5 Construction of the data "matrices" underlying the impact assessment

5.1. Introduction and overview

458. Having good data on the location of profit and economic activity of multinational enterprises (MNEs) is key to assessing the implications of international corporate tax reforms, such as the Pillar One and Pillar Two proposals currently being discussed by the OECD/G20 Inclusive Framework on Base Erosion and Profit Shifting (BEPS). However, while a range of data sources provide valuable insights on the profit and activities of MNEs, no existing data source is sufficiently comprehensive in its geographic coverage and in terms of variables available to be used in isolation for a comprehensive reform impact assessment covering all 137 jurisdictions in the OECD/G20 Inclusive Framework on BEPS (Inclusive Framework).

459. Against this background, the OECD Secretariat has undertaken to combine a range of existing data sources into a consistent framework, which serves as a central instrument supporting the impact assessment analysis in this report. The framework consists of a set of four matrices: a profit matrix, focusing on the location of the profit of MNEs across jurisdictions, and three matrices focusing on indicators of the economic activity of MNEs (turnover, tangible assets and payroll). Each matrix contains data spanning more than 200 jurisdictions (each jurisdiction corresponding to a matrix *row*) and broken down across more than 200 jurisdictions of ultimate parent of the MNE considered (each jurisdiction of ultimate parent being a matrix *column*). Each matrix therefore takes the form of a square table with more than 200 rows and more than 200 columns. For example, the France-United States cell in the profit matrix would represent the profit of US MNEs (i.e. MNEs with an ultimate parent in the United States) in France.

460. The matrices combine data from a range of sources, and build on earlier efforts to map the profit and activity of MNEs for the analysis of profit shifting (Tørsløv, Wier and Zucman, 2018[1]) and the study of global value chains (GVCs) (Cadestin et al., 2018[2]). A primary source of data used in the matrices is the newly available anonymised and aggregated Country-by-Country Report (CbCR) data,[1] which have been collected as a result of the implementation of the 2015 BEPS Action Plan and were published for the first time by the OECD in July 2020 (OECD, 2020[3]). Other sources include the ORBIS database of firm-level financial accounts (in jurisdictions where ORBIS coverage is good), the OECD AMNE database (which includes data from the Eurostat FATS database and from the US Bureau of Economic Analysis) and the OECD Analytical AMNE database (Cadestin et al., 2018[2]), which builds upon and complements the OECD AMNE database. The data considered focus essentially on year 2016, which is the latest available year across all the data sources used.

461. These various data sources complement each other as they have different geographic coverage and include different variables, meaning that the combined dataset is richer than any data source taken individually. These sources also have substantial overlap in their coverage. This overlap is used to benchmark sources against each other, in order to address the limitations of individual data sources and ensure the overall consistency of the approach, as further discussed below. The methodology aims to

make data across the four matrices as comparable as possible, so as to enable the joint use of the matrices (e.g. using simultaneously the profit and turnover matrices to compute average profitability). To this end, efforts have been made to rely as much as possible on comparable data sources for the same cell across the different matrices. For example, if a cell is filled with CbCR data in the profit matrix (e.g. profit of US MNEs in France), the aim has been to use CbCR data to fill the corresponding cell in the other matrices (e.g. turnover of US MNEs in France).

462. In matrix cells where no source of 'hard' data is available, estimates are based on extrapolations relying on macroeconomic data (e.g. FDI data, GDP, GDP per capita). The extrapolation methodology builds on the information contained in the matrix cells filled with hard data, which aims to ensure consistency *within* each matrix. The extrapolation methodology is also designed to make the data *across* the four matrices as comparable with each other as possible. For example, extrapolations in the tangible assets and payroll matrices are based on data from the turnover matrix.

463. Among the four variables considered in the matrices, profit is arguably the most difficult to extrapolate when it is not observed in hard data, because the profit of MNEs may not always be located in the same jurisdiction as their economic activity. To overcome this issue, a sophisticated extrapolation methodology based on foreign direct investment (FDI) data has been developed. This methodology, inspired by Damgaard and Elkjaer (2017[4]) and Casella (2019[5]), involves various steps to identify the ultimate foreign investor into a jurisdiction, based on successive iterations on the existing data on 'immediate' foreign investors, and to eliminate 'pass-through FDI' from the data. One of the intermediate outputs of this procedure is a full matrix of FDI by jurisdiction of ultimate investor, which is interesting in its own right.

464. Overall, the various extrapolations ensure that all of the cells in the matrices can be filled, which makes it much easier to use the matrices for economic analysis. Extrapolated data are more fragile than hard data, but extrapolations represent a moderate share of the total amounts in the matrices (on average 25% across the four matrices), meaning that the information in the matrices is based predominantly on hard data. There are important geographic differences in the share of extrapolated values. This share is relatively low in high-income jurisdictions, higher in middle-income jurisdictions, and very high in low-income jurisdictions.[2] In investment hubs, the share of extrapolated values, while substantial (e.g. close to 40% in the profit matrix), is much lower than it would have been in the absence of the CbCR data, highlighting the importance of CbCR as a key new source of data on the amount of profit in investment hubs.

465. The various data sources mobilised to build the matrices have limitations, as is the case for any source of economic data. More specifically, CbCR data on profit have issues related to ambiguities in the treatment of intra-company dividends as well as 'stateless' entities, these ambiguities being related to the fact that 2016 was the first year in which the data was collected (OECD, 2020[3]). This may give rise to cases of double-counting of profit and revenues.[3] Another limitation of the data sources is that ORBIS unconsolidated account data has uneven coverage across jurisdictions. Reflecting this, ORBIS is used to fill the matrices only in jurisdictions where coverage is deemed sufficiently good, but even in these jurisdictions, coverage is not always exhaustive. A limitation of the OECD Analytical AMNE database is that some values are based on imputations and alternative sources to fill coverage gaps in the underlying data (Cadestin et al., 2018[2]). Finally, a limitation of the OECD AMNE database is that it does not include the financial sector in its data on inward investment in European jurisdictions.

466. To assess the implications of these limitations, improve data quality and ensure consistency across the various data used in the matrices,[4] extensive benchmarking and quality checks have been undertaken in this chapter. The benchmarking primarily takes advantage of the fact that, in many matrix cells, several data sources are simultaneously available, making it possible to assess their consistency. Data in the matrices have also been cross-checked against other relevant sources, including tax or financial account data shared with the OECD Secretariat by jurisdiction representatives. Overall, the

consistency checks reveal some inconsistencies, but suggest good overall data comparability across sources. For example, the correlation between CbCR data and estimates based on ORBIS, computed across the matrix cells where both of these sources are available, exceeds 90% in the profit and turnover matrices, and the correlation of estimates based on extrapolations with those from hard data ranges between 64% and 96% across the four matrices and the various hard data sources considered.

467. The four matrices have been used extensively by the OECD Secretariat in its assessment of the estimated effect of Pillar One and Pillar Two on tax revenues (Chapters 2 and 3 of this report) and MNE investment behaviour (Chapter 4). In the case of Pillar One, the profit and turnover matrices were used primarily to assess the location of the residual profit of MNE groups (in the form of a 'residual profit matrix'), so as to identify jurisdictions that would provide double tax relief, i.e. from which residual profit would be reallocated under Pillar One (see Chapter 2).

468. In the case of Pillar Two, the profit matrix was used, in combination with data on effective tax rates, to assess the amount and the location of the 'low-taxed' profit of MNEs (i.e. profit that is currently taxed at an effective rate below the potential minimum tax rate). The profit and turnover matrices have also been used to assess the extent of MNE profit shifting and how profit shifting could be reduced by the introduction of Pillar Two (the tangible assets and payroll matrices have also been used instead of the turnover matrix for the purpose of robustness checks), and, in turn, how this could affect tax revenues across jurisdictions (see Chapter 3). In addition, the turnover matrix has been used to proxy where some of the revenues from the minimum tax would accrue (here as well, the tangible assets and payroll matrices have been used for the purpose of robustness checks).[5] Finally, the tangible assets and payroll matrices were used to model the implications of potential 'carve-outs' to the minimum tax based on economic substance.[6]

469. In the investment impact analysis, the matrices were used to calibrate the framework used to assess the impact of Pillar One and Pillar Two on forward-looking effective tax rates (see Chapter 4 and Hanappi and González Cabral (2020[6])).

470. This chapter contains a preliminary version of the four matrices, presented at a certain level of aggregation (i.e. by income groups and broad geographic regions). After extensive consultation with members of the Inclusive Framework, there was no consensus over whether or not jurisdiction-specific data in the four matrices should be publicly released as part of the economic impact assessment. In view of this lack of consensus, no jurisdiction-specific data are included in this chapter.

471. Looking ahead, the matrices presented in this chapter could be used in the future for a range of other purposes, in the area of tax policy analysis and beyond, as discussed in the conclusion of this chapter.

5.2. Main existing data sources on MNE profit and activities

472. A number of data sources are available to assess the location of the profit and economic activity of MNEs, with different strengths, limitations and coverage, as discussed for example in OECD (2015[7]) and OECD (2018[8]). This section gives a brief overview of the main existing data sources, with the aim to provide useful background to the methodology underlying the construction of the matrices. Indeed, all of the sources presented in this section are mobilised – to different degrees – to build and benchmark the matrices presented in this chapter. More precisely, CbCR data, ORBIS data and data from the OECD AMNE and Analytical AMNE databases are used directly in the matrices, and used to benchmark each other. FDI data are used for the purpose of the extrapolations in the profit matrix. Finally, data from the US Bureau of Economic Analysis (BEA) are used only for the purpose of benchmarking. This section gives general information on these data sources, while the precise way in which they are used in the matrices (e.g. preference order, extrapolation methodology) is presented in the following sections.

5.2.1. Anonymised and aggregated Country-by-Country Report (CbCR) data

473. The obligation for MNE groups with global revenues above EUR 750 million to report their profit and economic activities on a country-by-country basis was introduced in 2016 as part of the implementation of the OECD/G20 BEPS Project, in order to support jurisdictions in combating BEPS. While the main purpose of CbCRs is to support tax administrations in the high-level detection and assessment of transfer pricing and other BEPS-related risks, data collected from CbCRs also offer great potential for the economic analysis of BEPS and MNEs in general.

474. MNE groups file their CbCRs with tax administrations, typically in the jurisdiction of their ultimate parent entity. While the individual CbCRs of MNE groups are generally not public, it was decided as part of the work on Action 11 of the OECD/G20 BEPS Project that jurisdictions would compile and provide aggregated and anonymised CbCR statistics to the OECD for publication (OECD, 2015[7]).

475. The first vintage of aggregated and anonymised CbCR statistics was published in July 2020 as part of the 2020 edition of the OECD's Corporate Tax Statistics (OECD, 2020[3]). The dataset focuses on year 2016 and includes nearly 4 000 MNE groups from 26 ultimate parent jurisdictions (see list in Annex 5.A). This dataset contains a vast array of information on the global financial and economic activities of these MNE groups, including information on the number of employees, related and unrelated party revenues, profits and taxes paid (generally based on financial accounting data), as well as the main business functions across jurisdictions or jurisdiction groups.

476. The way the data is collected ensures that all activities and profits of the covered MNE groups are included, even in jurisdictions that are often subject to coverage issues (e.g. zero-tax jurisdictions and investment hubs). This makes CbCR data a key new source of information compared to existing sources, especially for the analysis of BEPS. As any data source, the CbCR data are subject to a number of limitations. One limitation of this first vintage is that several countries, including large ones, have not submitted aggregated CbCR statistics to the OECD for publication. Another issue is that due to lack of clarity in the expected treatment of intra-company dividends and 'stateless entities', the profit and to a lesser extent revenue variables may be subject to some double counting.[7] A full description of the CbCR dataset, including a presentation of the collection and aggregation methodology, a discussion of the main data limitations and high-level summary statistics based on the data can be found in OECD (2020[3]).

5.2.2. The ORBIS database

477. The ORBIS database, provided by Bureau van Dijk (BvD), is the largest cross-country database of ownership information and financial accounts of firms worldwide. It relies on information from various underlying sources, including credit rating agencies (e.g. Cerved in Italy) and national banks (e.g. National Bank of Belgium). ORBIS contains firm-level data for both publicly listed and privately owned companies. The available variables typically include balance sheet information (e.g. assets, liabilities), information from the profit and loss statement (e.g. turnover, cost of employees, earnings before interest and taxes (EBIT), profit before tax), data on the number of employees, and ownership information (e.g. direct and ultimate owners of an entity, ownership shares).

478. ORBIS contains financial account data both at the consolidated (i.e. MNE group) and unconsolidated (i.e. entity) level. The coverage of consolidated account data is good in most jurisdictions of ultimate parent, while the coverage of unconsolidated account data is very uneven across jurisdictions. For example, unconsolidated account data coverage is good in many European jurisdictions, while it is poor in the United States and most developing economies, zero-tax jurisdictions and investment hubs. In this chapter, ORBIS unconsolidated level data have only been used in jurisdictions with good coverage (see list in Annex 5.A). Despite the uneven coverage of unconsolidated account data, ownership information is comprehensive in ORBIS, as the global ultimate owner (GUO) of each entity is generally identified, even if it is located in a jurisdiction with poor coverage of unconsolidated financial accounts.[8]

This enables a comprehensive identification of the jurisdiction of ultimate parent of the MNE entities in ORBIS.

479. Given that ORBIS data are not primarily collected for statistical analysis, important processing and cleaning work is required to enhance data reliability (e.g. eliminating duplicates and reporting errors). The ORBIS dataset used to build the matrices in this chapter has benefitted from extensive cleaning of both ownership data and financial data, building on longstanding OECD expertise with ORBIS (see Annex 5.B for details).

5.2.3. The OECD AMNE database

480. The OECD AMNE database contains data on the activities of foreign-owned affiliates in OECD countries ('inward' data), and also on the foreign activities of affiliates of MNEs headquartered in OECD countries ('outward' data). The OECD AMNE database is based on data reported to the OECD and other institutions, including Eurostat (where it is included in the Eurostat FATS database) and the US BEA (where is it included in the US AMNE database), in the framework of annual surveys on the activities of foreign-owned enterprises and foreign affiliates abroad controlled by residents of the compiling country.

481. The AMNE database contains 17 variables broken down by country of origin (inward investment) or destination (outward investment) and by industrial sector (50 industries) for 31 OECD countries. The available variables include, among others, production, value added, employment, labour compensation, research and development expenditures, exports, gross investment in tangible goods and gross operating surplus. Gross operating surplus is the closest measure of profit among these variables, but it nevertheless has important conceptual differences with profit before tax as reported in firms' financial accounts. In particular, gross operating surplus is based on a national account methodology to account for depreciation, and interest paid is not subtracted from profit.

482. The main limitations of the AMNE database for the purpose of the analysis are (i) that it does not contain data on domestic-owned MNE entities, and (ii) that financial sectors are excluded from the scope of the data in certain jurisdictions (e.g. EU countries). In addition, there are a number of data gaps in the bilateral AMNE data, reflecting notably confidentiality issues as certain values at the country-pair level contain information related to a small number of MNEs. Inward and outward data also do not provide the same set of variables, as for instance outward data does not include gross operating surplus.

5.2.4. The OECD Analytical AMNE database

483. The OECD Analytical AMNE database contains a full bilateral matrix of the output of foreign affiliates in 59 countries plus a 'rest of the world' aggregate. Data is broken down by host and parent country and by industry (across 34 industrial sectors of the NACE Rev. 2 classification (Eurostat, 2008[9])). Analytical AMNE also contains data on value-added, exports and imports of intermediate inputs at the host country and industry level, which provides information on the contribution of foreign MNEs to those variables but without a breakdown by country of ultimate parent. In addition, the data contains a second set of tables providing information on output, value-added, exports and imports of intermediate inputs of domestic MNEs and non-MNE domestic firms.

484. The OECD Analytical AMNE database was constructed using the OECD AMNE database (see previous section) as a starting point. In order to estimate the missing information in the OECD AMNE database across countries and industries, additional national sources have been used and various statistical methodologies applied (see Cadestin et al. (2018[2]) for more details).

485. The main limitation of Analytical AMNE data for the purpose of this chapter is that it focuses on a limited number of variables, and does not contain an indicator of profit (unlike the OECD AMNE database, which contains data on gross operating surplus), tangible assets or payroll. Another limitation relates to the fact that the underlying AMNE data focuses on foreign-owned MNEs, which implies that the information

on domestic-owned MNEs in Analytical AMNE tends to be more fragile than the information on foreign-owned MNEs, as it tends to rely more on other (potentially less harmonised) sources as well as imputations.

5.2.5. Foreign Direct Investment (FDI) data

486. Several international organisations (e.g. OECD, IMF, UNCTAD) publish bilateral FDI data across a wide range of jurisdictions. These data are generally collected as part of the balance of payments statistics. FDI data contain information about investment positions (i.e. stocks) and investment flows across borders, focusing on investments involving a long-term relationship and reflecting a lasting interest and a degree of control (based on a 10% ownership threshold). Financial flows consist of equity transactions, reinvestment of earnings, and intercompany debt transactions. FDI data also contain information on investment income (dividends, interest) as well as royalty flows. Bilateral FDI data are typically reported by both the investor and the recipient jurisdiction, but different values may be reported for the same data point due to methodological differences between reporting jurisdictions.[9]

487. For the purpose of this chapter, one advantage of FDI data is their wide geographic coverage, as most pairs of jurisdictions with significant cross-border investment are covered. Data on FDI position and FDI income can be used to build proxy measures of MNE profit in foreign jurisdictions. In contrast, FDI data lack direct information on turnover, tangible assets or payroll. The well-known fact that BEPS behaviour can distort FDI data is an issue for the analysis of 'real' investment activity based on FDI (Damgaard, Elkjaer and Johannesen, 2019[10]), but it is not necessarily an issue to assess profit location across jurisdictions, which is the goal of the profit matrix. Indeed, FDI data are known to provide some information on the location of shifted profits of MNEs (Bolwijn, Casella and Rigo, 2018[11]).

488. One important limitation of FDI data is that they traditionally focus on direct investors into a jurisdiction, as opposed to ultimate investors. This can be an issue especially since certain investments can go through several jurisdictions before they reach their final destination (Borga and Caliandro, 2018[12]), especially in the context of profit shifting schemes. In recent years, the OECD started publishing inward FDI position statistics by ultimate (as opposed to immediate) investor for a subset of 15 recipient jurisdictions (OECD, 2015[13]). When using FDI data, this chapter makes use of these data by ultimate investor where available. In other recipient jurisdictions, it relies on a sophisticated methodology, inspired by Damgaard and Elkjaer (2017[4]) and Casella (2019[5]), to iterate on direct FDI data and eliminate 'pass-through' (or 'conduit') FDI to measure FDI positions by ultimate investor (see Annex 5.C).

5.2.6. Data from the US Bureau of Economic Analysis

489. National sources can contain additional data on the activity and profit of MNEs, even though they lack the cross-country perspective of the sources listed above. The most detailed national source is the US Bureau of Economic Analysis (BEA). As part of the Activities of US Multinational Enterprises database, the BEA provides statistics on the worldwide activities of US MNEs, including balance sheet data, income statement data, data on employment and compensation of employees, data on trade in goods and services, and expenditures for research and development. Data are disaggregated by both affiliate jurisdiction and economic sector, with different levels of disaggregation (geographic and sectoral) depending on the table considered. Part of the BEA data on US MNEs is the basis for the OECD AMNE data for the United States, but the BEA data is in many respects more detailed than the OECD AMNE data relative to US MNEs.

