.3.3 represents the distribution of accounts by the length of time series, within the pairs from Figure D.3.2. Within each combination, the upper bar shows the proportion of years when the first account has the longer time series, the middle bar shows the proportion of years when the second account has the longer time series, and the lowest bar the frequency when both accounts in the combination have the same number of years. For example, 32.24% of duplicate accounts corresponds to firm-year observations with the combination of C1&U1 (Figure D.3.2). According to Figure D.3.3, within this pair, C1 have the longest time series in 87.2% of cases, U1 accounts have the longest time series in 9.61% cases, and in 3.19 % of

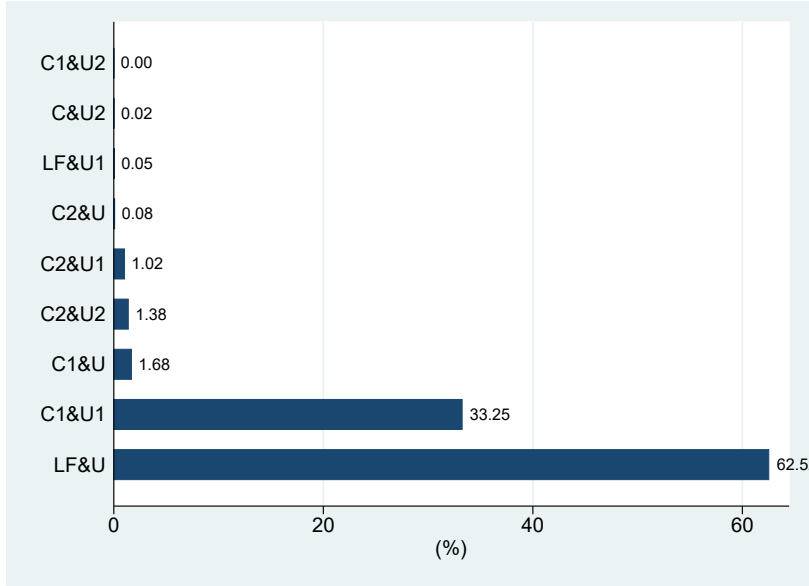
cases they have the same time series. To resolve the duplicates in Case 2, we either retain the accounts reporting the longer time series for a given company, or keep the unconsolidated account if it is reported the same number of years as the other available account type. We also check consolidation codes of duplicate accounts registered in “Historical Product” and verify the accuracy of this procedure.

Figure D.3.1: CONSOLIDATION IN BvD ACCOUNTS

Sector Code Account Type	<u>Parent Company, Headquarter</u>				
	<u>Koc Holding Inc.</u>				
			2910		
			C1		
	↙	↙	↓	↘	↘
	<b>Subsidiary 1</b>	<b>Subsidiary 2</b>	<b>Subsidiary 3</b>	<b>Subsidiary 4</b>	<b>Subsidiary 5</b>
	Ford Otosan Inc.	Opet Petroleum Inc.	Aygaz Inc.	Demir Export Inc.	Arcelik Inc.
Sector Code	2910	4930	4971	0710	2751
Account Type	C1 & U1	U1	C1	U1	C1
				↙	↘
				<b>Subsidiary 1</b>	<b>Subsidiary 2</b>
				<u>Beko Electronics Inc.</u>	<u>Arcelik-LG Inc.</u>
Sector Code				2611	2825
Account Type				C1	U1

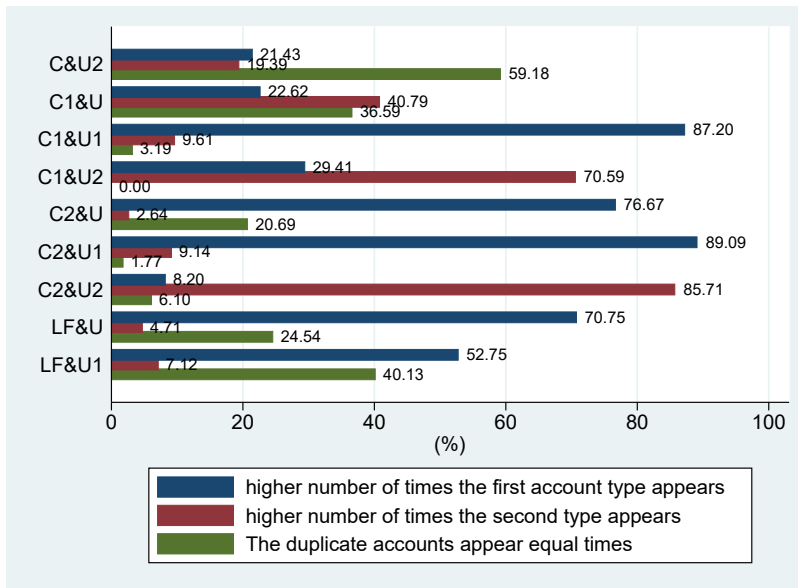
NOTES: The diagram illustrates an example of how the real activities of a headquarter and its subsidiaries are registered in BvD data. BvD registers the given firm in the sector to which largest portion of its revenue belongs to. The sector codes are classified according to NACE Revision 2, 4-digit Core Codes: 2910: Manufacture of Motor Vehicles; 4930: Retail sale of automotive fuel in specialised stores; 4971: Wholesale of solid, liquid and gaseous fuels and related products (4971); 0710: Mining of iron ores; 2751: Manufacture of electronic domestic appliances; 2611: Manufacture of electronic components; 2825: Manufacture of non-domestic cooling and ventilation equipment. The account types are as follows: C1–account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.), where the company headquarter has no unconsolidated account; U1–account of a company with no consolidated account. In this example, Ford Otosan Inc. reports both U1 and C1 accounts in our merged dataset because these accounts come from different vintages, therefore we drop C1 account to avoid from double-counting.

Figure D.3.2: THE DISTRIBUTION OF DUPLICATE ACCOUNTS BY ACCOUNT TYPE



NOTES: The figure represents distribution of duplicates by account type. In the figure, each bar corresponds to the share of a given combination of two different accounts appearing in the same financial year. BvD reports firm-level accounts in five types: C1: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.), where the company headquarter has no unconsolidated account, C2: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.) where the company headquarter also presents an unconsolidated account, U1: account of a company with no consolidated account, and U2: account of a company with a consolidated account. LF: accounts with financial information. For the firms with missing consolidation code, we derive 1-letter consolidation codes from the BvD account number and fill consolidation codes with either U or C.

Figure D.3.3: THE DISTRIBUTION OF DUPLICATE ACCOUNTS BY ACCOUNT TYPE



NOTES: The figure represents the distribution of duplicates by account type. In the figure, each bar within a given combination of account types corresponds to the share of the first duplicate account that appears higher times than the latter, the share of the latter duplicate account that appears higher times than the first, and the share of duplicate accounts that appear equal times, respectively. BvD reports firm-level accounts in five types: C1: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.), where the company headquarter has no unconsolidated account, C2: account of a company-headquarter of a group, aggregating all companies belonging to the group (affiliates, subsidiaries, etc.) where the company headquarter also presents an unconsolidated account, U1: account of a company with no consolidated account, and U2: account of a company with a consolidated account. LF: accounts with financial information. For the firms with missing consolidation code, we derive 1-letter consolidation codes from BvD Account Number, and fill consolidation code with U or C.

## D.4 Extra Material on Market Concentration

In this appendix we demonstrate that using the OECD STAN data to normalize the market shares of the largest firms delivers the same pattern of concentration as when we normalize by our Orbis data, provided we compute shares at 1-digit level as in the OECD STAN data using a subset of sectors and countries covered in the OECD STAN data. There are at least three differences from the baseline analysis based on solely our Orbis data as presented in the main text.

First, the OECD STAN data follows ISIC Revision 4 to disaggregate real variables at 2-digit and 1-digit industry levels. The OECD STAN data provides 2-digit industry level for the entire economy, the list of countries with non-missing data is relatively limited though. The OECD STAN data lacks 2-digit industry level data for Bulgaria, Croatia, Estonia, Great Britain, Greece, Latvia, Luxembourg, Poland, Portugal, Slovak Republic, and Spain, which is not acceptable for our purposes. Therefore, we compare the market shares based on the OECD STAN data and our Orbis data following *1-digit* industry level; see Table D.4.1 for the list of 1-digit industry codes.

Second, even at 1-digit level, the OECD STAN data lacks information for Bulgaria and Czech Republic, so we exclude them in the analysis.<sup>54</sup>

Third, the OECD STAN data includes annual measures of output, value added and its components, labour input, investment and capital stock. The OECD STAN data does not report “Turnover,” which corresponds to actual operating revenues in the year. It reports “Production” (Gross-Output) that represents the value of goods and/or services produced in a year, whether sold or stocked. Hence it might be smaller than actual operating revenues in a given period if all production is sold out with inventories from previous periods. In constructing market shares based on the OECD STAN data, we make use of gross-output instead of sales.

Keeping these differences in mind, we compute country-sector-year output shares to represent industry concentration, that uses our Orbis data, but is defined at 1-digit level,

---

<sup>54</sup>In our analysis, we focus on the sample of the OECD STAN countries that have non-missing information on “Production” at 1-digit level. These countries are Austria, Belgium, Croatia, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Italy, Latvia, Luxembourg, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, and Sweden.



comparable to the OECD STAN database. More specifically, we define measures of industry concentration using the operating revenue market shares of the top 4 (hereafter, MS4) or top 8 (hereafter, MS8) firms in a given 1-digit industry  $s1$ , country  $k$ , and year  $t$  relative to the total sectoral output in the  $s1, k, t$  cell, which is reported by either our Orbis data or the OECD STAN data. As explained in the main text, before constructing industry concentration measures, we categorize firms into two samples based on account type, available in Orbis. “*Sample 1*” comprises firms reporting *only one* type of account (unconsolidated (U) or consolidated (C)) in a given financial year, and “*Sample 2*” comprises firms reporting *both* unconsolidated and consolidated accounts in a given financial year.

In the following notation, the first argument in parenthesis defines the account type we use and the second argument denotes the sample of firms we aggregate from, and we omit the subscript  $s1, k, t$  for brevity:

- MS4(C,1) is based on consolidated accounts (C), using firms from *sample 1* that report only consolidated accounts. For this measure, the population of all firms is used in the denominator of market shares, so the denominator equals the aggregated output reported by all firms from *sample 1* and *sample 2* after consolidated accounts from *sample 2* are excluded to avoid double counting.<sup>55</sup>
- MS4(C,1)<sup>STAN</sup> is based on consolidated accounts (C), using firms from *sample 1* that report only consolidated accounts. For this measure, sectoral gross output *reported by the OECD STAN database* is used in the denominator.<sup>56</sup>

Figure D.4.1 reports the GDP-weighted average market concentration for manufacturing and services using these measures. Upper row (Panel A1 and Panel A2) compares the shares of top firms in the manufacturing sector, the bottom row (Panel B1 and Panel B2) depicts the shares in the services sector. In each row, the diagram on the left is based on the

---

<sup>55</sup>Numerically, the MS4(C,1) =  $\sum_{i \in F} y_i / Y$  where firm set  $F$  includes top 4 firms reporting consolidated accounts from *sample 1* in a given  $s1, k, t$  cell,  $y_i$  is firm-level revenue, and  $Y$  is aggregated output of all firms in  $s1, k, t$  from *sample 1* and *sample 2* after consolidated accounts from *sample 2* are excluded. Market share of the top 8 firms is computed similarly.

<sup>56</sup>Numerically, the MS4(C,1)<sup>STAN</sup> =  $\sum_{i \in F} y_i / Y^{\text{STAN}}$  where firm set  $F$  includes top 4 firms reporting consolidated accounts from *sample 1* in a given  $s1, k, t$  cell,  $y_i$  is firm-level revenue, and  $Y^{\text{STAN}}$  is gross output of  $s1, k, t$  as reported by the OECD STAN database. Market share of the top 8 firms is computed similarly.

normalization by aggregated firm revenue from our Orbis data, and on the right is based on the normalization by sectoral gross output from the OECD STAN data. Except for some disagreement in initial year, the patterns are very similar that gives us comfort about the quality of our data. Of note, the patterns which emerge when we aggregate to 1-digit sector level are fully consistent with those reported in the main text, based on more granular 2-digit sector shares. In particular, there is an increasing trend for concentration for manufacturing sector.

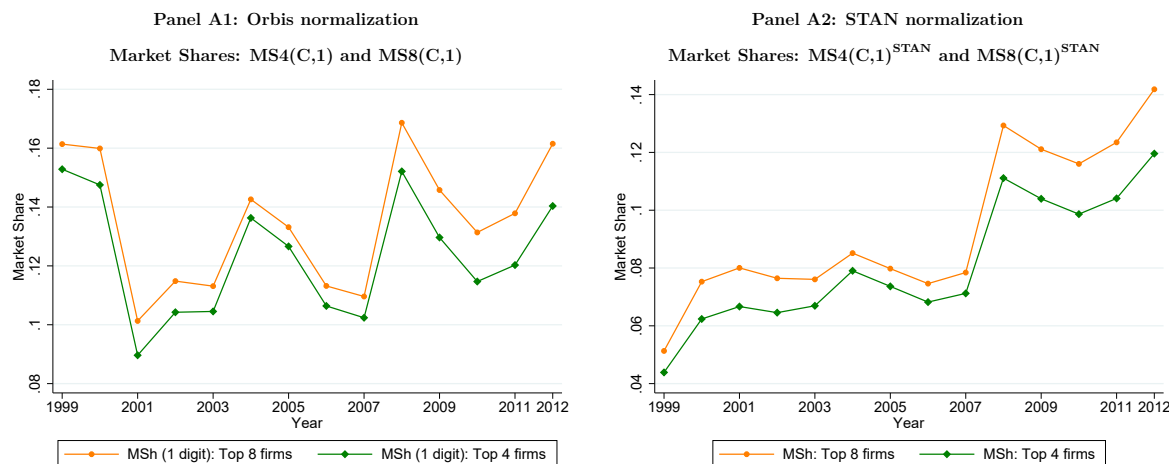
Table D.4.1: SECTORAL CLASSIFICATION OF ECONOMIC ACTIVITIES IN THE OECD STAN DATABASE

OECD STAN Codes	Definition	NACE Rev. 2 Level 2 Classification
A	Agriculture, forestry and fishing	01 to 03
B	Mining and quarrying	05 to 09
C	Manufacturing	10 to 33
D&E	Electricity, gas, steam and air conditioning supply (D); Water supply, sewerage, waste management and remediation (E)	35 to 39
F	Construction	41 to 43
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	45 to 47
H&J	Transportation and storage (H); Publishing, audiovisual and broadcasting activities (JA); Telecommunications (JB); IT (JC)	49 to 53
I	Accommodation and food service activities	58 to 63
K	Financial and insurance activities	55+56
L	Real-estate activities	66 to 66 68