490. Two indicators of profit are included in the BEA data: (i) net income, from the income statement (Table II.D 1 of the BEA database), which has been identified as involving some double counting of equity income (Blouin and Robinson, 2019[14]) and (ii) 'profit-type' return, from the decomposition of value-added (Table II.F 1 of the BEA database), which is the closest indicator to financial account profit and is not

subject to the double counting issue (see also Clausing (2020[15]) for a discussion). The BEA data also includes indicators of turnover ("Sales" in the income statement), tangible assets ("Property, plant and equipment" in the Balance sheet of affiliates in Table II.B 1-2 of the BEA database) and payroll ("Compensation of employees" in the decomposition of value-added).

5.3. Overview of the methodology underlying the matrices

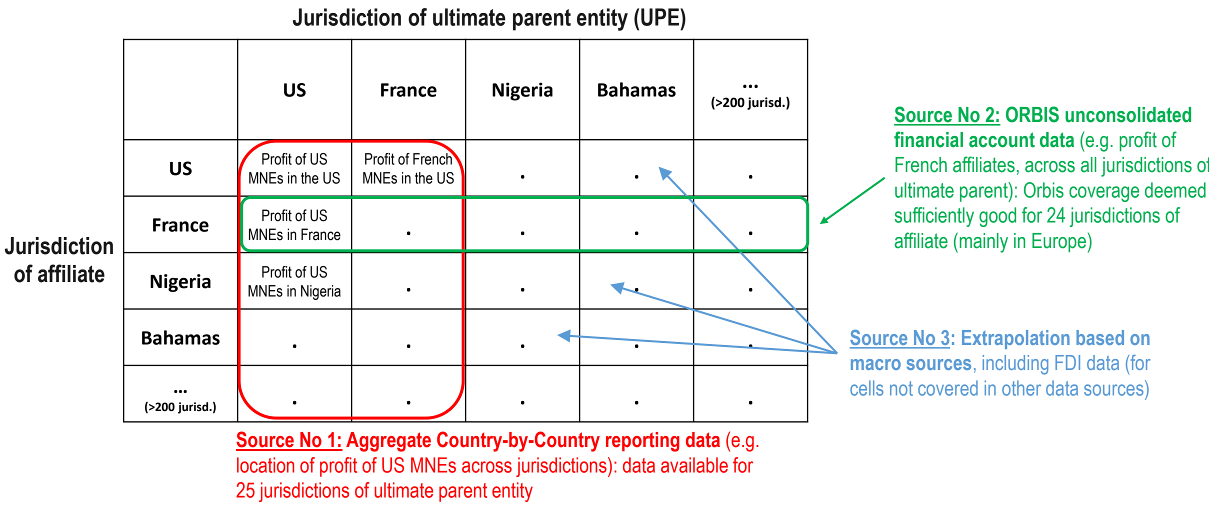
491. To overcome the coverage limitations of the existing data sources, the approach is to combine data from different sources in a consistent framework (i.e. a set of matrices). The aim is to obtain a global geographic coverage of the profit and economic activity of MNEs, while using for each data point the most reliable data source available. Another key feature of the approach is to take advantage of cases where several data sources are available for the same data point, to benchmark sources against each other in order to enhance data quality and consistency.

5.3.1. Stylised example illustrating the methodology

492. In practice, each matrix contains data across more than 200 jurisdictions of affiliate (matrix rows) and broken down across more than 200 jurisdictions of MNE ultimate parent (matrix columns). Each matrix therefore takes the form of a square table with more than 200 rows and more than 200 columns (the jurisdictions in rows and columns are the same). For example, the France-United States cell in the profit matrix contains the profit of US MNEs (i.e. MNEs with an ultimate parent in the United States) in France. Each matrix cell is filled with a certain data source, with a pre-defined order of preference when several sources are available, as discussed below.

493. As can be seen in Figure 5.1, which presents in a stylised way the data sources underlying the profit matrix, different data sources provide coverage of a different nature. In particular, CbCR data can be used to fill *columns* of the matrices, while ORBIS data can be used to fill *rows*. This is because CbCR data typically contain detailed information across affiliate jurisdictions, for a given jurisdiction of ultimate parent. In contrast, ORBIS unconsolidated account data contain, in the jurisdictions of affiliate with good ORBIS coverage, detailed information on MNE entities from any jurisdiction of ultimate parent.

Figure 5.1. Profit matrix: Stylised overview and underlying data sources



Note: CbCR data are used to fill *columns* of the profit matrix (e.g. profit of French MNEs across jurisdictions). ORBIS unconsolidated account data are used to fill *rows* of the profit matrix (i.e. MNE profit in France, split across ultimate parent jurisdictions). These two sources are used only where available, and in the case of ORBIS, where data coverage is sufficiently good. Other cells in the profit matrix are filled with extrapolations based on macroeconomic data, including FDI data. See Annex 5.A for the lists of jurisdictions where CbCR and ORBIS data are used.
Source: OECD Secretariat.

5.3.2. Definition of the variables included in the matrices

494. The four variables considered in the matrices (profit, turnover, tangible assets and payroll) are defined in Table 5.1. These definitions were chosen based on the needs of the impact assessment of Pillar One and Pillar Two, also taking into account the constraints around existing data sources. In particular, all four variables are based on concepts from financial accounting data (as opposed to tax or national accounting data). The focus is on year 2016, which is the latest available year across the various data sources used.

495. The definitions presented in Table 5.1 are 'targets', in the sense that the four matrices are filled with variables that are as close to these definitions as possible. However, the exact values in matrix cells can deviate from these targets due to limitations in the available data used to fill the matrices (e.g. some intra-group dividends are included in CbCR data). The exact variables considered in each of the underlying data source are presented in Section 5.4.

Table 5.1. Definition of the variables considered in the four matrices

Variable	Definition
Profit	Profit before tax, excluding dividends received from affiliates
Turnover	Revenues from sales to third-party and intra-group entities
Tangible assets	Property, plant and equipment, net of depreciation
Payroll	Expenditures for salaries and wages, including bonuses, social contributions and other employee benefits

Note: The exact definitions of variables across matrix cells can deviate from the 'targets' presented in this Table due to limitations in the available underlying data (see Section 5.4). All four variables are based on financial accounting data. Data related to entities from the same MNE group are consolidated at the jurisdiction level. Only data for MNE groups with a positive profit in the jurisdiction considered are included, as discussed in Section 5.4.
Source: OECD Secretariat

5.3.3. Overview of data sources and their preference order

496. All four matrices are filled using the same overall methodology. However, given that certain variables are available in some sources and not in other (e.g. CbCR data contain information on profit, turnover and tangible assets, but not on payroll), the combination of sources is not exactly the same across the four matrices (Table 5.2). In all four matrices, a 'last resort' method based on extrapolations has been employed when no hard data is available. This ensures that all matrix cells can ultimately be filled, although the extrapolated values obviously come with a greater degree of uncertainty than values based on hard data.

Table 5.2. Overview and preference order of data sources underlying the set of matrices

Data source (order of preference)	Profit matrix	Turnover matrix	Tangible assets matrix	Payroll matrix
1	CbCR data	CbCR data	CbCR data	ORBIS data
2	ORBIS data	ORBIS data	ORBIS data	OECD AMNE data
3	Extrapolations based on macro data (e.g. FDI)	OECD Analytical AMNE data	Extrapolations based on turnover matrix	Extrapolations based on turnover matrix
4		OECD AMNE data		
5		Extrapolations based on macro data (e.g. gravity model)		

Note: The combination of sources differs across matrices, reflecting differences in the available variables across sources. The order of preference is necessary to decide on which source to use to fill the matrix cells for which several sources are available. The ordering is as consistent as possible across the matrices to ensure greater consistency across the matrices. Finally, there is for each matrix an extrapolation method (involving more uncertainty than hard data) that ensures that all matrix cells can ultimately be filled.
Source: OECD Secretariat

497. As several data sources can be available for the same data point, it is necessary to decide on an order of preference between data sources. To ensure consistency between the four matrices, the approach has been to order data sources in as consistent a way as possible across the four matrices. Another general principle is that hard data has been given priority over extrapolations. With these principles, several ordering choices remained possible, with no undisputable way to decide on the best ordering between sources that have different strengths and limitations (e.g. CbCR vs. ORBIS data). Ultimately, the choice has been made to make CbCR data the preferred source, reflecting that it is the available source with the widest geographic coverage across several variables considered. The second preferred source is ORBIS, which offers the benefit of covering the four variables considered. After these two sources come the OECD Analytical AMNE database (in the turnover matrix) and the OECD AMNE database (in the turnover and payroll matrices), and ultimately extrapolations specific to each matrix. Robustness checks are presented in Section 5.8.5 to illustrate the potential implications of different source ordering choices.

498. The fact that several data sources cover the same data points is very useful for the purpose of benchmarking sources against each other and more broadly assessing the quality of the data. The extensive benchmarking that was undertaken as part of the preparation of the matrices is presented in Section 5.8.

5.4. Methodology underlying the use of hard data

499. Four sources of hard data are used to fill the matrices: (i) CbCR data (in all matrices except the payroll matrix), (ii) ORBIS data (in all four matrices), (iii) OECD Analytical AMNE data (in the turnover matrix) and (iv) OECD AMNE data (in the turnover and payroll matrices). This section describes more precisely how each of these sources are used. The extrapolation methods to fill cells for which no hard data source is available are described in the following section.

5.4.1. Aggregated and anonymised CbCR data

500. Aggregated and anonymised CbCR data have been used directly where available to fill cells in the profit, turnover and tangible assets matrices. Data are taken from Table I, 'Sub-groups with positive profits' panel. The variables used are respectively 'Profit (Loss) before Income Tax', 'Total Revenues' and 'Tangible Assets other than Cash and Cash Equivalents' (OECD, 2020[3]).

501. The data on 'sub-groups with positive profits' focus only on entities belonging to MNE groups that have positive profits in the jurisdiction considered, and therefore exclude entities from MNE groups that are in a loss position in the jurisdiction considered. An alternative option would have been to use the data for all MNE sub-groups regardless of their profit position. This would have led to lower profit amounts, as the profits of sub-groups with positive profits would have been netted of the losses of loss-making sub-groups. In contrast, it would have led to higher turnover and tangible assets, as the turnover and tangible assets of more MNE groups would have been included.[10]

502. The choice to focus on sub-groups with positive profits was driven by the aim of the matrices, which is to inform the impact assessment of Pillar One and Pillar Two. Given that both pillars would primarily affect the taxation of MNE groups in jurisdictions where they are in a profit position, the focus on sub-groups with positive profits makes the matrices more relevant tools for the impact assessment (see Chapters 2 and 3). This choice has been made consistently across the profit, turnover and tangible assets matrices to ensure consistency when the matrices are used in combination. To the extent possible, consistent assumptions to focus on sub-groups with positive profits have been made with other data sources (i.e. ORBIS) and implicitly in the extrapolations, as further discussed below.

503. The CbCR data have been used for 25 jurisdictions of ultimate parent, i.e. to fill *columns* in the matrices (see list of jurisdictions in Annex 5.A).[11] Certain jurisdictions of ultimate parent have reported the CbCR data at the level of each jurisdiction of affiliate, while others have reported the data for groups of jurisdictions (e.g. with groups by continent, or, in the most extreme cases of aggregation, only a split between their jurisdiction and foreign ones), or for a combination of individual jurisdictions and groups (e.g. with groups such as 'other European jurisdictions', 'other African jurisdictions'). The typical reason for this grouping is to avoid the potential breach of taxpayer confidentiality. For the purpose of filling the matrices, data for jurisdiction groups have not been used, since they cannot directly be attributed to individual jurisdictions. As a result, CbCR data have been considered missing for these cells, which have been filled based on the other available data sources (or extrapolations) following the order of preference presented in Table 5.2.

504. The CbCR data focus on year 2016, which is (for the moment) the only year available across a range of jurisdictions of ultimate parent, with the exception of the United States, which has already published 2017 CbCR data. The 2016 CbCR data in the United States was based on voluntary filing, while it was compulsory in 2017, leading to an increase by more than 40% in the number of reporting MNE groups.[12] Given this, the choice was made to use 2017 data for the United States instead of 2016 data. While this creates a small time inconsistency with the other sources of data in the matrices, it has been judged, on balance, to be an inconsistency worth accepting as it enables a better coverage of MNE groups.

505. CbCR data focus only on MNE groups with global revenues above EUR 750 million, since reporting is not compulsory for MNE groups below this threshold. In the context of the impact assessment of Pillar One and Pillar Two, the exclusion of smaller MNE groups from the data may be welcome, as it would provide more accurate estimates under the illustrative assumption that Pillar One and Pillar Two would include global revenue thresholds of the same order of magnitude as the CbCR threshold. However, the revenue threshold implies some inconsistency with some other data sources, where such a threshold cannot directly be applied. A detailed analysis based on consolidated account data from the ORBIS database (as described in Chapter 2) suggests that more than 90% of global MNE profit comes from MNE groups that have a turnover above the CbCR reporting threshold. The same holds for the other variables considered (turnover, tangible assets and payroll). Reflecting this, the overall impact on the matrices of the inconsistency created by the global revenue threshold in CbCR data is likely to be limited.

506. As discussed in section 5.2.1 above, CbCR data come with a number of limitations.[13] In particular, dividends from affiliates are potentially reported in profits in an inconsistent way (sometimes included and sometimes not). The extensive benchmarking of CbCR data against other sources presented below suggests that this does not seem to affect disproportionally the overall quality of CbCR data, although the issue may be more consequential in certain jurisdictions depending on the reporting guidance that was given and the importance of intra-company dividends across jurisdictions. The other main issue is that the profit and activity classified as 'stateless' in CbCR data may correspond to different situations, including 'pass-through' entities whose activities and profits are already included elsewhere in the data. Reflecting this, the profit and activity of 'stateless' entities have not been included in the matrices to avoid the risk of double counting.

507. The tangible assets variable in CbCR data includes property, plant and equipment, but it can also include inventories. As a result, it is not fully consistent with the definition of tangible assets in Table 5.1, nor with data from other sources, which exclude inventories. To address this issue, tangible assets values taken from CbCR data have been scaled down to exclude inventories, based on the share of inventories in the tangible assets of US MNEs (which are assumed to be representative of MNEs from other jurisdictions) computed using data from the US BEA.[14]

5.4.2. ORBIS data

508. ORBIS data have been used to fill cells in the profit, turnover, tangible assets and payroll matrices, using unconsolidated financial accounts of firms belonging to MNE groups (identified thanks to ORBIS ownership data, as detailed in Annex 5.B). The quality of coverage of ORBIS unconsolidated financial account data has been judged sufficient to use ORBIS as a data source to fill the matrices in 24 jurisdictions of affiliate (see list in Annex 5.A).[15] For each jurisdiction of affiliate, ORBIS contains financial information on MNE entities from any jurisdiction of ultimate parent (even when the ultimate parent is in a jurisdiction with poor ORBIS coverage of unconsolidated account data). As a result, ORBIS data can be used to fill the *rows* corresponding to these jurisdictions in the matrices. The variables used for the four matrices are respectively 'Profit and Loss before tax', 'Operating revenue turnover', 'Tangible fixed assets' and 'Cost of employees'.

509. ORBIS data have been cleaned and checked extensively before usage, building on OECD expertise from a range of past projects. Such cleaning is required to enhance data reliability because ORBIS data is not primarily collected for statistical analysis. A detailed cleaning of the ORBIS ownership data used in this project, including the identification of missing ownership links in the original ORBIS database, has been implemented by the OECD Directorate for Science, Technology and Innovation, following the methodology of Bajgar et al. (2019[16]). Regarding financial account data, the cleaning procedure is inspired by Gal (2013[17]), Johansson et al. (2017[18]) and Bailin et al. (2019[19]). The detailed cleaning procedure of ORBIS ownership and financial account data is presented in Annex 5.B.

510. ORBIS data contains two measures of profit: (i) profit before tax (PBT) and (ii) earnings before interest and tax (EBIT). The difference between the two measures is that PBT includes dividends received and is net of interest paid, while EBIT does not include dividends and is not net of interest. None of the two measures is exactly consistent with the concept of profit reported in CbCR data (which is net of interest and may or may not include dividends). While both measures are relevant indicators of profit, the choice was made to use PBT in the profit matrix as it is a more relevant indicator for the impact assessment of Pillar One and Pillar Two. For example, PBT is the measure of profit considered as the potential basis to define residual profit under Pillar One. Reflecting that EBIT data is also informative, consistency checks have been carried out with ORBIS data focusing on EBIT instead of PBT (see Section 5.8.2).

511. To enhance consistency with CbCR data, in which, as discussed above, the choice has been made to consider only sub-groups with positive profits (i.e. entities from MNE groups with a positive profit in the jurisdiction considered), the same assumption has been made in ORBIS data.[16] Also consistent with CbCR data, the identification of whether an entity belongs to a corporate group was based on a 50% ownership threshold. In contrast, the EUR 750 million global revenue threshold was not applied, reflecting that the consolidated financial information of the MNE group to which each entity belongs is not always available in the data.[17]

512. The coverage of tangible assets and payroll in ORBIS unconsolidated account data is less extensive than the coverage of turnover. Across the 24 jurisdictions where ORBIS data are used to fill the matrices, the turnover of MNE entities with missing information on tangible assets represents on average about 16% of the total turnover of MNE entities, with exact numbers depending on the jurisdiction considered. Coverage of payroll in ORBIS is generally weaker than coverage of the other variables considered in this chapter. As a result, ORBIS data are only used to fill the payroll matrix in 18 jurisdictions, against 24 jurisdictions for the other variables (see list in Annex 5.A). Across these 18 jurisdictions, the average share of missing information on payroll is 26%. To avoid that these gaps in coverage induce a negative bias in the data from ORBIS, tangible assets and payroll values taken from ORBIS have been scaled up in proportion to the estimated under-coverage rate in the jurisdiction and for the variable considered.[18]

5.4.3. OECD AMNE data

513. Data from the OECD AMNE database were used to fill matrix cells in the turnover and payroll matrices, where they are used respectively as the fourth and second possible source (see Table 5.2). AMNE contains data on turnover ('turnover') and payroll ('personnel costs') for 'ultimate parent'-'affiliate' pairs of jurisdictions, mainly across OECD economies. The OECD AMNE database encompasses data from Eurostat FATS and US BEA AMNE databases. Both inward and outward AMNE data are used, and a preference was given to inward data in cases where both are available for the same matrix cell.

514. One issue is that OECD AMNE data do not always cover the full economy, as for example financial sectors are excluded in the inward statistics of European countries. In the case of turnover, AMNE data have been used to fill the turnover matrix only when they cover the full economy. In the case of the payroll matrix, where fewer alternative data sources are available, AMNE data have been used even when they do not cover the full economy, but they have been rescaled to make them as consistent as possible with full economy data. More precisely, the values from AMNE data have been multiplied by the ratio of turnover in the whole economy (taken from the relevant cell in the turnover matrix) to turnover in the sectors covered by AMNE data. For example, the payroll of US MNEs in France sourced from AMNE data has been multiplied by the ratio of turnover of US MNEs in France, sourced from the turnover matrix, to turnover of US MNEs in France, sourced from AMNE data. When AMNE and the data underlying the turnover matrix have the same sectoral coverage, the ratio is close to one and the adjustment is not consequential. When sectoral coverage is narrower in AMNE, the ratio is above one, and the rescaling ensures that the payroll

matrix is filled with data that focus – at least in an approximated way – on the whole economy and that is consistent with the turnover matrix.[19]

5.4.4. OECD Analytical AMNE data

515. OECD Analytical AMNE data were used only in the turnover matrix, where they are the third source in the preference order, after CbCR data and ORBIS data. At the level of disaggregation required for the matrices (i.e. jurisdiction-pair level), Analytical AMNE contains data on the gross output of MNEs, but not directly on turnover. The data available at a more aggregated level (i.e. jurisdiction level), which contain both gross output and turnover, suggest that these two variables are generally closely related to each other, except in certain specific sectors (in particular wholesale and retail trade, see Figure 2 in Cadestin et al. (2018[2])).