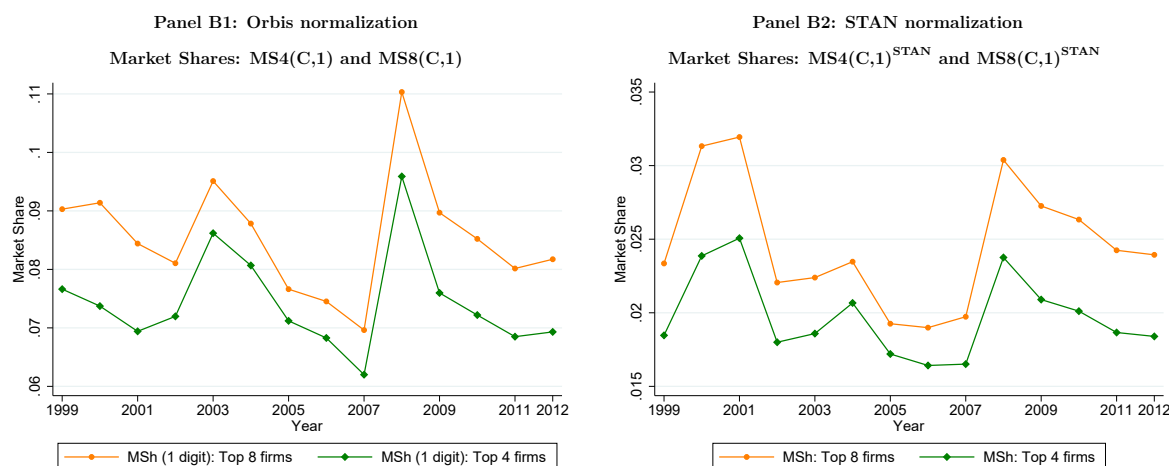
NOTES: The table lists 1-digit sector codes followed by the OECD STAN data. The OECD STAN data follows ISIC Revision 4 to disaggregate real variables at 1-digit industry level. The numbers available in the third column correspond to 2-digit sector codes based on NACE Revision 2, Level 2 Classification that are covered within a given 1-digit sector.

Figure D.4.1: CONCENTRATION IN EUROPE: MARKET SHARE OF TOP FIRMS, CONSOLIDATED STATEMENTS, NORMALIZATION BY ORBIS AND OECD-STAN DATA

**Panel A: Manufacturing, Companies reporting only Consolidated Statements (Sample 1)**



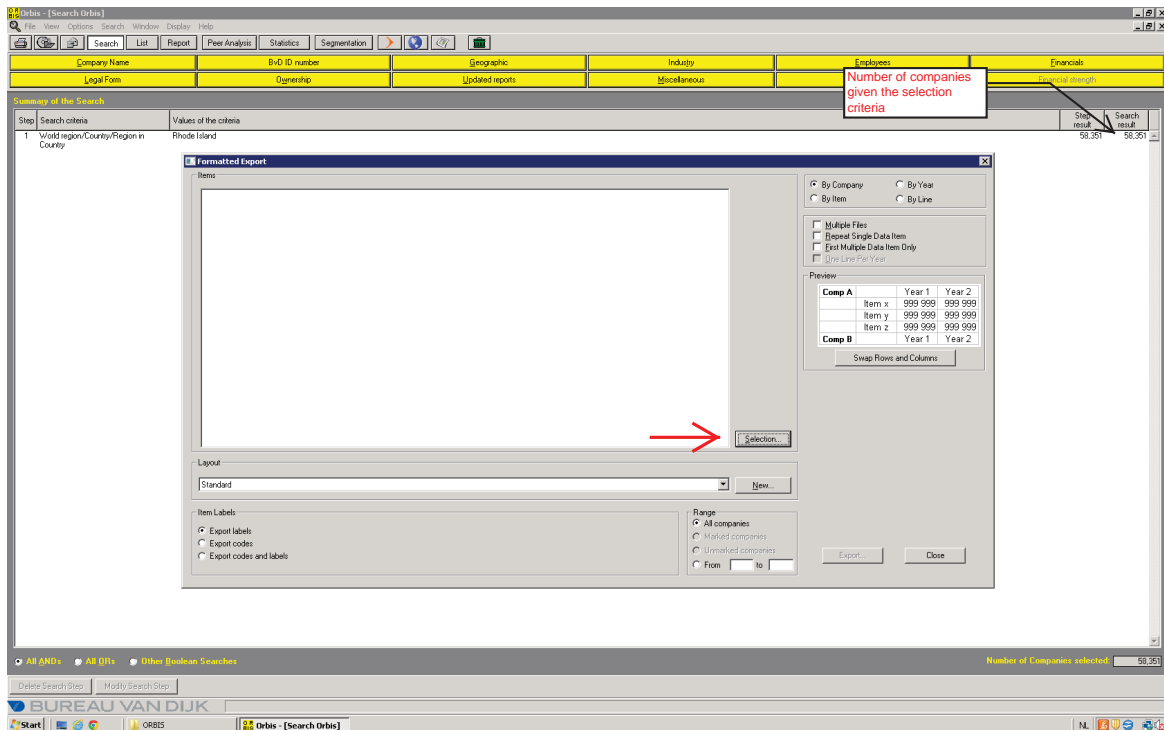
**Panel B: Services, Companies reporting only Consolidated Statements (Sample 1)**

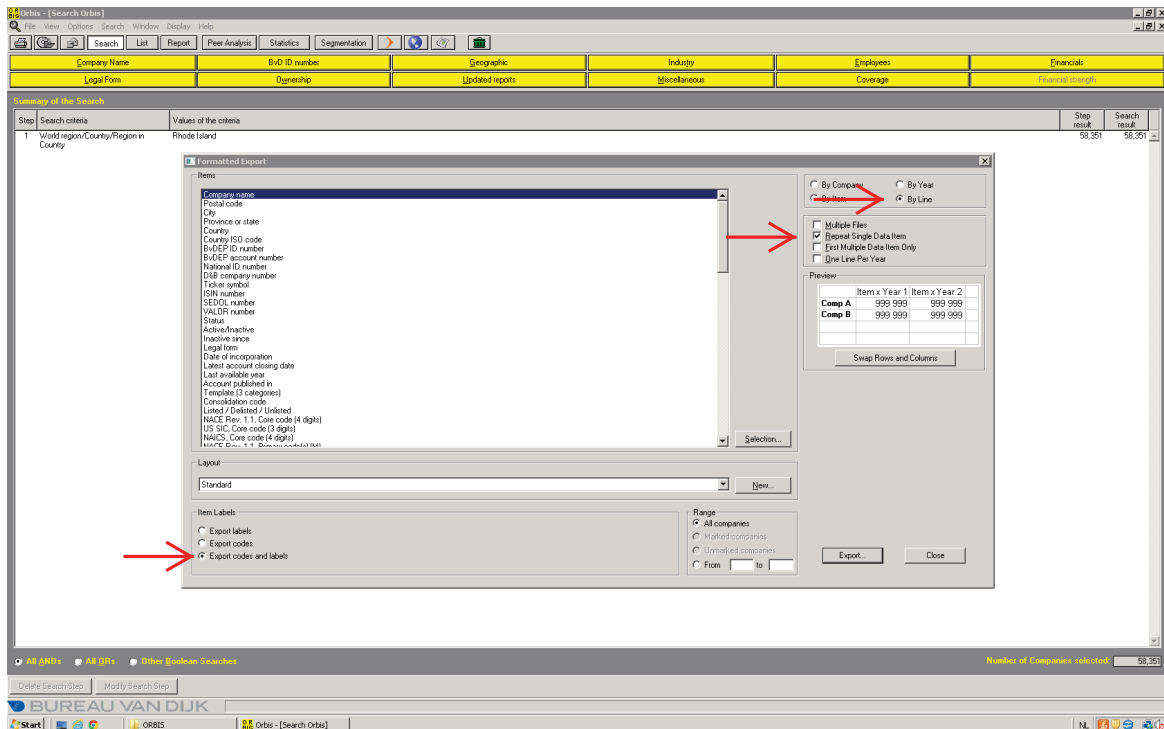
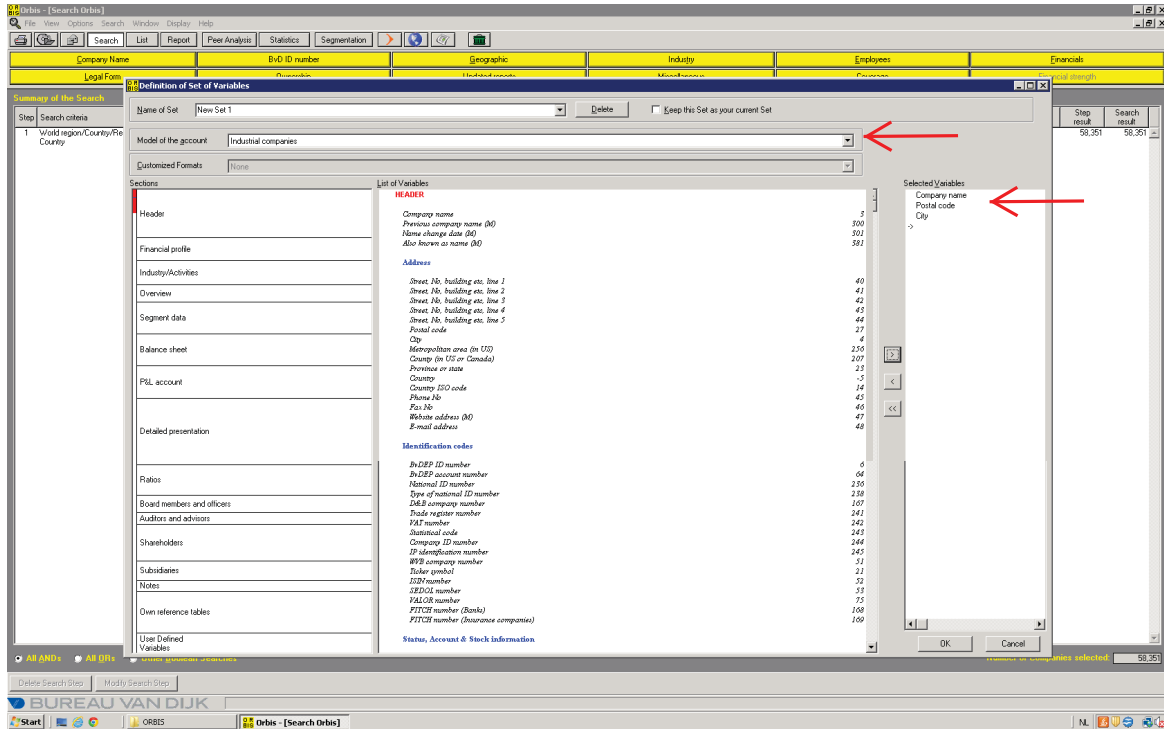


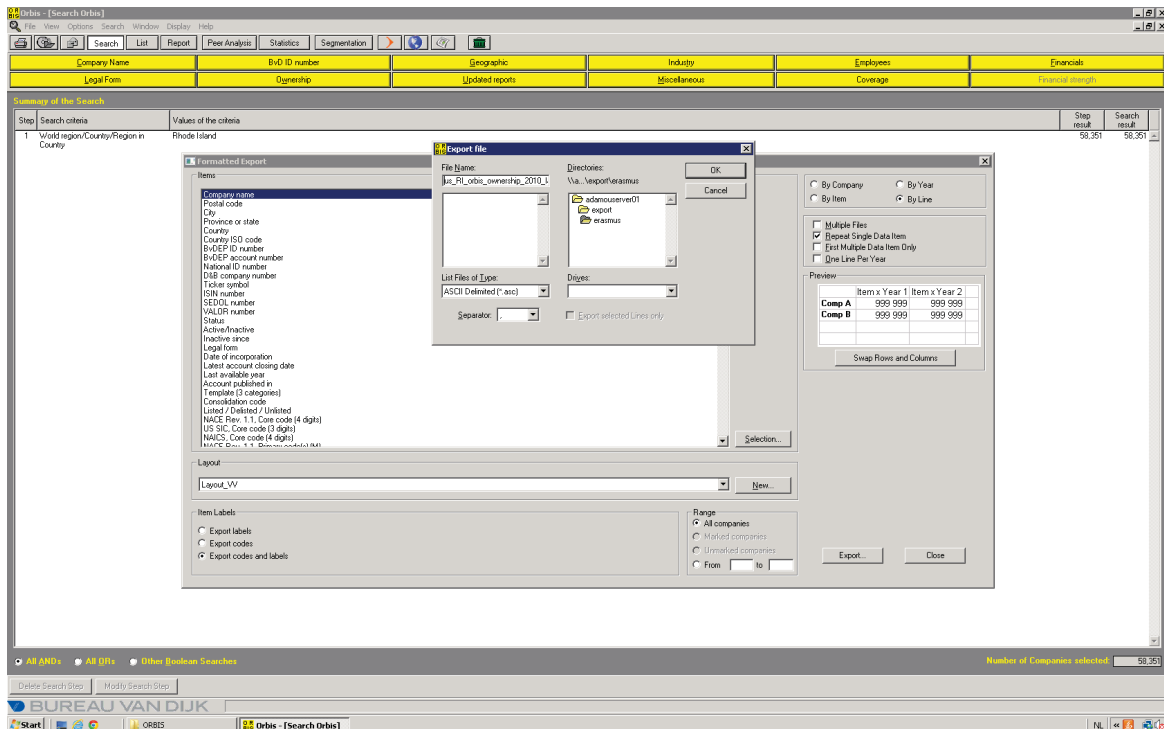
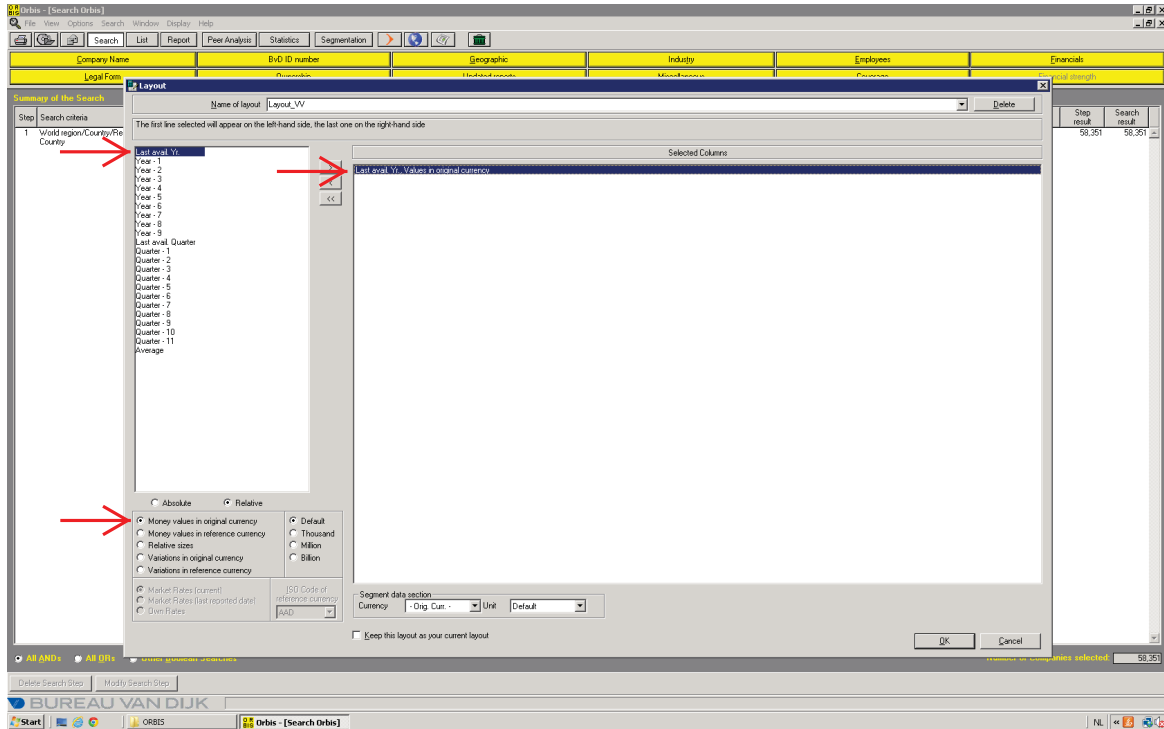
NOTES: The figure plots market concentration measures for the manufacturing industry and the services industry. The services industry is defined as “non-financial market services,” including “accommodation and food services,” “information and communication,” “real estate activities,” “professional, scientific and technical activities,” and “administrative and support services activities.” We measure country-sector-year market concentration by the market share of largest eight foreign (domestic) firms in terms of firm-level operating revenue from our Orbis dataset in the given country-sector-year, relative to the overall country-sector-year output. The output of the largest firms is determined from the subset of companies reporting *only* consolidated financial statements, drawn from *sample 1*. Under “Orbis normalization”, the output of the largest firms is normalized by aggregated firm revenue from our Orbis data, and under “STAN normalization”, it is normalized by sectoral gross output from the OECD STAN data. In all graphs, we plot “EU-weighted” average shares over the period 1999–2012, using a given country’s GDP weight in total sample, in the corresponding year.