516. To 'convert' the data on gross output into a measure of turnover at the jurisdiction-pair level, gross output in a given industry and jurisdiction-pair was multiplied by the ratio of gross output to turnover for this industry and jurisdiction-pair, based on (non-published) data underlying the OECD Analytical AMNE database. For jurisdiction pairs where this ratio was not available, the average ratio of gross output to turnover at the 'market jurisdiction'-'industry' level, or at the industry level depending on data availability, was applied.

5.5. Methodology underlying the data extrapolations

517. The extrapolation methodology depends on the matrix considered, as the nature of the gaps in hard data and the available proxies that can be used for the extrapolation depend on the variable considered. For example, FDI data can be used to extrapolate profit location, but it is less well suited to extrapolating the other variables of interest.

518. Still, the general approach has important similarities across the four matrices. One common element is that the extrapolation methodology covers all cells in the matrix considered, to ensure that there is no data gap in the final matrices. In addition, the methodology often builds on the information contained in the matrix cells filled with hard data, which aims to ensure that extrapolated values are as consistent with these other matrix cells as possible. Finally, the extrapolations aim at making the data in the four matrices as comparable with each other as possible, reflecting that the matrices may need to be used in combination with each other. For example, the extrapolations in the tangible assets and payroll matrices are building on the data in the turnover matrix.

5.5.1. Extrapolations in the profit matrix

519. The extrapolation strategy in the profit matrix differs between 'diagonal' cells (i.e. the profit of MNEs in their jurisdiction of ultimate parent) and 'non-diagonal' cells (i.e. the profit of MNEs in foreign jurisdictions). This reflects differences in the available data. For non-diagonal cells, the extrapolation is primarily based on bilateral FDI data, which are not available in diagonal cells as FDI is, by definition, focusing on foreign investment. Reflecting this, the extrapolation methodology for diagonal cells is based on other macroeconomic aggregates, as discussed below.

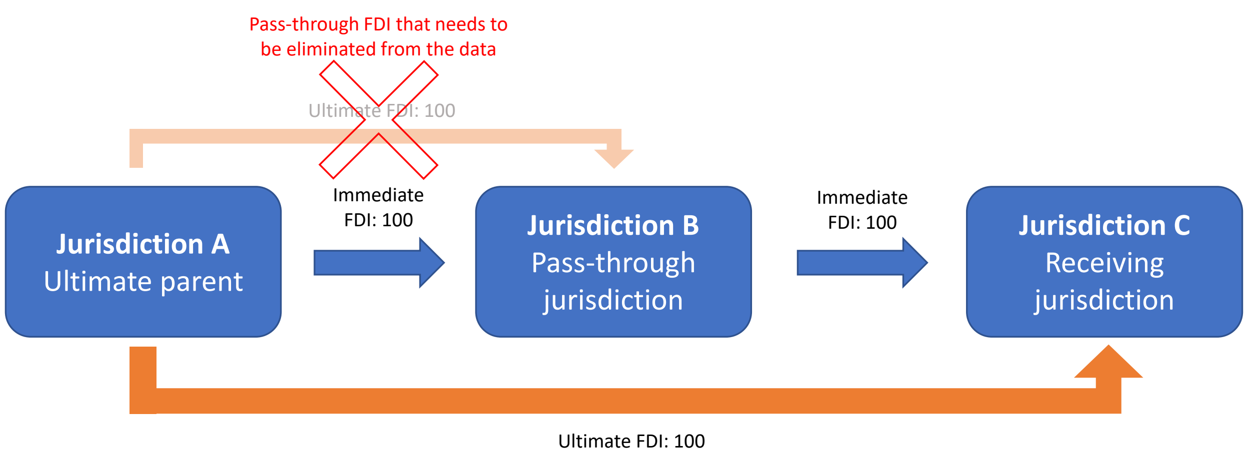
Extrapolations in 'non-diagonal' cells in the profit matrix

520. Profit in non-diagonal cells is extrapolated using bilateral FDI positions and an assumption regarding the rate of return on FDI, which takes into account heterogeneities in the rate of return across both investing and recipient jurisdictions. An overview of the methodology is presented in this section, and a fully detailed description of all its components is included in Annex 5.C.

521. The FDI positions considered are based on the location of the *ultimate* (as opposed to *immediate*) investor into a jurisdiction, for consistency with the other data sources in the profit matrix, which are based on the concept of ultimate investor (i.e. the location of the ultimate parent of the investing MNE group). The ultimate investor is reported in the available OECD data for 15 jurisdictions. In other jurisdictions, the location of the ultimate investor is identified via repeated iterations on immediate FDI data, following and refining the methodology of Casella (2019[5]).

522. This methodology is refined in two ways: (i) The probability that FDI from a certain jurisdiction is 'pass-through' (i.e. that the ultimate investor is not located in this jurisdiction), which is a central input to the methodology, is estimated based on the available OECD data on FDI by ultimate versus immediate investor, as explained in Annex 5.C. This is a more direct and geographically more widely available measure than the measure used by earlier studies (Damgaard and Elkjaer, 2017[4]; Casella, 2019[5]; Damgaard, Elkjaer and Johannesen, 2019[10]), which is based on the share of Special Purpose Entities (SPEs) in a jurisdiction's outward FDI investment. (ii) After identifying the ultimate investor corresponding to a given FDI position, the methodology eliminates from the data the intermediate positions corresponding to 'pass-through' (or 'conduit') FDI, as illustrated in Figure 5.2, which is an additional step compared to the methodology of Casella (2019[5]).

Figure 5.2. Stylised example on 'immediate', 'ultimate' and 'pass-through' FDI



Note: In this stylised example, an MNE group with an ultimate parent in jurisdiction A has invested 100 into jurisdiction C, passing through jurisdiction B. FDI statistics by immediate investor report an FDI position from jurisdiction A into jurisdiction B, and a similar position from jurisdiction B into jurisdiction C. FDI statistics by ultimate investor in jurisdiction C report that the ultimate investor is jurisdiction A. However, when considering the ultimate investors into jurisdiction B, the FDI position of jurisdiction A into jurisdiction B is also considered as an ultimate investment (A being the ultimate parent), while it is only a 'pass-through FDI' that should be eliminated from the data.
Source: OECD Secretariat.

523. FDI data by ultimate investor is sourced from the OECD FDI statistics, and FDI data by immediate investor from the OECD FDI statistics and the IMF Coordinated Direct Investment Survey (CDIS). For the pairs of jurisdictions where no immediate FDI data are available, FDI positions are imputed using a standard gravity model relying on variables such as distance, GDP and GDP per capita of the investor and recipient jurisdictions. These imputations based on a gravity model ensure that the matrix of immediate FDI data is filled completely and, in turn, that the matrix of ultimate FDI estimates can also be filled completely. While these imputations based on a gravity model come with more uncertainty than hard FDI data, their overall impact is limited as they represent only about 2% of FDI positions in the final matrix of FDI by immediate investor.

524. Based on this methodology, two full matrices of FDI positions have been built: (i) FDI positions by *immediate* investor, and (ii) FDI positions by *ultimate* investor. The ultimate investor matrix is presented, for aggregated groups of jurisdictions, in Table 5.3 (the immediate investor matrix is presented in Annex Table 5.C.3). While these two FDI matrices have only been used as intermediate inputs for the extrapolations in the profit matrix in this chapter, they are interesting in their own right and could be the subject of (or useful tools for) future studies. For example, it is interesting to note that eliminating 'pass-through FDI' mainly reduces the relative importance of investment hubs (both as recipient and investor jurisdictions) in the FDI matrix (see Annex 5.C).

Table 5.3. Matrix of FDI positions by ultimate investor, aggregated by broad jurisdiction groups

		Jurisdiction of ultimate investor				
	(USD billion of 2016)	High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of recipient	High income (64 jurisd.)	7787.9	591.0	1.4	2830.4	**11210.8**
	Middle income (105)	2559.7	367.9	2.2	2755.6	**5685.4**
	Low income (29)	11.1	7.2	0.1	2.3	**20.7**
	Investment Hubs (24)	4252.2	770.4	0.7	2228.8	**7252.1**
	Total	**14611.0**	**1736.4**	**4.4**	**7817.2**	**24169.0**

Note: Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP. The FDI positions considered in this matrix are based on the location of the *ultimate* (as opposed to *immediate*) investor into a jurisdiction. See methodology in Annex 5.C.
Source: OECD Secretariat.

525. The assumed rate of return on FDI is based on a global 'standard' return on FDI, which is computed as the median ratio of profit to FDI in the profit matrix cells filled with hard data (which yields a value of 7.9%). This approach offers the benefit of making the extrapolation methodology as consistent as possible with the hard data in the profit matrix. In addition to this 'standard' return, the rate of return takes into account how the average rate of return to FDI deviates (positively or negatively) from the global average across both investing and receiving jurisdictions. This deviation is computed based on the available data on FDI income[20] (see Annex 5.C for more details).

Extrapolations in 'diagonal' cells in the profit matrix

526. Less data are generally available on the activity and profit of domestic-owned MNEs than on foreign-owned ones. This might reflect the fact that foreign-owned MNEs tend to generate more policy interest than domestic-owned ones. In addition, it may be difficult from a statistical collection point of view to disentangle the activity of domestic-owned MNE entities from that of non-MNE firms. Indeed, identifying a domestic-owned MNE entity requires knowing the location of its ultimate owner (to make sure that it is not foreign-owned) but also having information about the activities of other subsidiaries of this ultimate owner (to make sure that at least one of them is located abroad). Reflecting this, extrapolations in diagonal matrix cells are overall less refined than in non-diagonal cells, and can be assumed to come with greater uncertainty.

527. The starting point for the extrapolation in diagonal cells of the profit matrix is the total profit of domestic-owned firms (MNEs and non-MNEs) as estimated by Tørsløv et al. (2018[1]). To obtain the diagonal cells in the profit matrix (i.e. profit of domestic-owned MNEs), the profit of domestic-owned firms needs to be split between MNE and non-MNE entities. This is done based on a regression of the share of profit of domestic-owned MNEs in the total profit of domestic-owned firms, which is estimated over the approximately 30 jurisdictions where hard data are available in the profit matrix diagonal (i.e. based on

data from CbCR or ORBIS) for the numerator (i.e. profit of domestic-owned MNEs) and in Tørsløv et al. (2018[1]) for the denominator (i.e. total profit of domestic-owned firms), and using GDP and GDP per capita as explanatory variables.

528. For jurisdictions not covered in Tørsløv et al. (2018[1]), where even less data are available, the diagonal profit is directly extrapolated based on a regression on GDP and GDP per capita. The regression results suggest that in general, profit of domestic-owned MNEs is positively correlated with GDP per capita, in line with the intuition that richer countries have more MNE headquarters than poorer ones. Overall, these extrapolations in the diagonal of the profit matrix are based on relatively limited information (e.g. compared to extrapolations in non-diagonal cells), and the extrapolation methodology aims primarily to ensure that cells are filled with plausible values that are consistent with the hard data in the rest of the profit matrix.

5.5.2. Extrapolations in the turnover matrix

529. With four available sources of hard data (CbCR, ORBIS, OECD Analytical AMNE and OECD AMNE), the turnover matrix is the one relying least on extrapolations across the four matrices. The extrapolation methodology follows a similar spirit as the extrapolations in the profit matrix, in the sense that the approach differs between diagonal and non-diagonal cells. In non-diagonal cells, turnover is extrapolated using a gravity model relating bilateral turnover to macroeconomic variables, the distance between countries and, where available, bilateral FDI data. In diagonal cells, turnover is extrapolated using macroeconomic variables in a way informed by the matrix diagonal numbers already available from the other data sources.

Extrapolations in 'non-diagonal' cells in the turnover matrix

530. The extrapolation of turnover in non-diagonal cells is based on a gravity equation. The level of bilateral MNE turnover is estimated using a Gamma pseudo-maximum-likelihood estimation (Santos Silva and Tenreyro, 2006[20]; Head and Mayer, 2014[21]), of the level of bilateral turnover on distance (mostly sourced from CEPII (Mayer and Zignago, 2011[22])), and both jurisdictions' GDP and GDP per capita. This estimation method, which is often used in a trade context, was chosen as it is well-suited to address the challenge arising from the fact that cells in the matrix of bilateral MNE turnover can take either positive or (frequently) zero values.[21] The results, which are reported in Table 5.4, are in line with intuition, in the sense that the size and income levels of both investor and recipient jurisdictions are positively related to bilateral MNE turnover, while distance is negatively related to it.

Table 5.4. Estimated gravity equation used for extrapolations in the turnover matrix

	Dependent variable: Bilateral MNE Turnover (in levels)
GDP of investor (log)	0.478***
	(22.12)
GDP of recipient (log)	0.635***
	(23.81)
Distance (log)	-0.566***
	(-9.51)
GDP per capita of investor (log)	0.684***
	(13.97)
GDP per capita of recipient (log)	0.172***
	(3.84)
Constant	-10.89***
	(-14.74)
N	6512

Note: T-statistics in parentheses. ***, **, *: denote significance at 1, 5, and 10% levels respectively.
Source: OECD Secretariat.

Extrapolations in 'diagonal' cells in the turnover matrix

531. Data on the diagonal of the turnover matrix are extrapolated in a similar fashion as in the profit matrix. The ratio of turnover of domestic-owned MNEs to GDP is regressed on GDP and GDP per capita over the approximately 30 diagonal matrix cells that were filled using CbCR and ORBIS.[22] The regression results are used to extrapolate diagonal cells where none of these data sources is available. Similar to the results on profit of domestic-owned MNEs, the ratio of turnover of domestic-owned MNEs to GDP is significantly positively correlated with GDP per capita, reflecting that richer countries tend to have more MNE headquarters than poorer ones.

5.5.3. Extrapolations in the tangible assets matrix

532. The extrapolations in the tangible assets matrix use the turnover matrix as a starting point. The idea is that tangible assets are generally an important input to production, meaning that turnover and tangible assets are likely to be correlated. Of course, capital intensity varies across firms, implying that the ratio of tangible assets to turnover will depend on the jurisdiction, the economic sector and the MNE considered.[23] To take this heterogeneity into account, at least to the extent possible with the available data, the approach considers how the average ratio of tangible assets to turnover varies across both ultimate parent and affiliate jurisdictions (i.e. across both *columns* and *rows* of the tangible assets matrix), as described in Figure 5.3.

Figure 5.3. Overview of methodology for extrapolations in the tangible assets matrix

Note: The Figure summarises in a stylised way the methodology to extrapolate data in cells of the tangible assets matrix where no hard data is available. The methodology to extrapolate one cell in the tangible assets matrix starts from the value in the corresponding cell in the turnover matrix, and multiplies it with the global average ratio of tangible assets to turnover, adjusted for deviations of this ratio from the global average both by jurisdiction of affiliate (i.e. matrix row) and of ultimate parent (i.e. matrix column). The global average and the deviations are computed based on the available data in CbCR and ORBIS, as further discussed below.
Source: OECD Secretariat

533. For example, a jurisdiction A that has an economy mainly focused on capital-intensive sectors (e.g. commodity extraction, manufacturing) will likely have a relatively high ratio of tangible assets to turnover. This relatively high ratio will be visible in the data relative to affiliates in A from the CbCR data of the ultimate parent jurisdictions listed in Annex 5.A. When extrapolating the tangible assets in A of MNEs from other ultimate parent jurisdictions, the methodology assumes that these MNEs also have a relatively high ratio in A (implicitly assuming that the nature of their activities is similar to other MNEs in A). More formally, the component ‘Delta 1’ in Figure 5.3 will be positive for jurisdiction A. Similarly, if MNEs from a certain ultimate parent jurisdiction tend to have a high ratio of tangible assets to turnover, this will also be taken into account in the extrapolation, thanks to the ‘Delta 2’ component. This ‘Delta 2’ component is computed with consolidated account data from ORBIS, which have good coverage worldwide.[24]

534. The global average ratio of tangible assets to turnover, computed based on CbCR data (after adjusting for inventories, as discussed above), is 33%. This is also the average ratio in ORBIS unconsolidated accounts. To ensure that the adjusted ratio does not take extreme values as a result of the combination of ‘Delta 1’ and ‘Delta 2’ and potential noise in the underlying data, the adjustments are capped. Namely, the individual adjustments (‘Delta 1’ *or* ‘Delta 2’) are such that the imputed ratio of tangible assets to turnover that would occur after each adjustment cannot be outside of a range between 15% and 100%, and the final combined adjustment (‘Delta 1’ *plus* ‘Delta 2’) is then capped so that the final ratio cannot be outside of that range either.

5.5.4. Extrapolations in the payroll matrix

535. The extrapolation methodology in the payroll matrix is very similar to the methodology in the tangible assets matrix. It starts from the turnover matrix, and applies to it an average ratio of payroll to turnover that takes into account heterogeneities across matrix rows and columns (Figure 5.4). The global average ratio of payroll to turnover, computed based on ORBIS unconsolidated data, is 15.3%.[25] Similar to the tangible assets matrix, the ‘Delta 1’ and ‘Delta 2’ parameters are capped both at the individual and combined level so that the imputed ratios are bounded between 5% and 25%.

Figure 5.4. Overview of methodology for extrapolations in the payroll matrix

Note: The Figure summarises in a stylised way the methodology to extrapolate data in cells of the payroll matrix where no hard data is available. The methodology to extrapolate one cell in the payroll matrix starts from the value in the corresponding cell in the turnover matrix, and multiplies it with the global average ratio of payroll to turnover, adjusted for deviations of this ratio from the global average both by jurisdiction of affiliate (i.e. matrix row) and of ultimate parent (i.e. matrix column). The global average and the deviations are computed based on the available data in ORBIS and AMNE.
Source: OECD Secretariat

5.6. Overview of the matrices at an aggregated level

536. The full matrices have more than 200 rows and columns each. To have an overview of the matrices, it is useful to consider compact versions, aggregated by broad groups of jurisdictions, as presented in Table 5.5. The matrices in Table 5.5 are aggregated at the level of four broad groups of jurisdictions (high, middle and low income jurisdictions, and investment hubs). More disaggregated versions of the matrices, with jurisdiction groups combining the income-level dimension and the geographic dimension are presented in Annex 5.D.

Table 5.5. The four matrices: Results aggregated by broad jurisdiction groups

Panel A: The profit matrix

		Jurisdiction of ultimate parent				
	(USD billion of 2016)	High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	3569.1	44.1	0.1	171.3	**3784.5**
	Middle income (105)	366.2	821.8	0.1	167.9	**1356.0**
	Low income (29)	1.3	1.3	3.1	0.2	**5.8**
	Investment Hubs (24)	650.9	69.5	0.0	314.3	**1034.7**
	Total	**4587.4**	**936.7**	**3.3**	**653.7**	**6181.1**

Panel B: The turnover matrix

		Jurisdiction of ultimate parent				
	(USD billion of 2016)	High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	37034.1	943.4	19.0	2602.3	**40598.7**
	Middle income (105)	4392.3	11281.2	11.5	1895.1	**17580.1**
	Low income (29)	50.4	22.4	45.4	11.3	**129.6**
	Investment Hubs (24)	3398.3	176.9	3.6	1487.3	**5066.2**
	Total	**44875.1**	**12423.9**	**79.6**	**5996.0**	**63374.6**

Panel C: The tangible assets matrix						
		Jurisdiction of ultimate parent				
	(USD billion of 2016)	High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	11463.1	314.5	6.2	614.8	**12398.7**
	Middle income (105)	1320.4	4357.9	5.0	757.4	**6440.7**
	Low income (29)	20.5	11.2	17.1	4.2	**53.1**
	Investment Hubs (24)	437.8	69.5	0.9	422.1	**930.3**
	Total	**13241.8**	**4753.2**	**29.2**	**1798.5**	**19822.8**

Panel D: The payroll matrix						
		Jurisdiction of ultimate parent				
	(USD billion of 2016)	High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	6967.3	153.6	3.0	472.2	**7596.2**
	Middle income (105)	497.8	1495.6	1.5	186.5	**2181.4**
	Low income (29)	7.0	3.1	6.8	1.8	**18.7**
	Investment Hubs (24)	225.3	18.2	0.4	170.3	**414.2**
	Total	**7697.5**	**1670.5**	**11.8**	**830.7**	**10210.5**

Note: Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. The number of jurisdictions in each group is indicated in parentheses. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat.