## E BvD Disks Interface

### E.1 Formatted Export Interface of Older BvD Disks

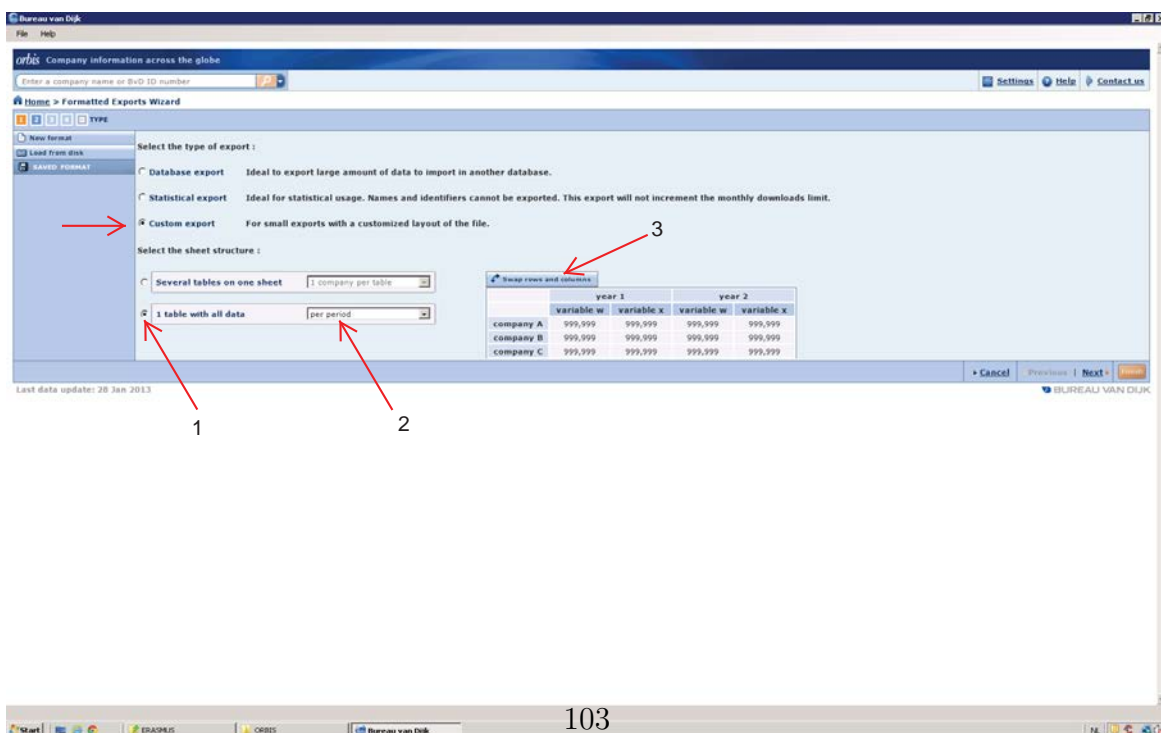
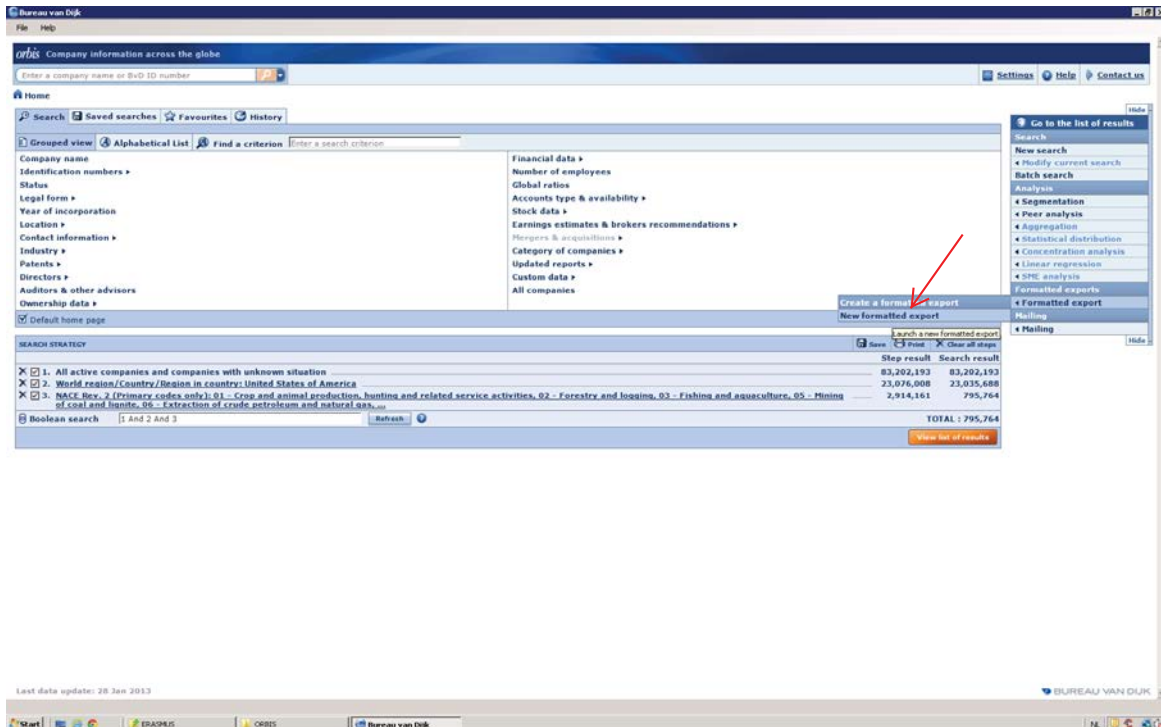




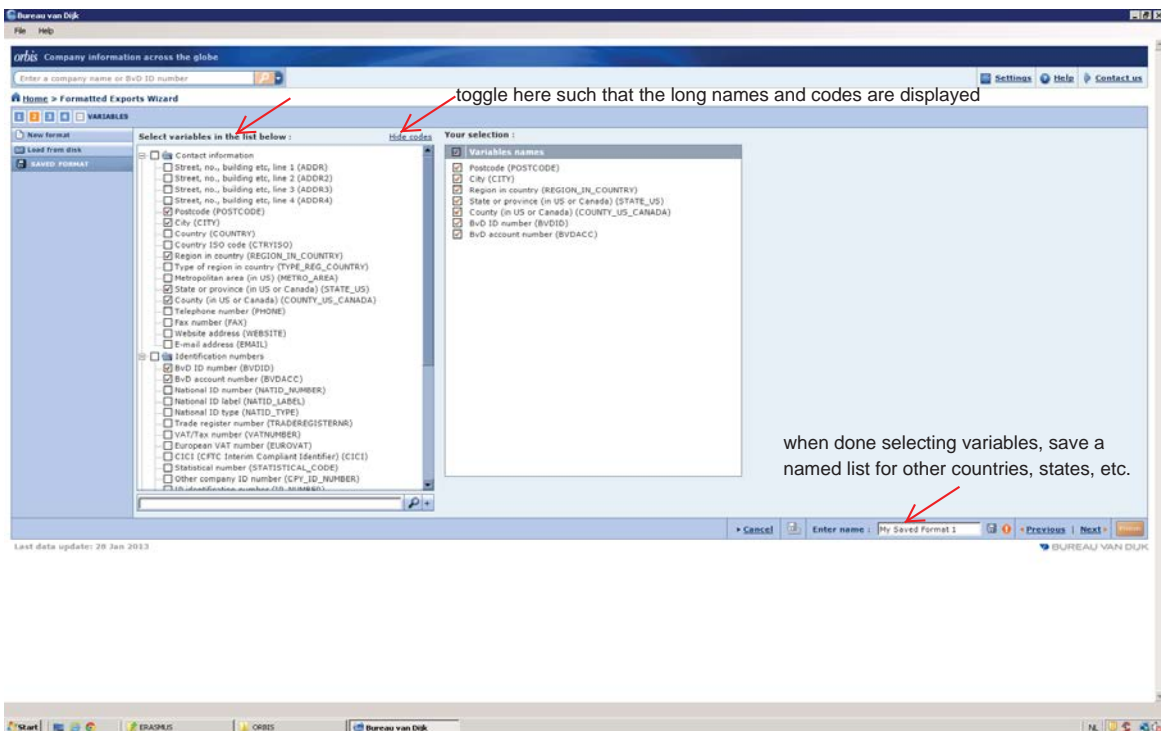
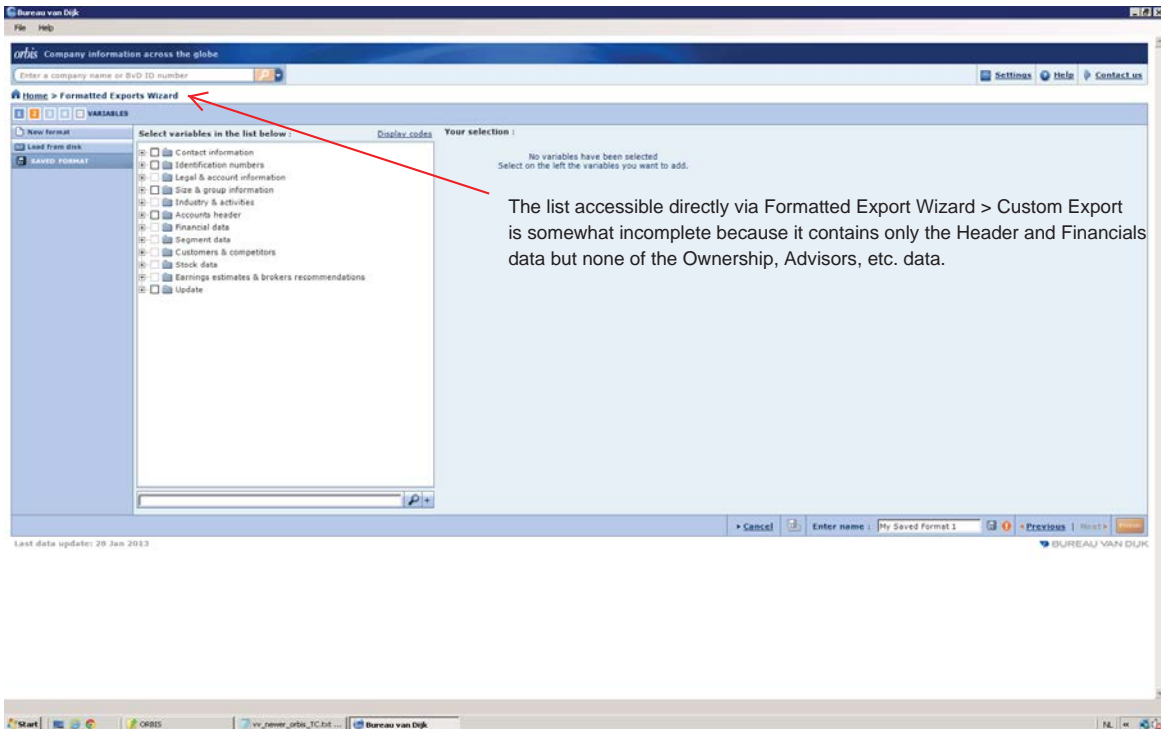