537. The columns totals from the matrices in Table 5.5 confirm that MNE groups with ultimate parents in high income jurisdictions represent the majority of the global activity and profit of MNE groups worldwide (e.g. 71% of turnover, 74% of profit). The row totals in Table 5.5 show that most of the profit and activity of MNE groups worldwide is located in high income jurisdictions (64% of turnover, 61% of profit), and to a lesser extent in middle income jurisdictions (e.g. 28% of turnover, 22% of profit). Interestingly, the share of global MNE profit located in investment hubs (17%) is significantly higher than the share of MNE activity in investment hubs (8% of turnover, 5% of tangible assets, 4% of payroll), consistent with earlier evidence of profit shifting behaviour by MNEs (Johansson et al., 2017[18]; Beer, de Mooij and Liu, 2019[23]).

5.7. Relative importance of data sources in the matrices

538. The relative importance of each data source in each matrix is summarised in Table 5.6. It is presented both (i) as a percentage of the total number of cells in each matrix, and (ii) as a percentage of the total amount of the variable considered (e.g. as a percentage of all profit in the profit matrix). This second measure is generally more informative. This is because the total number of cells in each matrix is very high (almost 50,000 cells), reflecting the wide range of jurisdictions covered (222 in total, some of which are very small, see full list in Annex Table 5.D.2). As a result, a large majority of cells contain very small values and have limited importance, either because there are few MNEs with an ultimate parent in the ultimate parent jurisdiction considered (as is the case for many small and developing economies) or because the 'ultimate parent'-'affiliate' pair of jurisdictions considered in the cell have little economic relationship with each other (e.g. because of high geographic distance between the two jurisdictions). Consequently, even though more than 90% of the cells in the matrices are filled in by extrapolations, those represent only about 25% of the total amounts on average across the four matrices.

Table 5.6. Relative importance of data sources in the matrices

	Profit matrix		Turnover matrix		Tangible assets matrix		Payroll matrix	
	% of cells	% of total profit	% of cells	% of total turnover	% of cells	% of total tangible assets	% of cells	% of total payroll
CbCR data	2%	63%	2%	58%	2%	60%	--	--
ORBIS data	3%	10%	3%	11%	3%	12%	3%	17%
Analytical AMNE data	--	--	4%	26%	--	--	--	--
AMNE data	--	--	5%	0%	--	--	2%	46%
Extrapolations	95%	27%	86%	4%	95%	28%	96%	36%
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Note: For example, 2% of cells in the profit matrix are filled with CbCR data. These cells contain 63% of the total profit in the profit matrix. Cells filled with "-- " correspond to data sources not used in the matrix considered.
Source: OECD Secretariat.

539. The statistics in Table 5.6 indicate that CbCR data are the primary source of data in the profit, turnover and tangible assets matrices, representing about 60% of the total amount of each of these variables in their respective matrices. The share of extrapolations is much lower in the turnover matrix (4%) than in the three other matrices (27-36%), suggesting that the turnover matrix may be the most reliable of the four matrices.

540. The relative importance of extrapolations across jurisdiction groups is presented in Table 5.7. In general, the matrices rely to a relatively low extent on extrapolations for data relative to high income jurisdictions, both across matrix rows (i.e. MNE entities in high income jurisdictions) and columns (MNE entities with an ultimate parent in a high income jurisdiction). For example, extrapolations represent 14% of data in high income jurisdiction rows in the profit matrix, and also 14% in high income jurisdiction columns in the profit matrix. The share of extrapolations is much higher in middle income jurisdictions (above 50% for all variables except turnover, where it is below 10%) and it approaches 100% in low income jurisdictions, reflecting the wide gaps in the existing hard data in these jurisdictions. Finally, the share of extrapolated profit in matrix rows corresponding to investment hubs is below 40%. While still substantial, this number highlights the importance of the newly available CbCR data as a source of data on profit (and other variables) in investment hubs, as other data sources have generally poor coverage of investment hubs.

Table 5.7. Relative importance of extrapolations in the matrices, by broad jurisdiction groups

Panel A: Share of extrapolated data by matrix rows

	Profit matrix	Turnover matrix	Tangible assets matrix	Payroll matrix
High income (64 jurisd.)	14%	3%	14%	19%
Middle income (105)	53%	8%	54%	93%
Low income (29)	77%	94%	93%	97%
Investment Hubs (24)	39%	4%	36%	60%
Global average	**27%**	**4%**	**28%**	**36%**

Panel B: Share of extrapolated data by matrix columns

	Profit matrix	Turnover matrix	Tangible assets matrix	Payroll matrix
High income (64 jurisd.)	14%	3%	15%	22%
Middle income (105)	72%	9%	64%	95%
Low income (29)	100%	100%	100%	100%
Investment Hubs (24)	51%	5%	31%	49%
Global average	**27%**	**4%**	**28%**	**36%**

Note: For example, 39% of data in *rows* corresponding to investment hubs in the profit matrix (i.e. profit of MNE affiliates in investment hubs) are extrapolated (Panel A), and 51% of data in *columns* corresponding to investments hubs in the profit matrix (i.e. profit of MNEs with an ultimate parent in an investment hub) are extrapolated (Panel B). Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP. The number of jurisdictions in each group is indicated in parentheses.
Source: OECD Secretariat.

5.8. Benchmarking and consistency checks

541. As the matrices combine data from a range of sources, as well as extrapolations, ensuring a sufficient degree of data consistency within and across the matrices is a key challenge. While the data sources used in the matrices all focus on measuring similar variables (e.g. MNE profits), there are inevitably some conceptual differences between them. In particular, the sources used can have different coverage (e.g. CbCR data includes only MNE groups with global revenues above the EUR 750 million reporting threshold, AMNE data excludes financial sectors), different variable definitions (e.g. profit in CbCR data may include some intracompany dividends received, while all dividends are included in the PBT measure from ORBIS and excluded in the EBIT measure) and different ownership rates to define multinationals (e.g. CbCR and ORBIS data are based on a 50% ownership threshold, while FDI data are based on a 10% threshold).

542. While the methodology to fill the matrices aims to address some of these inconsistencies (e.g. by using values in matrix cells filled with hard data in the procedures to extrapolate data to other cells, which enhances the consistency of the extrapolations with the available hard data), the persistence of some inconsistencies between sources is to some extent unavoidable. To gauge the magnitude of these inconsistencies, and to minimise their potential impact on final results, extensive benchmarking and quality checks have been undertaken. They are described in the following sections.

5.8.1. Correlation across the data sources used within the matrices

543. The correlation of data across sources within each matrix has been systematically tested over the cells where several sources are available. As the extrapolation methodologies employed in this chapter produce values for all cells, all cells filled with hard data can be compared at least to the extrapolation results. In addition, there is substantial coverage overlap across the hard data sources, which provides the possibility to consider correlation across these sources as well.

544. The results, presented in Table 5.8 suggest that pairwise correlation of sources is good, especially among hard data sources. For example, correlation reaches 92% between CbCR and ORBIS data in the profit matrix, and 93% in the turnover matrix. Correlation between hard data and extrapolated values is generally lower, as could be expected (between 64% and 96% depending on the variable and data source considered) but still sufficiently high to consider the extrapolations relevant. Even in the profit matrix, where extrapolation is arguably more challenging than in the other matrices because identifying the location of MNE profit is complicated by the fact that it can differ from the location of MNE economic activity, the correlation of extrapolated values with hard data reaches 75%.

Table 5.8. Pairwise correlation of sources in the matrices

Correlation between each pair of sources in each matrix, and number of observations where the sources overlaps

Panel A: Profit matrix

	CbCR data	ORBIS data	Extrapolations
CbCR data	--	0.92 (252 obs.)	0.75 (955 obs.)
ORBIS data	--	--	0.75 (1483 obs.)
Extrapolations	--	--	--

Panel B: Turnover matrix

	CbCR data	ORBIS data	Analytical AMNE data	AMNE data	Extrapolations
CbCR data	--	0.93 (254 obs.)	0.75 (561 obs.)	0.88 (417 obs.)	0.73 (1064 obs.)
ORBIS data	--	--	0.79 (941 obs.)	0.92 (365 obs.)	0.64 (1588 obs.)
Analytical AMNE data	--	--	--	0.95 (867 obs.)	0.75 (2583 obs.)
AMNE data	--	--	--	--	0.71 (1530 obs.)
Extrapolations	--	--	--	--	--

Panel C: Tangible assets matrix

	CbCR data	ORBIS data	Extrapolations
CbCR data	--	0.84 (251 obs.)	0.89 (1038 obs.)
ORBIS data	--	--	0.87 (1493 obs.)
Extrapolations	--	--	--

Panel D: Payroll matrix			
	ORBIS data	AMNE data	Extrapolations
ORBIS data	--	0.94 (598 obs.)	0.95 (1171 obs.)
AMNE data	--	--	0.96 (1321 obs.)
Extrapolations	--	--	--

Note: This table reports pairwise correlation of data sources in the four matrices, each panel corresponding to a matrix. For each pair of sources, the correlation is computed across the values in all matrix cells where both sources are available. For example, the correlation between the profit data from CbCR and ORBIS in the profit matrix, computed across the 252 profit matrix cells where both sources are available, is 92%. Given the extrapolation methodology employed, the extrapolations are available for comparison across all cells of all four matrices. Values are log-transformed before computing the correlations to avoid that extreme values have a disproportionate effect on the correlations. Correlations without this log-transformation are generally higher than those presented in this Table.
Source: OECD Secretariat.

545. Beyond the average correlation coefficients presented in Table 5.8, the consistency between data sources has been explored in more detail, based on an extensive set of comparison scatterplots. To remain readable, such scatterplots typically have to focus on comparing just one data source with another, in only one given matrix row or column. While these scatterplots would take too much space to be exhaustively reported in this chapter, illustrative examples are provided in Annex Figure 5.E.1, which compares CbCR data and ORBIS data in the columns of the profit, turnover and tangible assets matrices corresponding to ultimate parent jurisdictions where CbCR data have been reported for more than ten bilateral entries (payroll is not included in the Figure as it is not included in CbCR data).

546. This detailed exploration of the data has allowed the OECD Secretariat to identify a number of suspect values, defined as values for which several data sources were indicating very different outcomes. These values have been corrected manually, when necessary, after more detailed investigation to identify the most relevant data source – including via exchanges with jurisdiction representatives. As this exploration has primarily been based on a preliminary version of the CbCR data published in July 2020, it has allowed the identification of a few reporting errors in the CbCR data, which have been corrected before publication after exchanges with jurisdiction representatives.

5.8.2. Correlation with alternative data sources not used in the matrices

547. A number of data sources have not been used directly to fill the matrices, but nevertheless represent useful benchmarks to assess data quality and identify potential data consistency issues in the matrices. Three alternative sources have been considered in this chapter:

- **EBIT instead of PBT as a measure of profit in ORBIS data**: As discussed in section 5.4.2, two measures of profit are available in ORBIS (EBIT and PBT). PBT has been preferred to EBIT as the variable used to fill the profit matrix when using ORBIS data, but EBIT offers an interesting comparison point. In particular, the fact that EBIT excludes dividends received makes the comparison with CBCR data (where intra-group dividends are partly included) and ORBIS PBT data (where they are fully included) illustrative of the influence of dividends in the profit matrix. As shown in Table 5.9 (row 1), the correlation of EBIT with CbCR data is high (92%) and the correlation with ORBIS PBT data (98%) even higher, suggesting that the inclusion of dividends (partial or total) in these sources has only limited average consequences on the overall profit matrix, even though certain cells may be more affected than others (e.g. dividends are likely to be larger in jurisdictions with a larger share of parent or holding companies).
- **OECD AMNE data on gross operating surplus**: As discussed in section 5.2.3, the (inward and outward) OECD AMNE database contains bilateral data on MNE gross operating surplus. This data

has not been used to fill in the profit matrix because it adds little coverage beyond CbCR data and ORBIS, does not always cover all economic sectors, and because gross operating surplus has conceptual differences with the measure of profit in financial accounts due notably to differences in depreciation rules. Still, data on gross operating surplus provide a useful comparison point. To enhance comparability with other sources, adjustments inspired by Tørsløv et al. (2018[1]) are applied to adjust for depreciation.[26] The correlation of the adjusted variable with CbCR data exceeds 80% for both inward and outward AMNE, and the correlation with ORBIS exceeds 85% (Table 5.9, rows 2 and 3). These correlations are overall lower than the CbCR/ORBIS correlation in Table 5.8 (i.e. 92%), which tends to confirm that adjusted gross operating surplus based on OECD AMNE data was less suitable than these two sources for the purpose of filling the profit matrix.

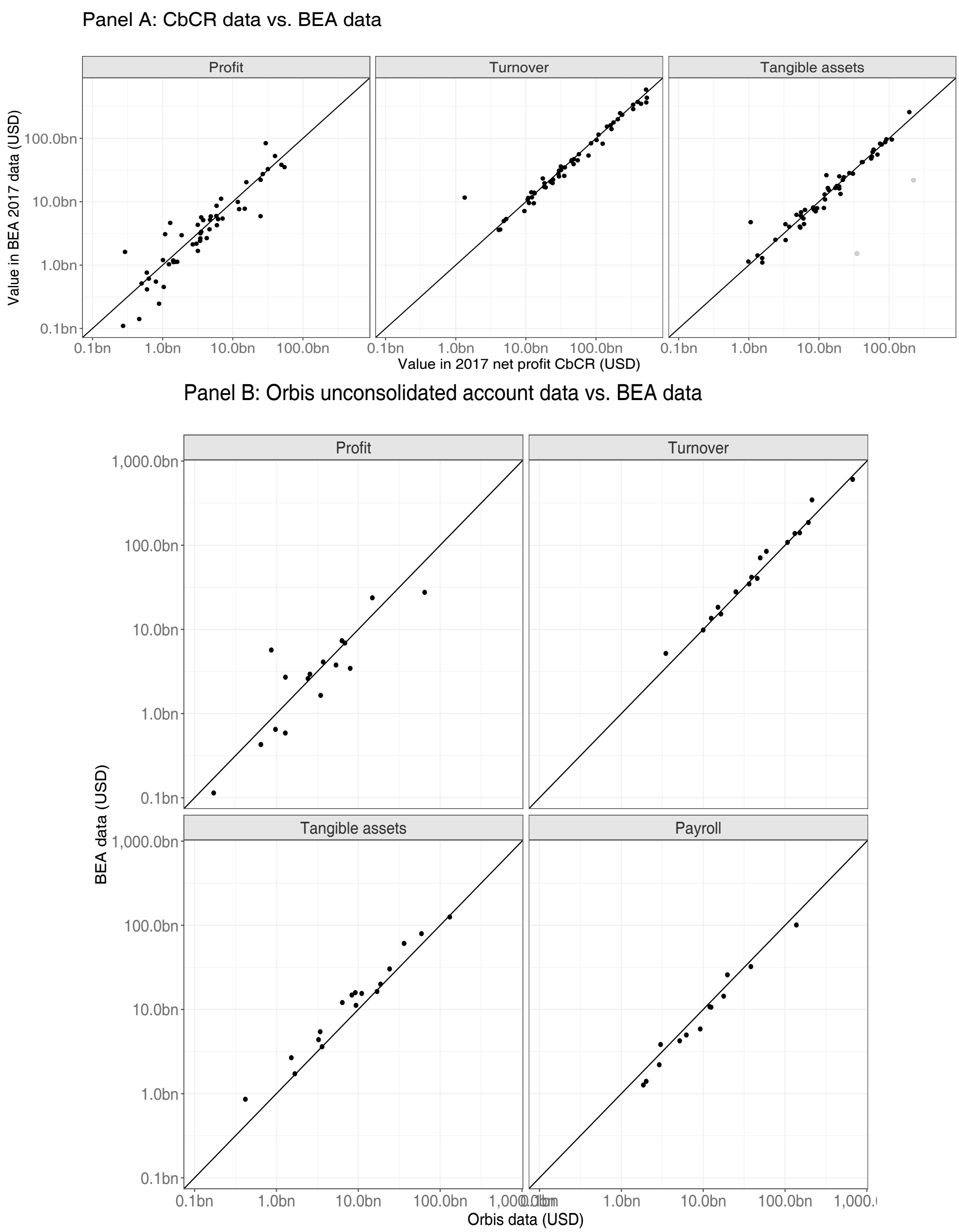
- **Data from the US BEA on US MNEs:** As discussed in section 5.2.6, the US BEA provides detailed data on the foreign activity of US MNEs. The BEA data on profit ('profit-type return'), turnover ('sales of US affiliates'), tangible assets ('property, plant and equipment'), and payroll ('compensation of employees') have been compared with US CbCR data (except for payroll, which is not available in CbCR) as well as ORBIS data on affiliates of US MNEs abroad. As BEA data does not distinguish 'positive' profits (i.e. profits from entities belonging to profit-making MNE sub-groups in the jurisdiction considered) from net profits, the comparison focuses on net profits (while only sub-groups with positive profits from CbCR and ORBIS are considered in the matrices). Overall, the correlation of CbCR and ORBIS data with the relevant BEA data is high (Figure 5.5). The comparison has also allowed for the identification of a few outlying values, corresponding to reporting errors or exceptional events in the data underlying the profit matrix, which have been corrected by replacing them with the corresponding values from the BEA data.

Table 5.9. Pairwise correlation of the data used in the profit matrix with alternative measures of profit from other data sources

		Matrix sources		
		CbCR data	ORBIS (PBT)	Extrapolations
Other sources	ORBIS (EBIT)	0.92	0.98	0.75
	OECD AMNE (inward)	0.84	0.86	0.76
	OECD AMNE (outward)	0.81	0.92	0.69

Note: This table reports pairwise correlation of data sources in the profit matrix (in column) and three alternative measures of profit from data sources which are not used in the profit matrix (in rows). The alternative measures considered are EBIT (sourced from ORBIS database), and gross operating surplus, adjusted for depreciation using data from Tørsløv et al. (2018[1]), in both inward and outward OECD AMNE data. For each pair of sources, the correlation is computed across all matrix cells where both sources are available. For example, the correlation between the profit data from CbCR and ORBIS EBIT data, computed across the profit matrix cells where both sources are available is 92%. Values are log-transformed before computing the correlations to avoid that extreme values have a disproportionate effect on the correlations. Correlations without this log-transformation are generally higher than those presented in this Table.
Source: OECD Secretariat.

Figure 5.5. US MNEs: Comparison of CBCR and ORBIS data with data from the US BEA



Note: These panels compare data from the US CbCR (Panel A) and ORBIS (Panel B) with data from the US BEA on the activity of multinationals. Data in Panel A relates to 2017 (the year for which US CbCR data are used in the analysis) and data in Panel B relates to 2016 (the year for which ORBIS data are used in the analysis). Each dot in the graphs corresponds to a jurisdiction where US MNEs have affiliates. A perfect correlation would imply that all dots are on the 45-degree line, which is overlayed on the graphs. The grey dots in the tangible assets panel for CbCR (Panel A), representing data for Luxembourg and Barbados, have been deemed outliers, suggesting reporting errors or exceptional events in the underlying CbCR data, and replaced manually by the corresponding values from the BEA data in the profit matrix. For greater comparability, the tangible assets variable considered in the BEA data includes inventories when comparing with CbCR data, and excludes inventories when comparing with ORBIS.

Source: OECD Secretariat.

5.8.3. Comparisons with financial data from MNE consolidated accounts

548. Column totals in the matrices have been benchmarked against information from consolidated financial accounts of MNE groups. For example, the total profit from US MNE groups – based on their consolidated financial accounts – should in theory correspond to the total in the US column of the profit matrix. However, there are several reasons why this comparison is imperfect: (i) the matrices focus on MNE sub-groups with positive profits (i.e. entities belonging to MNE sub-groups with positive profits in the jurisdiction considered), while data from consolidated financial accounts of MNEs consider all MNE entities regardless of their profit position in specific jurisdictions; as a result, overall profit is expected to be higher in the profit matrix than in consolidated financial accounts (where profit is net of the losses of loss-making entities); (ii) the turnover matrix focuses both on sales to third-party and related-party entities, while sales to related-party entities are not included in consolidated financial accounts.

549. Consolidated financial accounts from ORBIS have relatively good coverage of MNEs worldwide (contrary to unconsolidated accounts, where coverage is uneven across jurisdictions). To ensure even better coverage, the ORBIS dataset has been complemented by the OECD Secretariat with: (i) data from the Worldscope database, which contains the financial accounts of (mainly listed) firms worldwide; (ii) data from the EU Industrial R&D Investment Scoreboard, which covers the 2,500 companies with the highest level of R&D spending worldwide (Hernández et al., 2017[24]); and (iii) data from the Fortune Global 500 list (i.e. 500 firms with the highest turnover globally) (see Annex 2.A of Chapter 2 for more details).

550. Overall, there is a correlation between the column totals in the matrices and financial data from consolidated accounts, but it is imperfect for the reasons listed above (Annex Figure 5.E.2). While the differences between the two are generally moderate, they seem implausibly large in some jurisdictions of ultimate parent, even accounting for potential gaps in the coverage of consolidated account data. This may relate to fragilities in the extrapolation of diagonal cells in the profit and turnover matrices, which as discussed above entails a greater degree of uncertainty than data and extrapolations in other cells due to scarcer data availability. To address this issue, estimates in the profit and turnover matrix diagonal have been adjusted – when they relied on extrapolations, and when the difference between the column total in the profit matrix and the corresponding total from consolidated financial account data was deemed implausible – to bring back this difference to a more plausible level.[27] While inevitably imperfect due to data limitations, this adjustment aims to reduce the risk of implausible values in the matrices.[28]

5.8.4. Interactions with jurisdiction representatives

551. Bilateral interactions with jurisdiction representatives have helped refine the data in the matrices in two main ways. First, the data in the matrices has been benchmarked against tax and financial account data on the profit and activity of domestic-owned and foreign-owned MNEs provided to the OECD Secretariat on a confidential basis through a questionnaire filled by Delegates to Working Party No. 2 of the OECD's Committee on Fiscal Affairs in 2019. This has contributed to the identification of potential inconsistencies with data in the matrices.

552. Second, the values in some specific cells of the matrices have been the subject of bilateral discussions with jurisdiction representatives, either to identify the most relevant data source in the case of conflicting signals across sources (as discussed above), to eliminate from the data the effect of large one-off events (e.g. effect of large mergers and acquisitions on the amount and location of profit across jurisdictions) or to address limitations in the existing data sources (e.g. issues related to intra-company dividends in CbCR data). In this context, some representatives have been able to provide the OECD Secretariat with their estimates for specific matrix cells (e.g. diagonal cells in the profit matrix relative to their jurisdiction) based on detailed analysis of their national data. These estimates have been integrated in the relevant matrices, reflecting that, notwithstanding potential consistency issues with other data sources in the matrices, these estimates were deemed to be more accurate than the estimates based on the sources presented in this chapter.

5.8.5. Robustness check: Changing the order of preference between data sources

553. As the data sources underlying the matrices have a certain degree of overlap, the matrices are built with a preference order between sources, which is presented in Table 5.2. An interesting robustness check is to modify the ordering of sources and to compare the result with the original matrices. The result of this robustness check is presented in Annex 5.F, at a relatively aggregate level (i.e. the level of aggregation of the matrices presented in Table 5.5) for the experiment of inverting the order of preference between the first two data sources in each matrix. For example, in the profit matrix, the baseline order of preference is (i) CbCR data, (ii) ORBIS data, (iii) extrapolations. The experiment in Annex 5.F is to apply the following order instead (i) ORBIS data, (ii) CbCR data, (iii) extrapolations.

554. Overall, the results of this robustness check are reassuring, in the sense that most values (at the aggregate level presented in Annex 5.F) are modified by less than 10% with the alternative source ordering compared to the baseline.

5.9. Conclusion

555. This chapter describes the methodology and data sources underlying a set of matrices mapping the profit and activity of MNEs worldwide. These matrices have been central instruments in supporting the assessment of Pillar One and Pillar Two presented in this report. The matrices combine a range of existing data sources into a consistent framework, enabling a more accurate and comprehensive reform impact assessment than any individual source taken in isolation. While the matrices inevitably rely to some extent on extrapolations, sophisticated methods have been used to make the extrapolated values as accurate as possible. Another benefit of the methodology is to make these extrapolations explicit, which makes it possible to assess their quality.

556. Given the limitations of the underlying data sources, the methodology cannot go as far as offering a precise and certain estimate of each individual data point in each matrix. Rather, the aim of the methodology has been to ensure that the data points in the matrices have the right order of magnitude and offer a good level of consistency within and across matrices. For the purpose of the impact assessment in this report, overall consistency is even more important than the precise accuracy of individual points, as the assessment is generally based on combining data within and across matrices. To ensure the maximum level of consistency, extensive data checks and benchmarking have been undertaken, including via interactions with jurisdiction representatives.

557. Looking ahead, the matrices presented in this chapter could be used in the future for a range of other purposes. For example, they could be used for the analysis of MNE profit shifting behaviour, where they have potential to provide a more comprehensive and detailed picture of profit shifting than the earlier studies predating the publication of CbCR data (e.g. Johansson et al. (2017[18]) Tørsløv et al. (2018[1]) Beer et al. (2019[23])) or recent studies using the publicly available CbCR data of US MNEs (e.g. Clausing (2020[15])).[29] Preliminary estimates on profit shifting based on the matrices presented in this chapter and undertaken in the analysis of the potential MNE behavioural reactions to Pillar Two (presented in Chapter 3) suggest an overall intensity of profit shifting that is broadly consistent with these earlier studies, but offer a more detailed mapping of profit shifting flows. Another potential use of the matrices would be the analysis of potential fundamental corporate tax reforms. Beyond tax-related analysis, the matrices could also be used, in combination with data from the OECD Analytical AMNE database, to assess the links between global value chains and MNE profitability, for example to inform future analyses on the restructuring of those global value chains in a post-COVID-19 environment.

558. Another avenue for future work would be to update and refine these matrices as more data become available and to reflect the ongoing changes in the economic situation, including most prominently the COVID-19 crisis. The current matrices focus primarily on year 2016, which was the latest available year in

both CbCR and ORBIS data when the matrices were constructed. In particular, CbCR data are expected to gradually cover a wider set of ultimate parent jurisdictions in coming years, reflecting that 2016 was the first year where it was collected and that some jurisdictions were not yet in a position to compile and provide CbCR data for that year to the OECD for publication. The quality of CbCR data is also expected to improve in the future, since some issues related to its collection, which have been mentioned in this chapter, have been identified and are being addressed. As the quality and coverage of CbCR data improves, it may be possible to build more precise matrices in the future.

References

Bailin, A. et al. (2019), "Like it or not? The impact of online platforms on the productivity of service providers", *OECD Economics Department Working Papers*, No. 1548, OECD Publishing, Paris, http://dx.doi.org/10.1787/080a17ce-en. [19]

Bajgar, M. et al. (2020), "Coverage and representativeness of Orbis data", *OECD Science, Technology and Industry Working Papers*, No. 2020/06, OECD Publishing, Paris, https://dx.doi.org/10.1787/c7bdaa03-en. [26]

Bajgar, M. et al. (2019), "Industry Concentration in Europe and North America", *OECD Productivity Working Papers*, No. 19, OECD, https://www.oecd-ilibrary.org/economics/industry-concentration-in-europe-and-north-america_2ff98246-en. [16]

Beer, S., R. de Mooij and L. Liu (2019), "International Corporate Tax Avoidance: A Review of the Channels, Magnitudes, and Blind Spots", *Journal of Economic Surveys*, pp. 1-29, http://dx.doi.org/10.1111/joes.12305. [23]

Blouin, J. and L. Robinson (2019), "Double Counting Accounting: How Much Profit of Multinational Enterprises Is Really in Tax Havens?", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3491451. [14]

Bolwijn, R., B. Casella and D. Rigo (2018), *An FDI-driven approach to measuring the scale and economic impact of BEPS*, United Nations Conference on Trade and Development. [11]

Borga, M. and C. Caliandro (2018), "Eliminating the Pass-Through: Towards FDI Statistics that Better Capture the Financial and Economic Linkages between Countries", *NBER Working Paper*, No. 25029, http://dx.doi.org/10.3386/w25029. [12]

Cadestin, C. et al. (2018), "Multinational enterprises and global value chains: the OECD Analytical AMNE database", *OECD Trade Policy Papers*, No. 211, OECD Publishing, Paris, https://dx.doi.org/10.1787/d9de288d-en. [2]

Casella, B. (2019), "Looking through conduit FDI in search of ultimate investors-a probabilistic approach", No. 26(1), United Nations Conference on Trade and Development, https://unctad.org/en/PublicationChapters/diaeia2019d1a5_en.pdf. [5]

Clausing, K. (2020), "Profit Shifting Before and After the Tax Cuts and Jobs Act", *SSRN Electronic Journal*, http://dx.doi.org/10.2139/ssrn.3274827. [15]

Coppola, A. et al. (2020), "Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens", *NBER Working Paper*, No. 26855, https://www.nber.org/papers/w26855. [27]

Damgaard, J. and T. Elkjaer (2017), “The Global FDI Network: Searching for Ultimate Investors”, *IMF Working Papers*, No. 17/258, https://www.imf.org/en/Publications/WP/Issues/2017/11/17/The-Global-FDI-Network-Searching-for-Ultimate-Investors-45414. [4]

Damgaard, J., T. Elkjaer and N. Johannesen (2019), “What Is Real and What Is Not in the Global FDI Network?”, *IMF Working Papers*, Vol. 19/274, http://dx.doi.org/10.5089/9781513521527.001. [10]

Eurostat (2008), “NACE Rev. 2 - Statistical classification of economic activities in the European Community”, *Eurostat Methodologies and Working papers*, Eurostat, Luxembourg, https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF. [9]

Gal, P. (2013), “Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS”, *OECD Economics Department Working Papers*, No. 1049, OECD Publishing, Paris, https://dx.doi.org/10.1787/5k46dsb25ls6-en. [17]

Hanappi, T. and A. González Cabral (2020), “The impact of the pillar one and pillar two proposals on MNE’s investment costs : An analysis using forward-looking effective tax rates”, *OECD Taxation Working Papers*, No. 50, OECD Publishing, Paris, https://dx.doi.org/10.1787/b0876dcf-en. [6]

Head, K. and T. Mayer (2014), “Gravity Equations: Workhorse,Toolkit, and Cookbook”, *Handbook of International Economics*, Vol. 4, pp. 131-195, http://dx.doi.org/10.1016/B978-0-444-54314-1.00003-3. [21]

Hernández, H. et al. (2017), *The 2017 EU Industrial R&D Investment Scoreboard*, European Commission, https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/2017-eu-industrial-rd-investment-scoreboard. [24]

Johansson, Å. et al. (2017), “Tax planning by multinational firms: Firm-level evidence from a cross-country database”, *OECD Economics Department Working Papers*, No. 1355, OECD Publishing, Paris, https://dx.doi.org/10.1787/9ea89b4d-en. [18]

Mayer, T. and S. Zignago (2011), “Notes on CEPII’s Distances Measures: The GeoDist Database”, *CEPII Working Papers*, No. 2011-25, CEPII, http://dx.doi.org/10.2139/ssrn.1994531. [22]

OECD (2020), “Corporate Tax Statistics - Second Edition”, OECD Publishing, https://www.oecd.org/tax/tax-policy/corporate-tax-statistics-second-edition.pdf. [3]

OECD (2018), *Multinational enterprises in the global economy: Heavily debated but hardly measured*, OECD Publishing, Paris, http://www.oecd.org/industry/ind/MNEs-in-the-global-economy-policy-note.pdf. [8]

OECD (2015), *Implementing the latest international standards for compiling foreign direct investment statistics - FDI Statistics by the Ultimate Investing Country*, https://www.oecd.org/daf/inv/FDI-statistics-by-ultimate-investing-country.pdf. [13]

OECD (2015), *Measuring and Monitoring BEPS, Action 11 - 2015 Final Report*, OECD/G20 Base Erosion and Profit Shifting Project, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264241343-en. [7]

Santos Silva, J. and S. Tenreyro (2006), "The log of gravity", *Review of Economics and Statistics*, Vol. 88/4, pp. 641-658, http://dx.doi.org/10.1162/rest.88.4.641. [20]

Tørsløv, T., L. Wier and G. Zucman (2018), "The Missing Profits of Nations", *NBER Working Paper*, No. 24701, http://www.nber.org/papers/w24701.pdf. [1]

UNCTAD (2018), *World Investment Report 2018: Investment and New Industrial Policies*, https://unctad.org/en/PublicationsLibrary/wir2018_en.pdf. [25]

Annex 5.A. List of jurisdictions covered by the main data sources

Annex Table 5.A.1. List of jurisdictions of ultimate parent for which anonymised and aggregated CbCR data are used in the matrices

Australia	India	Norway
Austria	Indonesia	Poland
Belgium	Ireland	Singapore
Bermuda	Italy	Slovenia
Canada	Japan	South Africa
Chile	Korea	Sweden
Denmark	Luxembourg	United States (2017)
Finland	Mexico	
France	Netherlands	

Note: Out of the 26 jurisdictions of ultimate parent included in the OECD publication of 2016 aggregated and anonymised CbCR data in July 2020, data from one ultimate parent jurisdiction (China) were not used in the analysis in this chapter. This is because Chinese CbCR data for 2016 are based only on a subsample of 82 CbCRs, while it is estimated that many more CbCRs were filed in China for the fiscal year 2016. CbCR data for Brazil have recently been added to the online CbCR database of the OECD, but was not available at the time when this analysis was undertaken. For all jurisdictions in this Table except the United States, CbCR data for 2016 are used, as 2017 data are not available yet. For the United States, CbCR data for 2017 are used, as they are more complete than 2016 data (in 2016, CbCR were filed on a voluntary basis in the United States).
Source: OECD Secretariat.

Annex Table 5.A.2. List of jurisdictions of affiliate for which ORBIS unconsolidated account data are used

Good coverage for both domestic-owned and foreign-owned MNE entities		Good coverage only for foreign-owned MNE entities
Australia	Latvia	Bulgaria
Belgium	Lithuania	China
Croatia	Norway	
Czech Republic	Poland	
Denmark	Portugal	
Estonia	Russia	
Finland	Slovak Republic	
France	Slovenia	
Greece	Spain	
Italy	Sweden	
Korea	United Kingdom	

Note: For each jurisdiction of affiliate, the quality of ORBIS coverage has been assessed based on a comparison with CbCR data and aggregate numbers from the Analytical AMNE database. Jurisdictions considered as having good coverage in ORBIS are those having at least 750 observations, where coverage of MNE turnover (assessed against Analytical AMNE data) is above 70%, and for which a comparison against available CbCR data (across all jurisdictions of ultimate parent where CbCR data are available to the OECD Secretariat) does not suggest major discrepancies. The coverage of the cost of employees variable is relatively poor in China, Greece, Latvia, Lithuania and Russia, and ORBIS data has not been used for this specific variable in these countries. In Korea, the definition of cost of employees in ORBIS may not be fully consistent with other jurisdictions, so ORBIS data have not been used for this specific variable.
Source: OECD Secretariat.

Annex 5.B. ORBIS data cleaning

559. The ORBIS database, provided by Bureau van Dijk (BvD), is the largest cross-country database on ownership and financial accounts of firms worldwide. It relies on information from various underlying sources, and contains data for both publicly listed and privately owned companies.

560. Given that ORBIS data is not primarily collected for statistical analysis, important processing and cleaning work is required to enhance data reliability (e.g. eliminating duplicates and reporting errors). This concerns ownership data and financial data. The main data cleaning steps in each area build on OECD expertise with ORBIS and follow as much as possible procedures used in previous OECD studies, while adapting them when necessary to the needs of the current exercise. These main cleaning steps are detailed in the sections below.

Cleaning of ORBIS ownership data

561. The ORBIS historical ownership database contains extensive information on ownership links between firms, which can be used to identify entities belonging to the same corporate group. Following Bajgar et al. (2019[16]), entities in ORBIS are assigned to corporate groups based on their Global Ultimate Owner (GUO), using a 50% ownership threshold, and considering GUOs of corporate nature (i.e. Industrial companies, Banks, Financial companies, Insurance companies, or Financial companies) to avoid for example assigning to the same group two independent firms owned by the same individual or government entity.

562. In turn, MNE groups are defined as corporate groups having entities in at least two jurisdictions. For each MNE group, only the consolidated accounts of the GUO are kept in the sample, to avoid potential double counting.

563. The procedure to clean and extend ownership links in ORBIS has been implemented by the OECD Directorate for Science, Technology and Innovation, following Bajgar et al. (2019[16]) and updating it for year 2016. The procedure focuses on all entities with a turnover of at least EUR 10 million, and focuses on ownership links above a 50% threshold. Missing links are identified, or (in a smaller number of cases) existing links are corrected, using the following steps:

- Using the BvD Zephyr database on Merger and Acquisition (M&A) to identify changes in immediate (rather than global ultimate) owners not available from ORBIS.
- Using ORBIS historic ownership linkages to identify changes in immediate owners not available from ORBIS.
- Translating the changes in immediate owners (from the first two steps above) to changes in ultimate ownership.
- Imputing missing ownership information by using data on M&A or changes in ownership in earlier or later years.
- Correcting ultimate owners that are in fact majority owned by another firm, since by definition they cannot be an ultimate owner.
- Removing temporary (one or two year) changes in ultimate owner that reverse themselves – as such cases seem highly unlikely to occur in reality and probably reflect gaps in in ownership data.
- Detecting missing linkages for large firms that change from having no subsidiaries to having a large number of subsidiaries one year to the next.

- Identifying missing links for large firms that never have any subsidiaries, and for large groups of subsidiaries that never have a parent with financials.
- Using name-matching algorithms to identify potential links, in combination with detailed manual inspection (e.g. against firms' annual reports) to check if these potential links are correct or not.
- Manually checking the 300 largest firms, using the subsidiary structure in their financial statements to cross-check the ownership data.

564. Overall, this procedure identified the GUO of about 50,000 entities for which it was not reported in the raw ORBIS data, and corrected the GUO of about 4,000 entities. Overall, these entities (added and corrected GUO) represent about 4% of global MNE turnover (as measured with the ORBIS dataset of consolidated MNE group accounts).

Cleaning of ORBIS unconsolidated financial account data

565. The sample of unconsolidated account data is restricted to entities belonging to MNE groups, i.e. corporate groups that have entities in at least two jurisdictions, as identified with the ownership data cleaned with the procedure above.