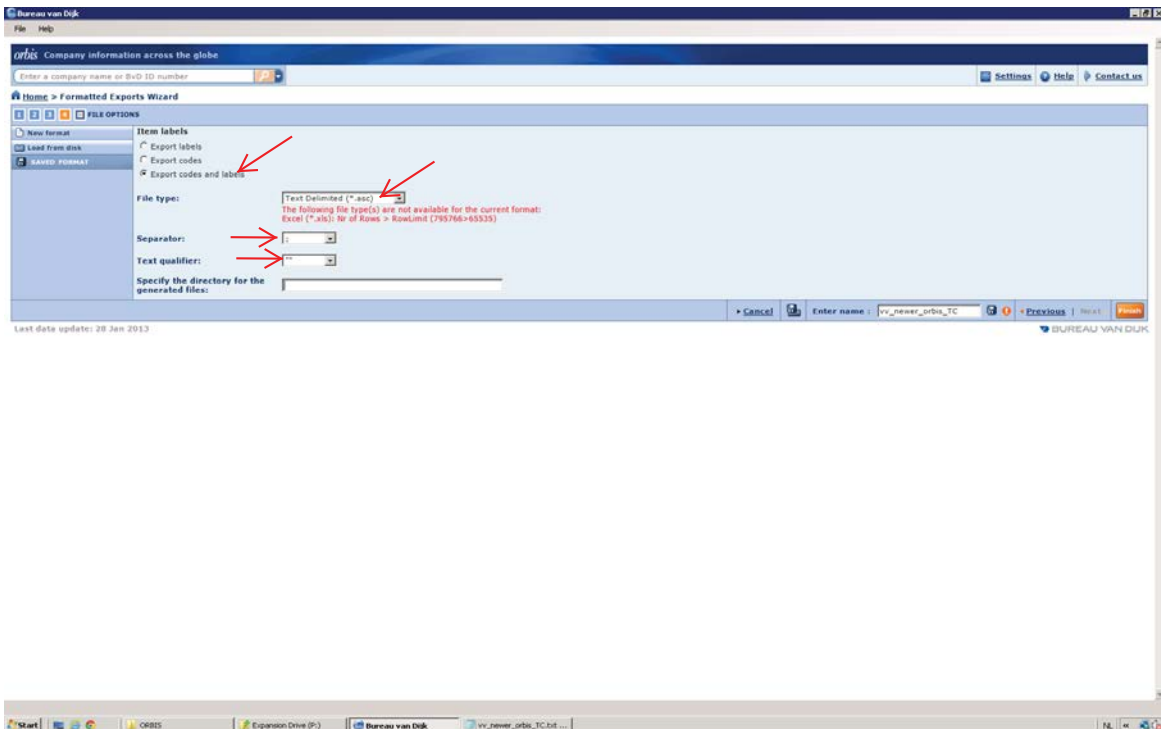
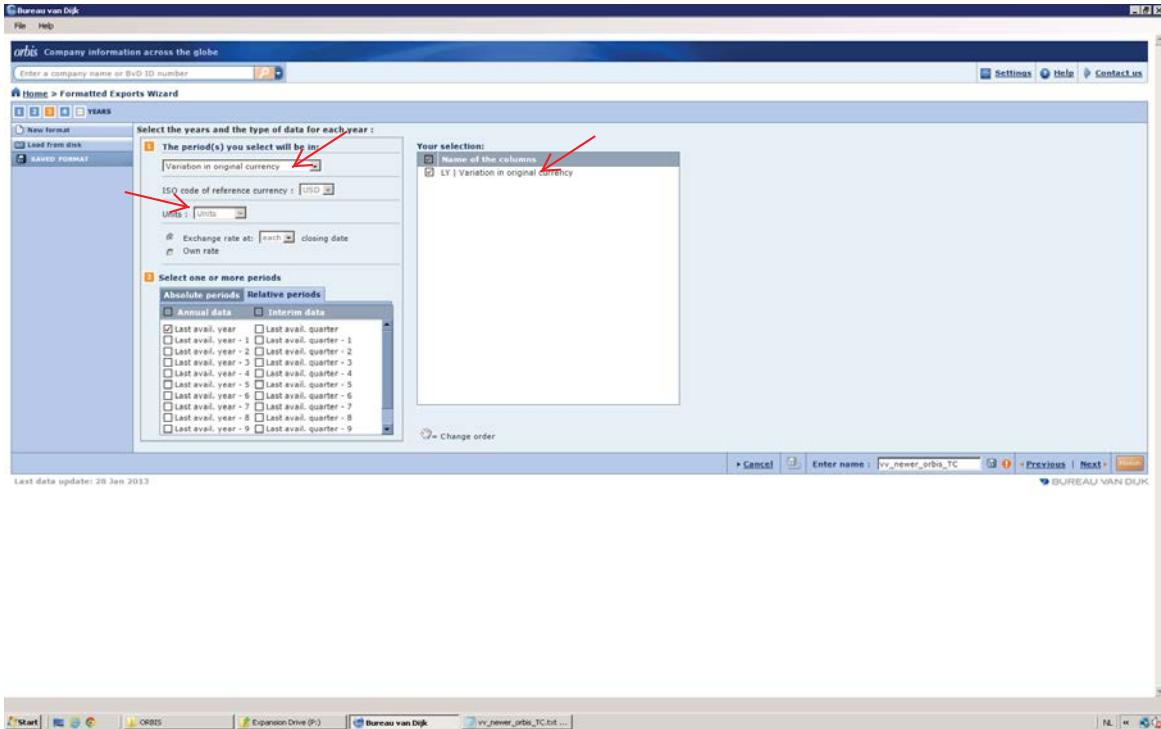


## E.2 Custom Export Interface of Newer BvD Disks



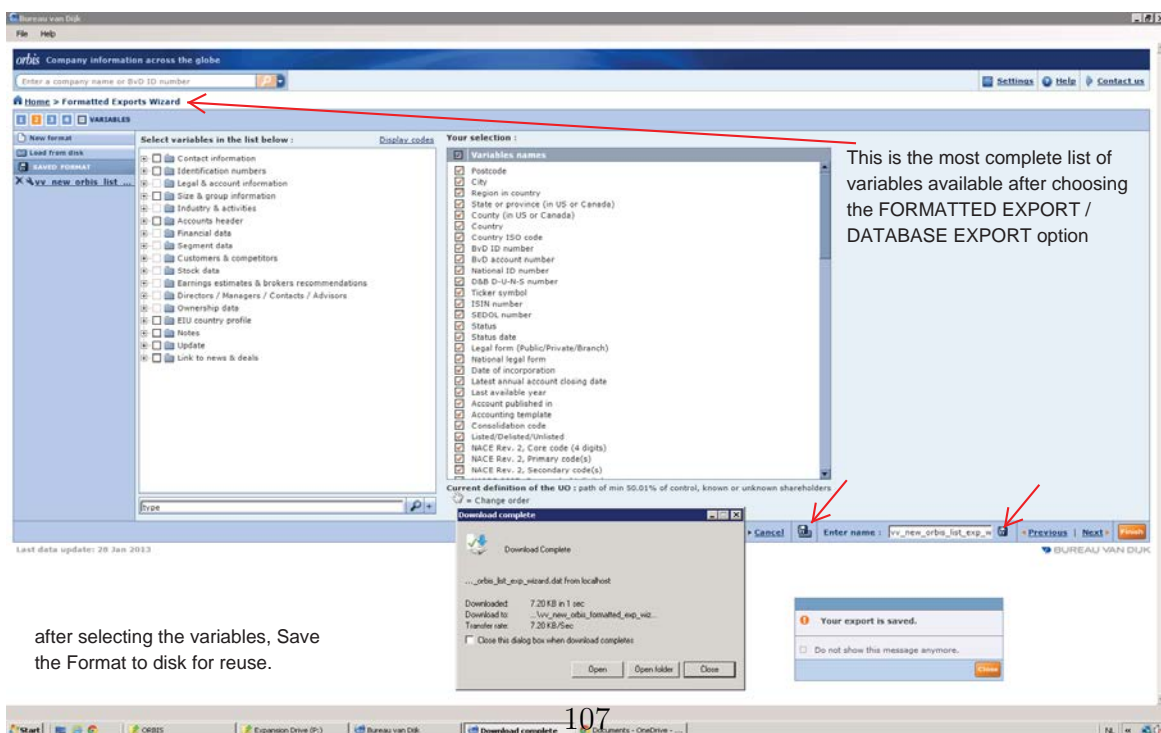
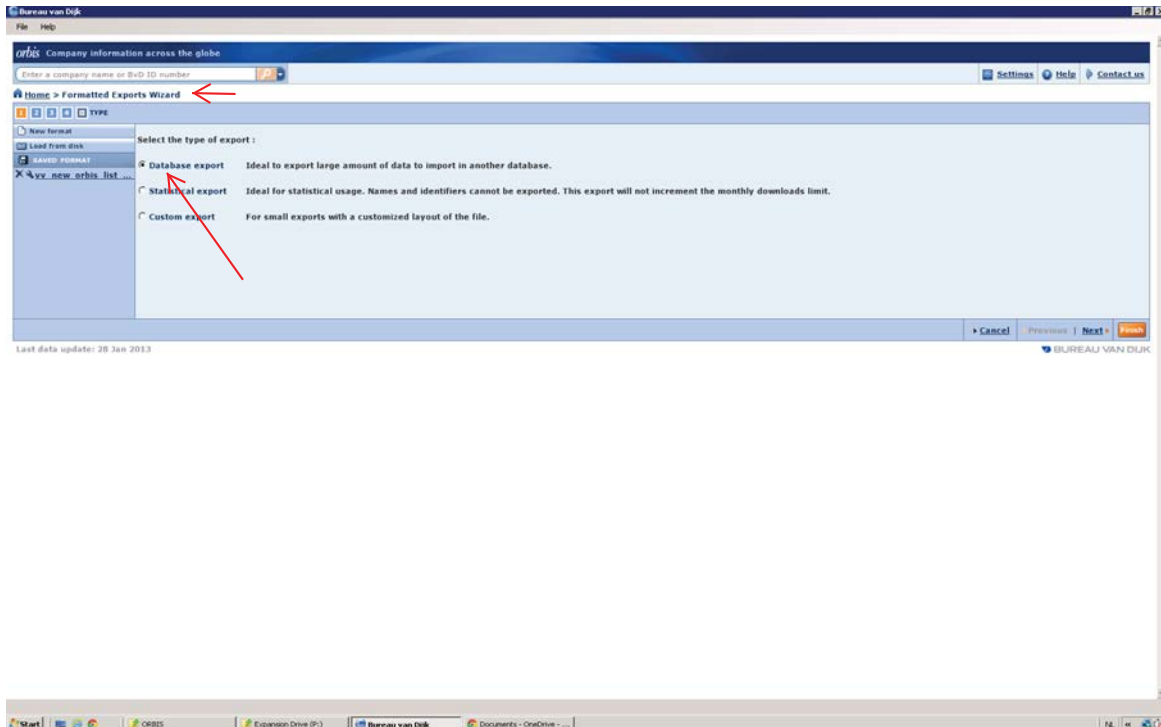