566. The procedure for cleaning unconsolidated account data comprises the following steps, inspired by the cleaning steps in Gal (2013[17]), Johansson et al. (2017[18]) and Bailin et al. (2019[19]) that are relevant for this exercise:

- Selecting full-year accounts with closing date around December 2016 (from July 2016 to June 2017);
- Filtering duplicate firm-year observations, favouring those with non-missing key financial variables and with closing date equal to or closest to 31st of December;
- Eliminating implausible values: negative tangible assets, turnover, or cost of employees, implausibly high profit or turnover, tangible assets values above fixed assets or total assets values;
- Eliminating jumps in the turnover variable, i.e. situations where this variable is multiplied or divided by more than 5 in one year over 2014-2016;
- Eliminating implausible values: tax expenses higher than pre-tax profit, pre-tax profit minus tax expenditure inconsistent with post-tax profit;
- Eliminating outliers based on the following ratios of interest: EBIT to turnover, Profit before tax to turnover, Tangible assets to turnover and Cost of employees to turnover (keeping observations between the 2.5th and 97.5th percentiles of the distribution).

Annex 5.C. Detailed methodology for the extrapolations based on FDI in the profit matrix

567. As described in the main text, the profit matrix is filled primarily with CbCR and ORBIS data. For matrix cells not covered by these sources, the approach is to rely on extrapolations based on macroeconomic variables (e.g. FDI positions). This annex describes in detail the sophisticated extrapolation procedure employed for non-diagonal cells of the profit matrix, i.e. for profit of MNEs outside of their jurisdiction of ultimate parent. This extrapolation consists of four steps, which are further described in the following sections:

- **Step 1**: building a full matrix of FDI positions by immediate investor, by (i) combining the available data in bilateral FDI statistics and (ii) extrapolating FDI positions to fill the data gaps in bilateral FDI statistics, based on a standard gravity model;
- **Step 2**: building a full matrix of FDI positions by ultimate investor jurisdiction. Existing OECD data on FDI by ultimate investor are used in the subset of jurisdictions where they are available. In other jurisdictions, estimates of FDI by ultimate investor are derived from the full matrix of FDI positions by immediate investor obtained in step 1, applying a methodology developed by Casella (2019[5]);
- **Step 3**: adjusting the matrix obtained in step 2 to eliminate double counting resulting from ‘pass-through FDI’ (also called ‘conduit’ FDI);
- **Step 4**: applying an estimated rate of return on FDI in each matrix cell, taking into account how the average rate of return to FDI deviates from the global average across both investing and receiving jurisdictions.

Step 1: Building a full matrix of FDI positions by *immediate* investor jurisdiction

568. The bilateral FDI matrix by immediate investor jurisdiction is first filled with the existing bilateral FDI statistics from the OECD and the IMF, with the following order of preference: (i) OECD inward FDI statistics, (ii) OECD outward FDI statistics, (iii) IMF CDIS inward FDI, (iv) IMF CDIS outward FDI.

569. Even after combining these sources, there remain data gaps in the FDI matrix. These missing values are extrapolated using a standard gravity model. The gravity equation is estimated using a Poisson pseudo-maximum-likelihood estimation (Santos Silva and Tenreyro, 2006[20]) of the level of bilateral FDI on distance and both jurisdictions’ GDP and GDP per capita, as well as the statutory CIT rate of the recipient jurisdiction. In the regression, all the independent variables have been transformed in logs. The results are reported in Annex Table 5.C.1. The results from the third column are the ones used for the extrapolation.

Annex Table 5.C.1. Estimated gravity equation used to extrapolate bilateral FDI positions

	Dependent variable: Bilateral FDI position (in levels)		
GDP of investor (log)	0.380***	0.551***	0.551***
	(4.03)	(40.86)	(41.25)
GDP of recipient (log)	0.370***	0.506***	0.525***
	(6.71)	(34.06)	(29.37)
Distance (log)	-1.121***	-0.770***	-0.766***
	(-15.46)	(-33.03)	(-33.12)
GDP per capita of investor (log)		0.725***	0.725***
		(25.56)	(25.75)
GDP per capita of recipient (log)		0.164***	0.148***
		(7.74)	(6.47)
Statutory CIT rate of recipient			-0.963**
			(-2.50)
Constant	10.68***	-2.934***	-2.685***
	(9.03)	(-6.78)	(-6.30)
N	20151	19742	19602

Note: T-statistics in parentheses. ***, **, *: denote significance at 1, 5, and 10% levels respectively.
Source: OECD Secretariat.

570. The amount of FDI derived from each of the sources in the FDI matrix by immediate investor is presented in Annex Table 5.C.2. For example, 61% of the FDI matrix was filled with OECD inward FDI data, while only 2% was filled with the gravity model. Annex Table 5.C.3 presents an aggregated version of the FDI matrix by immediate investor. The global level of FDI is found to be slightly above USD 35.3 trillion, which is comparable to the 2016 numbers from Damgaard et al. (2019[10]).[30]

Annex Table 5.C.2. Sources of bilateral FDI data by immediate investor

Source of immediate FDI data	Total FDI (USD billion)	Share
OECD inward	21654	61%
OECD outward	6659	19%
IMF inward	5137	15%
IMF outward	1240	4%
Extrapolation (gravity model)	638	2%
Total	**35329**	**100%**

Source: OECD Secretariat.

Annex Table 5.C.3. FDI by *immediate* investor matrix, aggregated by broad jurisdiction groups

		Jurisdiction of investor				
	(USD billion of 2016)	**High income**	**Middle income**	**Low income**	**Investment Hubs**	**Total**
Jurisdiction of recipient	High income (64 jurisd.)	7218	512	4	5231	**12965**
	Middle income (105)	2320	381	5	3487	**6193**
	Low income (29)	32	18	1	20	**72**
	Investment Hubs (24)	7241	1109	6	7743	**16099**
	Total	**16811**	**2021**	**16**	**16480**	**35329**

Note: Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. The number of jurisdictions in each group is indicated in parentheses. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat calculations, based on OECD and IMF FDI statistics, complemented with extrapolations based on a gravity model.

Step 2: Building a full matrix of FDI positions by *ultimate* investor jurisdictions

571. For the purpose of filling the profit matrix, it is preferable to use FDI positions by *ultimate* rather than *immediate* investor, i.e. based on the jurisdiction of the ultimate parent of the MNE investing in a jurisdiction (Annex Figure 5.C.1). This is because the profit matrix is structured by jurisdiction of ultimate parent. For example, each column in the profit matrix corresponds to MNEs from a given jurisdiction of ultimate parent. Other sources used to fill the profit matrix (e.g. CbCR data, ORBIS) are consistent with this 'ultimate investor' approach.

Annex Figure 5.C.1. Stylised example on FDI by ultimate versus immediate investor

Note: In this stylised example, a MNE with an ultimate parent in jurisdiction A has invested 100 into jurisdiction C, passing through jurisdiction B. FDI statistics by immediate investor report a FDI position from jurisdiction A into jurisdiction B, and a similar position from jurisdiction B into jurisdiction C. FDI statistics by ultimate investor in jurisdiction C report that the ultimate investor is jurisdiction A.
Source*:* OECD Secretariat.

572. While FDI data are traditionally reported by "immediate investing country", the OECD recently started publishing inward FDI statistics by ultimate investor for a subset of 15 receiving jurisdictions,[31] whose combined inward FDI represents 23% of global FDI.

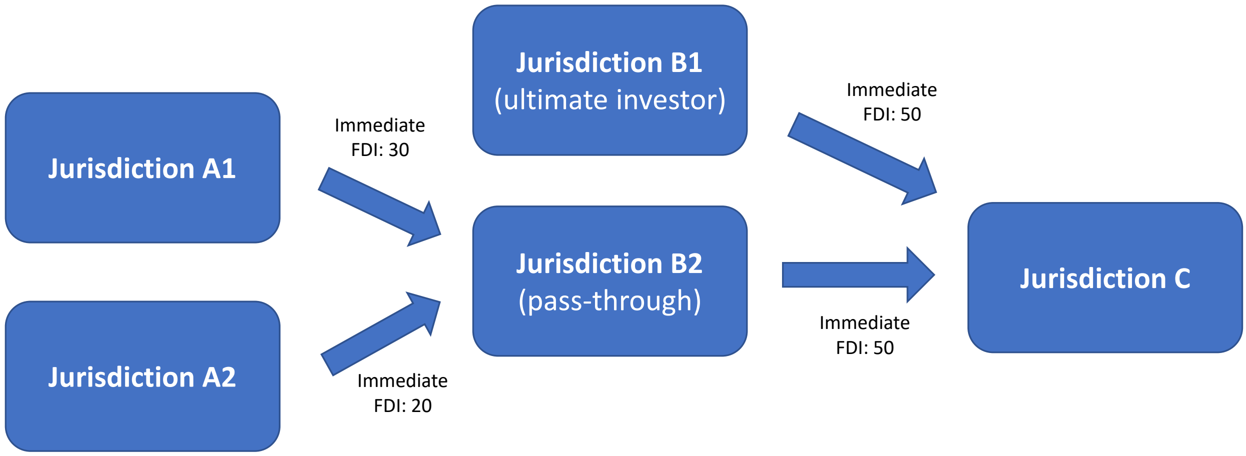
573. There are important differences between inward FDI data by immediate versus ultimate investor, which reflect that some FDI may pass through a jurisdiction before reaching its final destination.[32,33] For instance, in 2018, OECD FDI data indicate that the United States accounted for 3.2% of immediate investors in Iceland (USD 0.3 billion out of USD 9.5 billion) but 28.6% of the ultimate investors (USD 2.7 billion out of the same USD 9.5 billion), reflecting that most investment from the United States to Iceland were channelled through third-party jurisdictions.

574. To build a full matrix of FDI positions by ultimate investor jurisdiction, the approach in this chapter is to use the available OECD data for the jurisdictions where it is available, and to complement it with extrapolations using data on FDI positions by immediate investor, applying the methodology developed by Casella (2019[5]).

575. Casella (2019[5]) uses a probabilistic method in order to obtain the distribution of ultimate investors in a given jurisdiction based on data on immediate investors. The intuition is the following. The first step is to assess if an FDI position is "pass-through" or not (or, more precisely, the probability that it is "pass-through"). If a position is not "pass-through", the immediate investor is considered to be the ultimate investor. If it is "pass-through", the second step is to go up in the investing chain, based on FDI data by immediate investor, to try to find the ultimate investor. This procedure is repeated until the ultimate investor is found. Identifying the ultimate investor may require going up several steps in the investing chain if the investment has been channelled successively through several jurisdictions. In practice, this procedure is implemented by using absorbing Markov chains (see Casella (2019[5]) for more details).

576. For example, in Annex Figure 5.C.2, jurisdictions B1 and B2 are immediate investors into jurisdiction C. Jurisdiction B2 is identified as a pass-through jurisdiction (based on assumptions described below), while jurisdiction B1 is not. Therefore, jurisdiction B1 is considered to be the ultimate investor into jurisdiction C for the investment observed from B1 to C in the statistics by immediate investors. In the case of jurisdiction B2, one needs to go one step up in the investing chain to identify the ultimate investor(s), which are here A1 and A2. Ultimate investors into C are finally found to be B1 (investment of 50), A1 (30) and A2 (20). If A1 had itself been a pass-through jurisdiction, one would have needed to look at the jurisdictions investing into A1 to identify the ultimate parent.

Annex Figure 5.C.2. Stylised example on the methodology to identify ultimate investors



Note: In this stylised example, both jurisdictions B1 and B2 are immediate investors in jurisdiction C, and both jurisdictions A1 and A2 are immediate investors in jurisdiction B1. However, the immediate investment observed from B2 to C is identified as being a "pass-through" investment corresponding to the immediate investments made by A1 and A2 into B. The "ultimate investors" in C are thus B1, A1 and A2 and the respective levels of ultimate investments are 50, 30 and 20.
Source*:* OECD Secretariat.

577. In this simplified example, B1 is fully an ultimate investor jurisdiction and B2 is fully a pass-through jurisdiction. In practice however, many jurisdictions are simultaneously ultimate investors (for certain positions) and pass-through (for other positions). The methodology therefore relies on the *probability* that an FDI position from a jurisdiction is pass-through. For example, if the probability that positions from jurisdiction B are pass-through is 40%, 60% of positions from B are considered as having B as ultimate investor, and one looks one step up in the investing chain to identify the ultimate investor for the remaining 40%.

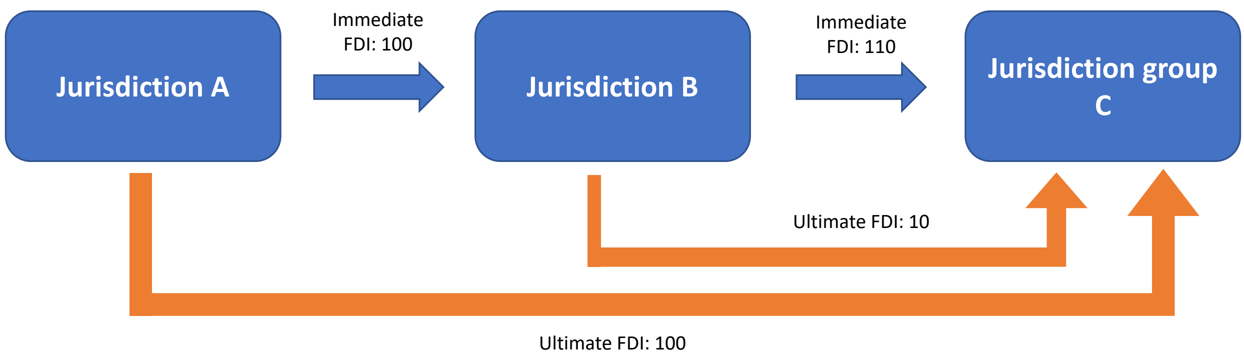
578. A key input to the procedure is therefore the probability that a FDI position is "pass-through". Several approaches have been proposed to assess this probability. In particular, Bolwijn et al. (2018[11]), Casella (2019[5]), Damgaard et al. (2019[10]) suggest using the share of Special Purpose Entities (SPEs) in a jurisdiction's outward FDI investment.[34] However, an issue with this approach is that only around 30 jurisdictions report FDI statistics for SPEs separately, and that extrapolating the share of SPEs beyond these jurisdictions entails important uncertainties.

579. As an alternative, the present analysis assesses the probability that investment from a given jurisdiction is pass-through based on the available data on inward FDI by ultimate versus immediate investor, across the subset of 15 recipient jurisdictions where it is available. The intuition is that if a jurisdiction is often an immediate investor without being an ultimate investor, an important share of its outward FDI is likely to be pass-through. Reflecting this, it is assumed that the share of investment from a jurisdiction that is *not* pass-through corresponds to the ratio between this jurisdiction's outward FDI as an *immediate* investor and its outward FDI as an *ultimate* investor.[35] To ensure that the numerator and denominator of the ratio are comparable, only FDI into the subset jurisdictions reporting FDI statistics by ultimate investor is included. Finally, if the ratio exceeds one, it is assumed that the jurisdiction is never pass-through.

580. For example, in Annex Figure 5.C.3 below, jurisdiction group C is assumed to be the set of jurisdictions reporting FDI data by ultimate investor. Outward FDI of jurisdiction B as an *immediate* investor into C is 110, out of which jurisdiction B is the ultimate investor for 10, and a pass-through jurisdiction for investment from A for 100. The probability that jurisdiction B is *not* "pass-through" is estimated to be 10/110 = 9%, and therefore the probability that it is pass-through is 91%. Hence, for 9% of observed FDI positions

from jurisdiction B, jurisdiction B will be deemed to be the ultimate investor. For the other 91%, the procedure will assume that jurisdiction B is not the ultimate investor, and go up one step in the investment chain to search for an ultimate investor, based on the data on immediate investors into jurisdiction B.

Annex Figure 5.C.3. Stylised example on probability to be "pass-through"



Note: In this stylised example, jurisdiction group C is the group of jurisdictions reporting FDI data by ultimate investor. Based on data from these jurisdictions, it appears that jurisdiction B is often an immediate investor without being an ultimate investor. This happens for 100/110=91% of jurisdiction B's immediate investments into C. As a result, it is assumed that jurisdiction B is a pass-through jurisdiction 91% of the time.
Source: OECD Secretariat.

581. The resulting FDI matrix by ultimate investor jurisdiction based on the methodology described above is presented for aggregate jurisdiction groups in Annex Table 5.C.4. The main difference compared to Annex Table 5.C.3 reflects the fact that investment hubs are more often intermediate than ultimate investors. The aggregate level of FDI is slightly lower than in Annex Table 5.C.3, which is mainly due to the elimination of 'self FDI', where a jurisdiction is the ultimate investor of foreign direct investment into itself (while the immediate investor is in another jurisdiction).[36] Annex Table 5.C.4 excludes this 'self-investment'. However, the difference in global FDI between Annex Table 5.C.4 and Annex Table 5.C.3 is relatively small given the scale of 'pass-through FDI'. This is because pass-through FDI has not been fully dealt with at this stage, which is why Step 3 is required.

Annex Table 5.C.4. FDI by *ultimate* investor matrix, aggregated by broad jurisdiction groups

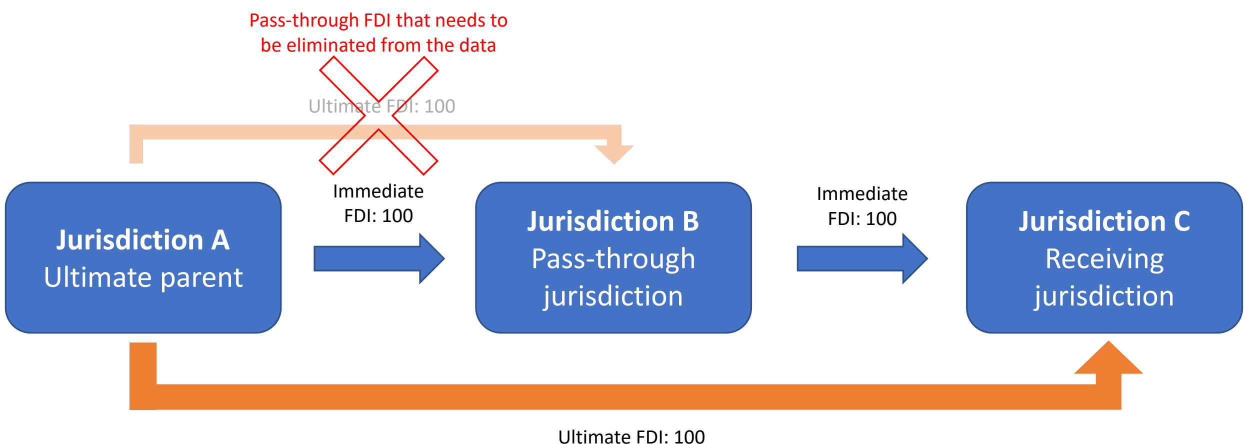
		Jurisdiction of ultimate investor				
	(USD billion of 2016)	**High income**	**Middle income**	**Low income**	**Investment Hubs**	**Total**
Jurisdiction of recipient	High income (64 jurisd.)	8281	655	1	3114	**12051**
	Middle income (105)	2841	462	3	2863	**6169**
	Low income (29)	40	18	1	13	**72**
	Investment Hubs (24)	9137	1335	2	5159	**15632**
	Total	**20299**	**2470**	**7**	**11148**	**33924**

Note: Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. The number of jurisdictions in each group is indicated in parentheses. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat calculations, based on OECD FDI statistics and estimates based on the FDI matrix by immediate investor (Annex Table 5.C.3) and a methodology following Casella (2019[5]).

Step 3: Eliminating double counting from “pass-through FDI”

582. Considering FDI data by jurisdiction of ultimate investor as done in Step 2 addresses some issues related to “pass-through FDI” (also called “conduit FDI”) compared to FDI data by jurisdiction of immediate investor, but not all of them. This is illustrated by the example in Annex Figure 5.C.4. In this example, the ultimate investor into jurisdiction C is correctly identified to be jurisdiction A. However, when considering the ultimate investors into jurisdiction B, the FDI position of jurisdiction A into jurisdiction B is still considered as an ultimate investment, while it should be eliminated as it is “pass-through FDI” for which jurisdiction B is not the final destination.