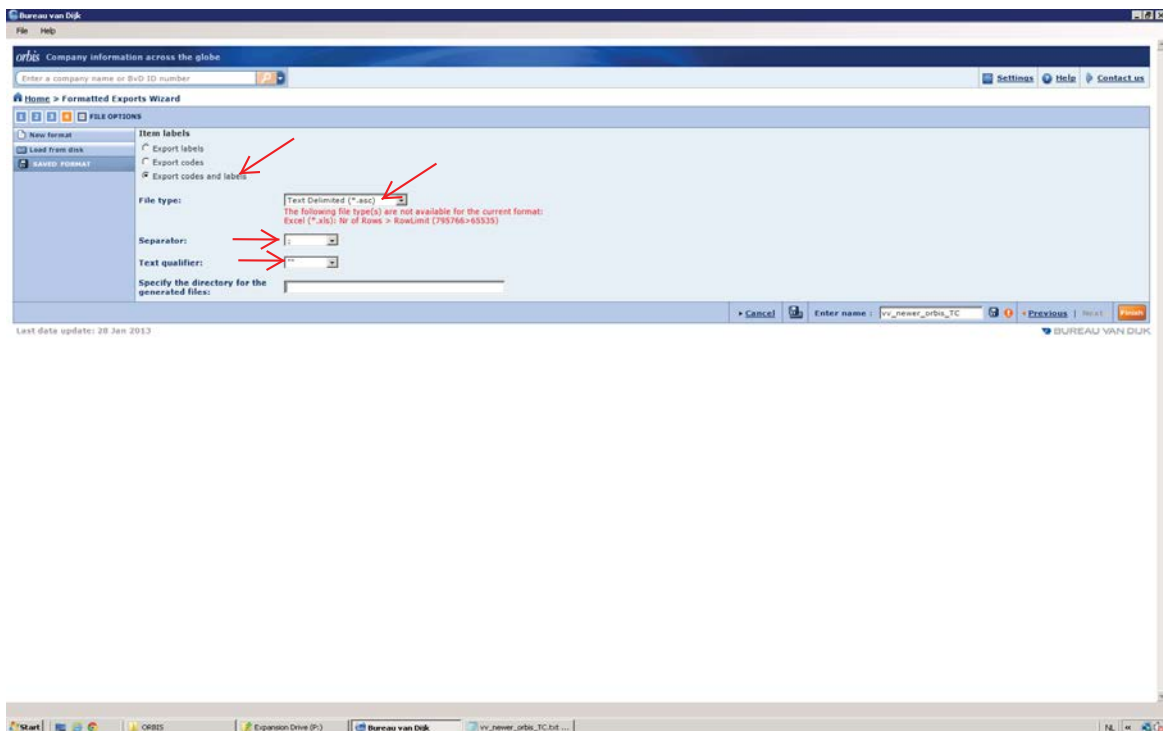
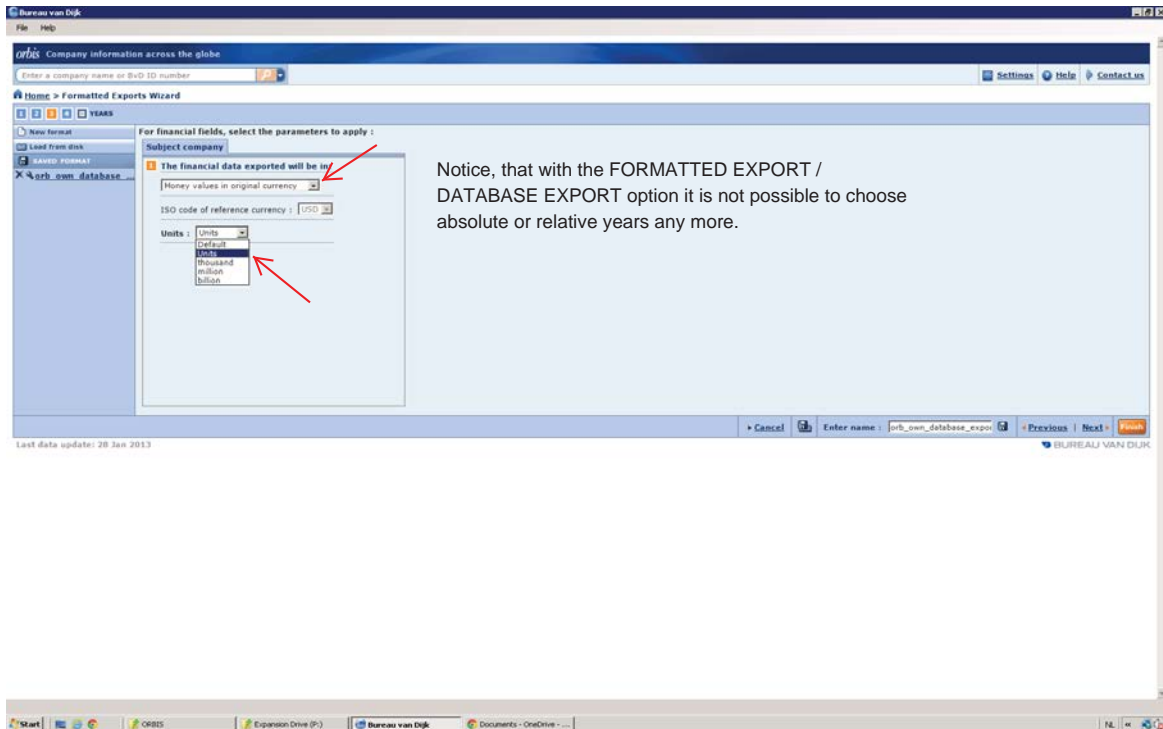






## E.3 Database Export Interface of Newer BvD Disks







## E.4 List Export Interface of Newer BvD Disks