Annex Figure 5.C.4. Stylised example on “pass-through FDI”



Note: This stylised example shows that FDI data by ultimate investor jurisdiction contain redundant FDI positions resulting from pass-through FDI that needs to be eliminated to avoid double counting. In this example, the ultimate investor into jurisdiction C is correctly identified to be jurisdiction A. However, when considering the ultimate investors into jurisdiction B, the FDI position of jurisdiction A into jurisdiction B is still considered as an ultimate investment, while it is only a “pass-through FDI” that should be eliminated from the data.
Source: OECD Secretariat.

583. The underlying reason for this issue is that FDI data by ultimate investor does not consider whether a given FDI position stops in jurisdiction B or is pass-through to another jurisdiction (in this case, jurisdiction C).

584. Another way to see this double counting issue is that the global total of FDI positions by jurisdiction of *immediate* investor is only slightly lower than the global total of FDI positions by jurisdiction of *ultimate* investor (see Annex Table 5.C.3 and Annex Table 5.C.4), while global FDI positions by jurisdiction of *ultimate* investor should be substantially lower if pass-through FDI were fully dealt with (see also Damgaard et al. (2019[10])).

585. To address the issue, the approach in this chapter is to adjust downwards the FDI positions into a jurisdiction in proportion to its probability to be pass-through. For example, if a jurisdiction is deemed to be pass-through for 91% of FDI positions into it (as was the case of jurisdiction B in the example of Annex Figure 5.C.3 above), FDI positions into jurisdiction B are reduced by 91%, reflecting that jurisdiction B is the ultimate *destination* of the FDI in only 9% of the cases.

586. The resulting FDI matrix is presented in Annex Table 5.C.5. As expected, total global FDI is reduced, from USD 35.3 trillion in the immediate investor matrix, to USD 24.2 trillion in Annex Table 5.C.5. This represents a 32% reduction, which can be interpreted as suggesting that 32% of global FDI is pass-through. This is broadly consistent with estimates by Damgaard and Elkjaer (2017[4]) and Damgaard et al.

(2019[10]), who both use data on FDI towards Special Purpose Entities to identify pass-through FDI. Damgaard and Elkjaer (2017[4]) assess that 34% of FDI is pass-through, and Damgaard et al. (2019[10]) assess that the share of "phantom" (i.e. pass-through) FDI was about 30% in 2009 and increased to almost 40% in 2017.

Annex Table 5.C.5. FDI by *ultimate* investor matrix, after adjustment for "pass-through FDI", aggregated by broad jurisdiction groups

		Jurisdiction of ultimate investor				
	(USD billion of 2016)	**High income**	**Middle income**	**Low income**	**Investment Hubs**	**Total**
Jurisdiction of recipient	High income (64 jurisd.)	7788	591	1	2830	**11211**
	Middle income (105)	2560	368	2	2756	**5685**
	Low income (29)	11	7	0	2	**21**
	Investment Hubs (24)	4252	770	1	2229	**7252**
	Total	**14611**	**1736**	**4**	**7817**	**24169**

Note: Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. The number of jurisdictions in each group is indicated in parentheses. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat.

587. An interesting way to assess the quality of the methodology to identify ultimate investors is to compare the predicted distribution of ultimate investors to their actual distribution in the jurisdictions where this distribution is observed in the data, i.e. in the 14 jurisdictions reporting FDI data by ultimate investor and a global aggregate inward FDI position.[37] The predicted values in these 14 jurisdictions are not used to build the matrix in Annex Table 5.C.5 given that actual data is available (and therefore used) in these jurisdictions, but comparing these predicted values to actual data is a way to test the performance of the methodology to identify ultimate investors.

588. This comparison is done by computing the "distance" between the predicted and the actual distribution of ultimate investors into a jurisdiction. This distance is measured, for each recipient jurisdiction, by the sum of the absolute deviations in the shares of each ultimate investor in the total FDI into this recipient jurisdiction considered (i.e. the so-called "L1 norm") as done by Casella (2019[5]). The results are presented in Annex Table 5.C.6 (column 2). These distances are positive, reflecting that the predicted distributions differ from the actual distributions, but they are substantially lower than the distances between the distribution of immediate investors and the distribution of ultimate investors (column 1). This suggests that the methodology clearly improves the identification of ultimate investors compared to a methodology that would simply assume that the distribution of ultimate investors is the same as the distribution of immediate investors. The median improvement across the 14 jurisdictions in Annex Table 5.C.6 is 34%.

Annex Table 5.C.6. Distance between predicted and actual distribution of ultimate investors

Recipient jurisdiction	(1) Distance between distribution of immediate and ultimate investors	(2) Distance between actual and predicted distribution of ultimate investors	% difference between (1) and (2)
Canada	0.46	0.45	-3%
Czech Republic	0.31	0.21	-32%
Estonia	0.16	0.16	-2%
Finland	0.25	0.23	-8%
France	0.25	0.14	-46%
Germany	0.32	0.21	-35%
Hungary	0.75	0.48	-36%
Iceland	0.99	0.55	-45%
Italy	0.31	0.24	-23%
Lithuania	0.22	0.19	-12%
Poland	0.28	0.18	-38%
Switzerland	0.46	0.24	-48%
Turkey	0.09	0.18	96%
United States	0.21	0.12	-41%
Median difference			-34%

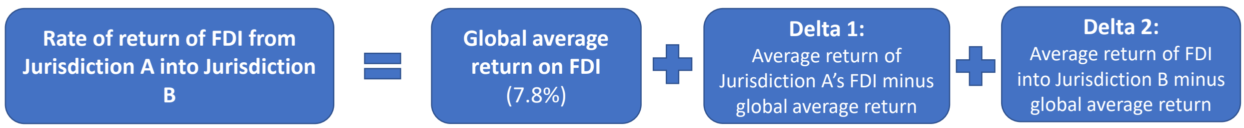
Note: This table presents, for the 14 recipient jurisdictions reporting FDI data by jurisdiction of ultimate investor and a global inward position in the OECD data, the "distance" between the distribution of ultimate FDI investors into the recipient jurisdiction considered with (i) the distribution of immediate FDI investors into the recipient jurisdiction considered (column 1), and (ii) the predicted distribution of ultimate FDI investors into the recipient jurisdiction considered, as computed with the methodology described in this annex (column 2). Distance is measured as the sum of the absolute deviations in the shares of each investor in the total FDI of the recipient jurisdiction considered (i.e. the so-called "L1 norm") similar to Casella (2019[5]). Recipient jurisdictions do not always report the jurisdiction of the ultimate investor for 100% of their inward FDI (e.g. due to confidentiality concerns) and the sum is computed only over the jurisdictions of ultimate investor covered on the data. As a result, the *level* of the distances are not necessarily comparable across jurisdictions, but the relative differences between columns 1 and 2 are comparable.
Source: OECD Secretariat.

Step 4: Computing jurisdiction-specific rates of return to FDI

589. The final step to obtain a measure of profit that can be used to fill cells in the profit matrix is to apply a rate of return to the bilateral FDI positions obtained in Step 3. The rate of return on FDI can vary across both investing and receiving jurisdictions for a range of reasons, such as the sectoral composition of investment, the riskiness of investments and tax planning schemes (e.g. returns on FDI generated by profit shifting strategies can differ from returns on 'real' investments). For instance, average rates of return have been relatively low in low income jurisdictions over the last five years compared to other jurisdiction groups (UNCTAD, 2018[25]).

590. The approach in this chapter takes into account that average rates of return can differ from the global average both at the investor and receiving jurisdiction level. The method is first to compute a global average rate of return to FDI, and then to apply to this average (i) a (positive or negative) delta corresponding to the difference between the average rate of return on FDI of the *investing* jurisdiction and the global average (Delta 1), and (ii) a (positive or negative) delta corresponding to the difference between the average rate of return on FDI in the *receiving* jurisdiction and the global average (Delta 2) (Annex Figure 5.C.5).

Annex Figure 5.C.5. Assumption on the rate of return to FDI



Source: OECD Secretariat.

591. For example, in the case of FDI from a jurisdiction A into a jurisdiction B, if global FDI from jurisdiction A has an average return that is 3 percentage points above the global average, but that global FDI into jurisdiction B has an average return that is 1 percentage point below the global average, the rate of return on FDI from jurisdiction A into jurisdiction B is assumed to be equal to the global average plus 3-1=2 percentage points. To avoid potential noise from extreme observations, both deltas are bounded at +/- 5 percentage points, and the sum of the two deltas as well.

592. For consistency with the other data sources used in the profit matrix (CbCR and ORBIS data), the global average rate of return on FDI is computed based on the cells of the profit matrix that are already filled with these sources. In each of these cells, profit (measured with CbCR or ORBIS) is divided by the bilateral FDI position obtained in Step 3. After excluding cells with outlying rates of return from the sample (based on the Cook's distance of each observation in the regression of profit on FDI), a "standard" rate of return is computed by taking the median ratio of bilateral profit to bilateral position across remaining matrix cells. This results in a standard rate of return of 7.8%.

593. Return differentials at the level of both investing and receiving jurisdictions (Delta 1 and Delta 2) are computed based on statistics on FDI positions and FDI income in the OECD FDI statistics.[38] For each investing and each receiving jurisdiction, a differential is computed for each year between 2013 and 2016 by comparing the rate of return of the jurisdiction to the global rate of return. The jurisdiction-specific differential is the median of the differential over the four years to reduce volatility, capped between -5 and +5 percentage points to avoid generating extreme values, as mentioned above.

Annex 5.D. Matrices aggregated by broad income groups and regions

Annex Table 5.D.1. Matrices aggregated by broad income groups and regions

Panel A: The profit matrix

In USD billion	A. Americas - High income	B. Europe & Central Asia - High income	C. East Asia & Pacific - High income	D. Middle East & North Africa - High income	E. Latin America & Caribb. - Middle and low income	F. Europe & Central Asia - Middle and low income	G. East Asia & Pacific - Middle and low income	H. Middle East & North Africa - Middle and low income	I. South Asia - Middle and low income	J. Sub-Saharan - High and middle income	K. Sub-Saharan - Low income	L. Americas invest. hubs	M. Europe invest. hubs	N. Other invest. hubs	Total
A. Americas - High income	1527	126	53	4	12	1	5	0	3	1	0	5	52	2	**1791**
B. Europe & Central Asia - High income	158	884	34	5	2	3	4	1	3	1	0	11	74	3	**1184**
C. East Asia & Pacific - High income	63	28	605	2	1	0	3	0	0	0	0	2	10	4	**720**
D. Middle East & North Africa - High income	14	7	2	56	0	1	1	1	1	0	0	0	5	1	**89**
E. Latin America & Caribbean - Middle and low income	49	33	4	0	110	1	1	0	0	0	0	3	18	1	**221**
F. Europe & Central Asia - Middle and low income	10	27	3	2	1	109	1	0	1	0	0	6	42	1	**203**
G. East Asia & Pacific - Middle and low income	52	37	89	1	0	2	472	0	2	1	0	49	11	21	**736**
H. Middle East & North Africa - Middle and low income	5	8	0	2	0	0	0	15	0	0	0	0	1	0	**31**
I. South Asia - Middle and low income	15	8	3	1	0	0	1	0	80	0	0	0	3	3	**114**
J. Sub-Saharan - High and middle income	8	10	1	0	0	0	1	0	0	24	0	2	4	1	**52**
K. Sub-Saharan - Low income	0	0	0	0	0	0	0	0	0	1	2	0	0	0	**4**
L. Americas Investment hubs	115	11	2	0	5	1	0	0	1	0	0	31	19	10	**196**
M. European Investment hubs	265	115	14	4	3	6	2	0	1	1	0	4	136	7	**558**
N. Other Investment hubs	78	28	18	2	5	3	39	0	2	1	0	34	16	56	**281**
Total	**2358**	**1322**	**829**	**78**	**140**	**128**	**529**	**17**	**94**	**30**	**2**	**150**	**391**	**112**	**6181**

Panel B: The turnover matrix

In USD billion	A. Americas - High income	B. Europe & Central Asia - High income	C. East Asia & Pacific - High income	D. Middle East & North Africa - High income	E. Latin America & Caribb. - Middle and low income	F. Europe & Central Asia - Middle and low income	G. East Asia & Pacific - Middle and low income	H. Middle East & North Africa - Middle and low income	I. South Asia - Middle and low income	J. Sub-Saharan - High and middle income	K. Sub-Saharan - Low income	L. Americas invest. hubs	M. Europe invest. hubs	N. Other invest. hubs	Total
A. Americas - High income	13909	1364	1153	45	121	14	50	13	56	17	3	82	826	36	**17686**
B. Europe & Central Asia - High income	1589	9162	530	115	54	93	159	42	63	39	7	111	1236	56	**13254**
C. East Asia & Pacific - High income	521	340	7534	37	16	16	91	12	15	17	2	78	97	42	**8818**
D. Middle East & North Africa - High income	77	68	22	564	8	14	8	14	17	5	1	2	30	7	**840**
E. Latin America & Caribbean - Middle and low income	489	363	125	20	1576	13	31	7	6	8	1	32	126	8	**2804**
F. Europe & Central Asia - Middle and low income	114	315	73	32	38	1316	19	10	6	8	1	68	342	24	**2367**
G. East Asia & Pacific - Middle and low income	544	411	1253	24	15	76	7058	9	18	8	2	808	119	248	**10594**
H. Middle East & North Africa - Middle and low income	32	85	24	32	8	15	9	171	2	4	1	2	30	6	**420**
I. South Asia - Middle and low income	96	78	52	16	6	7	15	4	653	3	1	6	26	9	**970**
J. Sub-Saharan - High and middle income	61	123	31	14	9	7	12	4	4	155	1	6	32	5	**464**
K. Sub-Saharan - Low income	5	18	6	6	3	3	2	2	2	5	32	1	6	2	**92**
L. Americas Investment hubs	184	6	6	1	2	1	1	0	1	0	0	36	4	1	**243**
M. European Investment hubs	1100	747	114	24	19	15	17	11	9	9	2	37	1041	15	**3160**
N. Other Investment hubs	694	210	303	9	7	4	59	2	16	5	0	24	110	220	**1663**
Total	**19415**	**13289**	**11226**	**940**	**1881**	**1593**	**7531**	**302**	**867**	**280**	**55**	**1293**	**4024**	**679**	**63375**

Panel C: The tangible assets matrix

In USD billion	A. Americas - High income	B. Europe & Central Asia - High income	C. East Asia & Pacific - High income	D. Middle East & North Africa - High income	E. Latin America & Caribb. - Middle and low income	F. Europe & Central Asia - Middle and low income	G. East Asia & Pacific - Middle and low income	H. Middle East & North Africa - Middle and low income	I. South Asia - Middle and low income	J. Sub-Saharan - High and middle income	K. Sub-Saharan - Low income	L. Americas invest. hubs	M. Europe invest. hubs	N. Other invest. hubs	Total
A. Americas - High income	4567	360	290	19	46	7	20	5	19	17	1	39	185	19	**5594**
B. Europe & Central Asia - High income	348	2674	97	29	15	23	30	11	16	9	2	35	249	22	**3559**
C. East Asia & Pacific - High income	162	80	2443	13	6	7	29	4	8	7	1	17	17	17	**2810**
D. Middle East & North Africa - High income	32	26	10	312	4	7	4	7	13	2	1	1	10	3	**433**
E. Latin America & Caribbean - Middle and low income	143	130	39	10	703	7	12	3	2	3	1	238	42	4	**1335**
F. Europe & Central Asia - Middle and low income	39	93	17	12	13	695	7	4	4	2	0	23	97	6	**1013**
G. East Asia & Pacific - Middle and low income	123	90	255	8	5	42	1900	3	4	2	0	197	23	65	**2717**
H. Middle East & North Africa - Middle and low income	16	50	11	19	5	9	5	108	1	2	1	1	15	4	**245**
I. South Asia - Middle and low income	25	51	41	12	4	5	12	3	672	2	0	1	17	5	**851**
J. Sub-Saharan - High and middle income	45	73	17	10	6	5	7	3	2	110	1	1	16	4	**300**
K. Sub-Saharan - Low income	2	6	2	3	1	1	1	1	1	2	10	0	2	1	**34**
L. Americas Investment hubs	28	2	2	1	1	0	0	0	1	0	0	18	1	2	**57**
M. European Investment hubs	99	148	34	7	12	4	4	3	2	4	1	10	221	4	**554**
N. Other Investment hubs	50	35	29	3	2	2	16	1	15	2	0	26	12	126	**319**
Total	**5679**	**3819**	**3286**	**457**	**823**	**814**	**2049**	**156**	**760**	**165**	**18**	**608**	**908**	**282**	**19823**

Panel D: The payroll matrix

In USD billion	A. Americas - High income	B. Europe & Central Asia - High income	C. East Asia & Pacific - High income	D. Middle East & North Africa - High income	E. Latin America & Caribb. - Middle and low income	F. Europe & Central Asia - Middle and low income	G. East Asia & Pacific - Middle and low income	H. Middle East & North Africa - Middle and low income	I. South Asia - Middle and low income	J. Sub-Saharan - High and middle income	K. Sub-Saharan - Low income	L. Americas invest. hubs	M. Europe invest. hubs	N. Other invest. hubs	Total
A. Americas - High income	3704	219	109	5	16	2	8	2	11	2	0	7	136	4	**4226**
B. Europe & Central Asia - High income	279	1653	58	11	9	20	15	9	18	9	1	22	257	11	**2373**
C. East Asia & Pacific - High income	60	46	710	4	2	2	12	1	1	4	0	7	15	5	**870**
D. Middle East & North Africa - High income	14	12	3	81	1	2	1	2	2	1	0	0	6	1	**127**
E. Latin America & Caribbean - Middle and low income	55	54	11	2	198	2	5	1	1	1	0	3	21	1	**354**
F. Europe & Central Asia - Middle and low income	12	45	7	4	5	136	3	1	1	1	0	4	44	3	**265**
G. East Asia & Pacific - Middle and low income	48	55	100	2	2	7	938	1	2	1	0	51	17	22	**1246**
H. Middle East & North Africa - Middle and low income	4	13	3	4	1	2	1	23	0	0	0	0	5	1	**58**
I. South Asia - Middle and low income	22	17	10	3	1	1	3	1	137	0	0	1	6	2	**203**
J. Sub-Saharan - High and middle income	7	18	3	2	1	1	2	0	0	20	0	0	5	1	**61**
K. Sub-Saharan - Low income	1	2	1	1	0	0	0	0	0	1	5	0	1	0	**13**
L. Americas Investment hubs	4	1	1	0	0	0	0	0	0	0	0	2	0	0	**8**
M. European Investment hubs	52	83	9	2	2	2	2	1	1	1	0	2	137	1	**296**
N. Other Investment hubs	33	22	19	1	1	0	7	0	1	0	0	1	12	13	**111**
Total	**4294**	**2240**	**1041**	**122**	**239**	**178**	**997**	**43**	**175**	**42**	**8**	**102**	**664**	**64**	**10211**

Note: The composition of the jurisdiction groups is presented in Annex Table 5.D.2.
Source: OECD Secretariat

Annex Table 5.D.2. Jurisdiction groups in the aggregated matrices

Panel A: Jurisdiction groups A to G

A. Americas - High income	B. Europe & Central Asia - High income	C. East Asia & Pacific - High income	D. Middle East & North Africa - High income	E. Latin Am. & Caribbean - Middle and low income	F. Europe & Central Asia - Middle and low income	G. East Asia & Pacific - Middle and low income
Antigua and Barbuda	Andorra	Australia	Bahrain	Argentina	Albania	American Samoa
Aruba	Austria	Brunei Darussalam	Israel	Belize	Armenia	Cambodia
Bonaire	Belgium	Cook Islands	Kuwait	Bolivia	Azerbaijan	China (People's Republic of)
Canada	Croatia	French Polynesia	Oman	Brazil	Belarus	DPRK
Chile	Czech Republic	Guam	Qatar	Colombia	Bosnia and Herzegovina	Fiji
Curaçao	Denmark	Japan	Saudi Arabia	Costa Rica	Bulgaria	Indonesia
Montserrat	Estonia	Korea	United Arab Emirates	Cuba	Georgia	Kiribati
Panama	Faroe Islands	Macau (China)		Dominica	Kazakhstan	Lao PDR
Puerto Rico	Finland	New Caledonia		Dominican Republic	Kosovo	Malaysia
Saint Kitts and Nevis	France	New Zealand		Ecuador	Kyrgyzstan	Micronesia
Sint Maarten	Germany	Northern Mariana Islands		El Salvador	Moldova	Mongolia
Trinidad and Tobago	Greece	Palau		Grenada	Montenegro	Myanmar
United States	Greenland	Chinese Taipei		Guatemala	North Macedonia	Nauru
United States Virgin Islands	Iceland			Guyana	Romania	Papua New Guinea
Uruguay	Italy			Haiti	Russia	Philippines
	Latvia			Honduras	Serbia	Samoa
	Liechtenstein			Jamaica	Tajikistan	Solomon Islands
	Lithuania			Mexico	Turkey	Thailand
	Monaco			Nicaragua	Turkmenistan	Timor-Leste
	Norway			Paraguay	Ukraine	Tonga
	Poland			Peru	Uzbekistan	Tuvalu
	Portugal			Saint Lucia		Vanuatu
	San Marino			Saint Vincent and the Grenadines		Viet Nam
	Slovak Republic			Suriname		
	Slovenia			Venezuela		
	Spain					
	Sweden					
	United Kingdom					

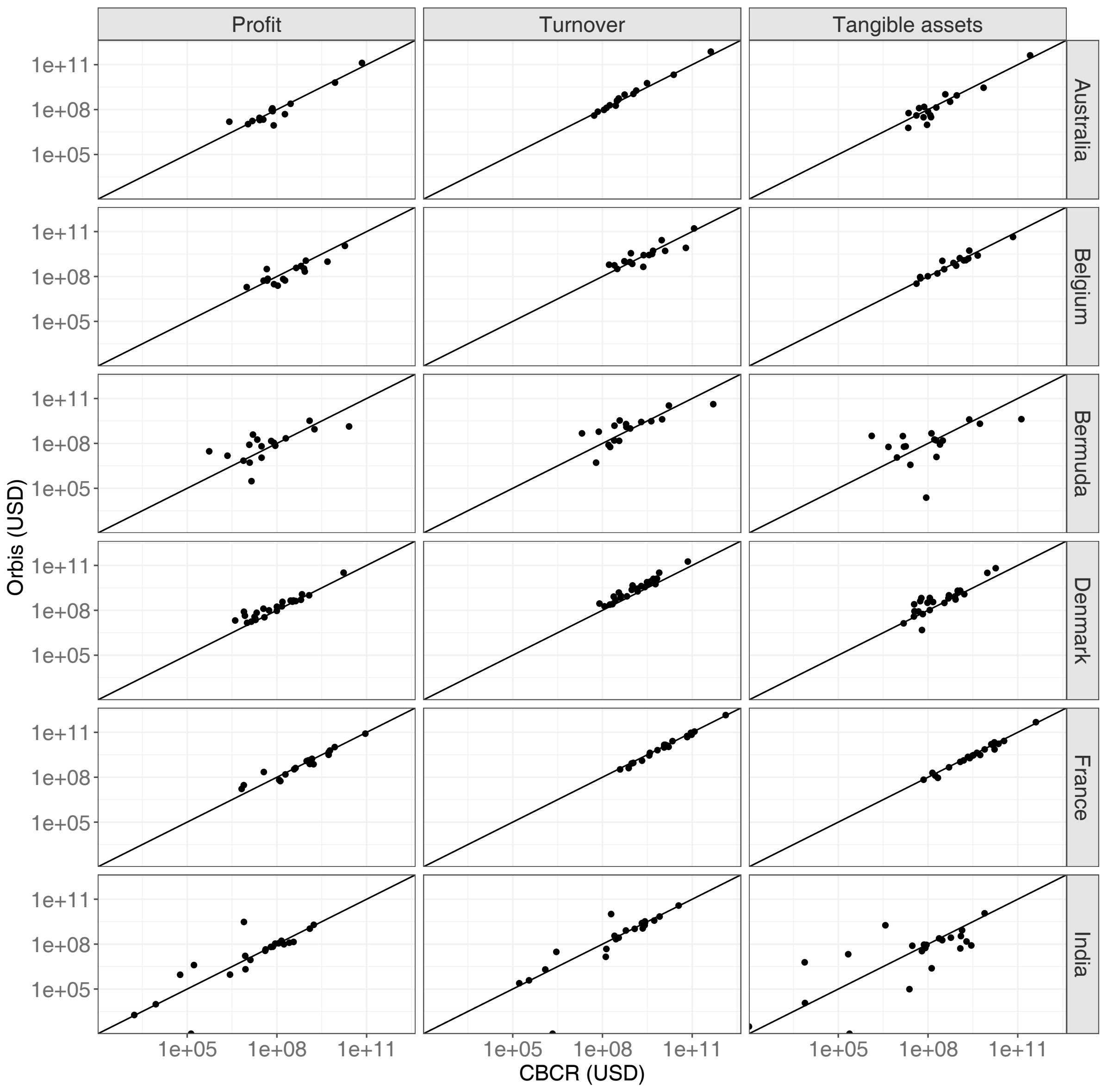
Panel B: Jurisdiction groups H to N

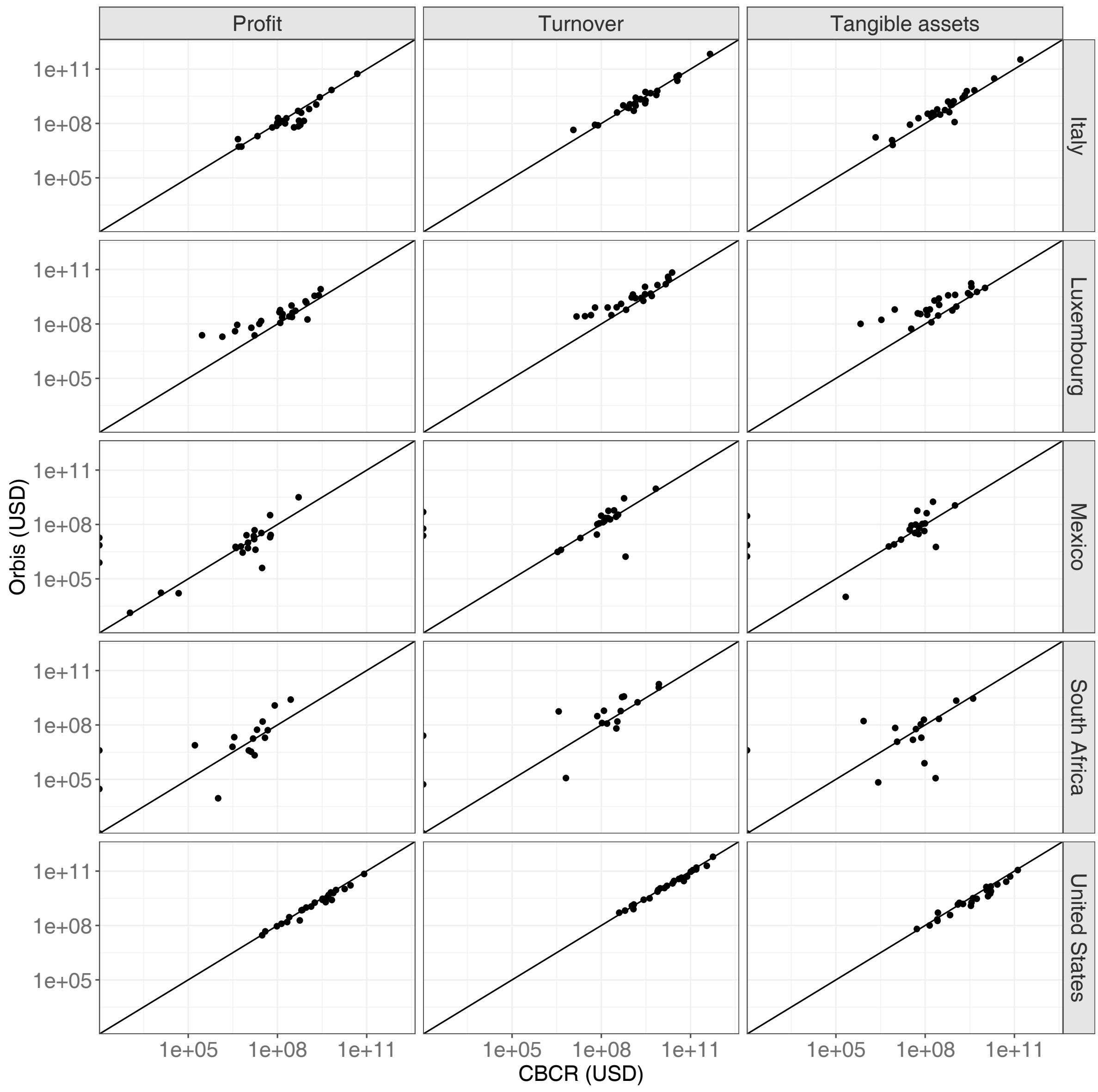
H. Middle East & North Africa - Middle and low income	I. South Asia - Middle and low income	J. Sub-Saharan - High and middle income	K. Sub-Saharan - Low income	L. Americas investment hubs	M. European investment hubs	N. Other investment hubs
Algeria	Afghanistan	Angola	Benin	Anguilla	Bailiwick of Guernsey	Hong Kong (China)
Djibouti	Bangladesh	Botswana	Burkina Faso	Bahamas	Cyprus	Liberia
Egypt	Bhutan	Cabo Verde	Burundi	Barbados	Gibraltar	Malta
Iran	India	Cameroon	Central African Republic	Bermuda	Hungary	Marshall Islands
Iraq	Maldives	Comoros	Chad	British Virgin Islands	Ireland	Mauritius
Jordan	Nepal	Congo	DRC	Cayman Islands	Isle of Man	Mozambique
Lebanon	Pakistan	Côte d'Ivoire	Eritrea	Turks and Caicos Islands	Jersey	Singapore
Libya	Sri Lanka	Equatorial Guinea	Ethiopia		Luxembourg	
Morocco		Eswatini	Gambia		Netherlands	
Palestinian Authority		Gabon	Guinea		Switzerland	
Syria		Ghana	Guinea-Bissau			
Tunisia		Kenya	Madagascar			
Yemen		Lesotho	Malawi			
		Mauritania	Mali			
		Namibia	Niger			
		Nigeria	Rwanda			
		Sao Tome and Principe	Sierra Leone			
		Senegal	Somalia			
		Seychelles	South Sudan			
		South Africa	Tanzania			
		Sudan	Togo			
		Zambia	Uganda			
		Zimbabwe				

Note: The groups are based on World Bank classifications of jurisdictions by income levels and geographic regions. Certain categories are grouped to ensure a sufficient number of jurisdictions in each group in order to preserve confidentiality of the jurisdiction-specific data. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat

Annex 5.E. Additional figures on benchmarking across data sources

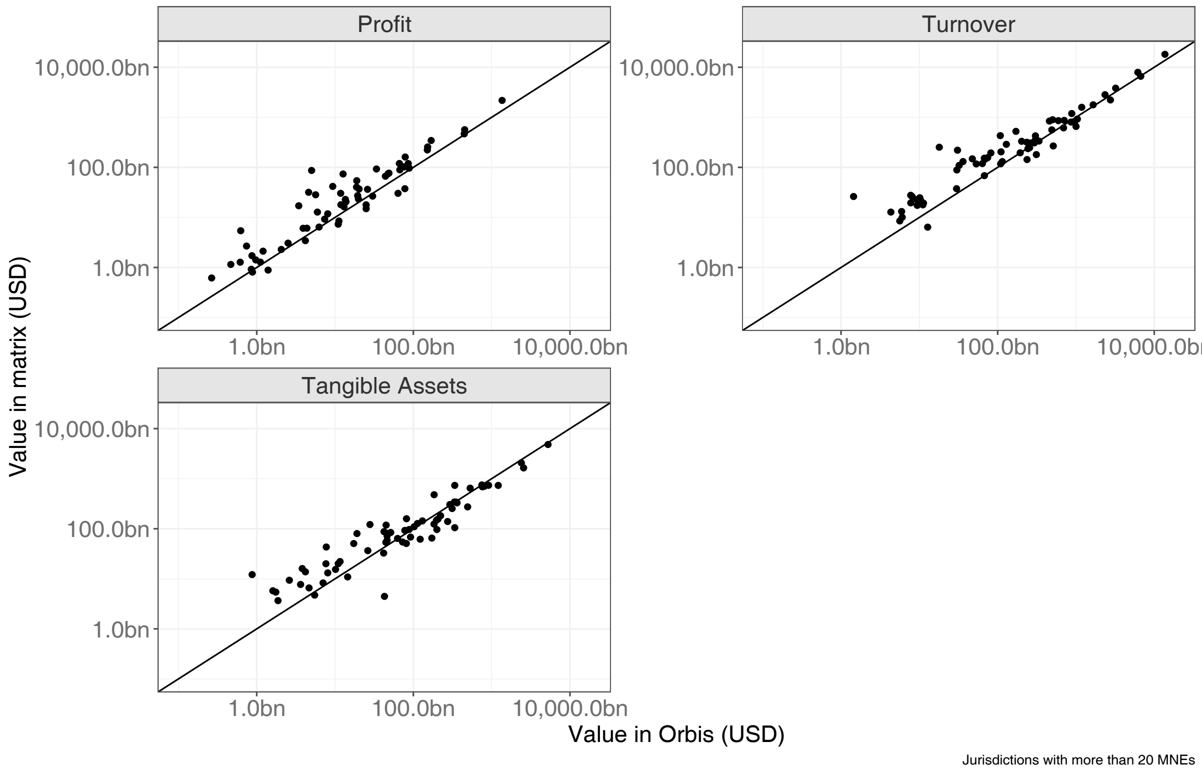
Annex Figure 5.E.1. Comparison between CbCR data and ORBIS unconsolidated account data, by jurisdiction of ultimate parent





Note: These figures compare data on profit, turnover, and tangible assets, of MNEs with an ultimate parent in a given jurisdiction across a range of affiliate jurisdictions, from two sources: CbCR data and data from ORBIS (unconsolidated financial accounts). For comparability, both data sources focus on 'sub-groups with positive profits' in 2016, except for the US data which is from the 2017 CbCR. Each dot corresponds to one jurisdiction of affiliate. The comparison is restricted to jurisdictions of affiliate with sufficiently good coverage of unconsolidated accounts in ORBIS (see list in Annex 5.A), and to ultimate parent jurisdictions that reported more than ten bilateral entries in their CbCR data (i.e. a subset of the list in Annex 5.A). The line corresponds to the 45-degree line.
Source: OECD Secretariat calculations based on ORBIS and CbCR data.

Annex Figure 5.E.2. Comparison between the column totals in the matrices and consolidated financial account data



Note: These figures compare data on total profit, turnover, and tangible assets of MNEs by jurisdiction of ultimate parent (each dot corresponds to a jurisdiction of ultimate parent), from two sources: totals by column (i.e. by jurisdiction of ultimate parent) in the profit, turnover, and tangible assets matrices, and consolidated financial account data from ORBIS combined with other sources such as Worldscope (total of consolidated financial accounts by jurisdiction of ultimate parent). Only jurisdictions with at least 20 observations of MNE consolidated accounts in ORBIS are included. In the case of turnover, data in the matrix includes intra-group transactions, while they are netted out in consolidated ORBIS data, explaining why turnover tends to be relatively higher in the data from the matrix. Likewise, in the case of profits, data in the matrix focus on 'sub-groups with positive profits', while profits are net in consolidated ORBIS data, explaining why profit tends to be relatively higher in the data from the matrix. The coverage of payroll in consolidated account data from ORBIS was judged insufficient to include payroll in the comparison.
Source: OECD Secretariat.

Annex 5.F. Robustness of the matrices to a different data source ordering rule

Annex Table 5.F.1. Alternative order of data sources to fill the matrices: Differences with the baseline matrices

Differences with the amounts in the baseline matrices when the order of preference of the first and second preferred data sources are inverted

Panel A: Profit matrix

		Jurisdiction of ultimate parent				
		High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	+2%	+23%	+6%	+11%	**+3%**
	Middle income (105)	-12%	-3%	+5%	-15%	**-7%**
	Low income (29)	+3%	+2%	+9%	+4%	**+6%**
	Investment Hubs (24)	+1%	+3%	+6%	+2%	**+1%**
	Total	**+1%**	**-1%**	**+9%**	**+0%**	**+1%**

Panel B: Turnover matrix

		Jurisdiction of ultimate parent				
		High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	+2%	+4%	+1%	+6%	**+2%**
	Middle income (105)	-6%	-0%	+0%	-26%	**-4%**
	Low income (29)	-0%	-0%	-21%	-0%	**-7%**
	Investment Hubs (24)	+0%	+0%	+1%	+0%	**+0%**
	Total	**+1%**	**+0%**	**-11%**	**-5%**	**+0%**

Panel C: Tangible assets matrix

		Jurisdiction of ultimate parent				
		High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	+3%	+1%	-0%	+9%	**+3%**
	Middle income (105)	-5%	-0%	-1%	-19%	**-3%**
	Low income (29)	-0%	+0%	+0%	+1%	**-0%**
	Investment Hubs (24)	-0%	+0%	+6%	-0%	**+0%**
	Total	**+2%**	**+0%**	**-0%**	**-5%**	**+1%**

Panel D: Payroll matrix						
		Jurisdiction of ultimate parent				
		High income	Middle income	Low income	Investment Hubs	Total
Jurisdiction of affiliate	High income (64 jurisd.)	-1%	-6%	0%	-8%	**-2%**
	Middle income (105)	-0%	0%	0%	0%	**-0%**
	Low income (29)	0%	0%	0%	0%	**0%**
	Investment Hubs (24)	0%	0%	0%	0%	**0%**
	Total	**-1%**	**-1%**	**0%**	**-5%**	**-1%**

Note: For example, when the first and second data sources (i.e. CbCR and ORBIS) are switched in the preference order rule to fill the profit matrix, the amount of profit in the 'high income'-'high income' cell is increased by 2% compared to the amount in the baseline matrix presented in Table 5.5. Groups of jurisdictions (high, middle and low income) are based on the World Bank classification. The number of jurisdictions in each group is indicated in parentheses. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.
Source: OECD Secretariat.

Notes

[1] In the rest of this chapter, anonymised and aggregated Country-by-Country Report (CbCR) data are simply referred to as CbCR data.

[2] In this report, groups of jurisdictions (high, middle and low income) are based on the World Bank classification of jurisdictions by income group. Investment hubs are defined as jurisdictions with a total inward FDI position above 150% of GDP.

[3] For a detailed discussion of these issues, see OECD (2020[3]) and https://www.oecd.org/tax/tax-policy/anonymised-and-aggregated-cbcr-statistics-disclaimer.pdf.

[4] For example, a potential source of inconsistency across sources is that CbCR data focuses only on MNE groups with global revenues above EUR 750 million, which is not the case for the other data sources in this report. As more than 90% of the worldwide profit and turnover of MNE groups is generated by groups that are above this revenue threshold, this difference is assumed not to be overly consequential, which is confirmed by the benchmarking undertaken in this report.

[5] More specifically, the turnover matrix was used as a proxy measure to identify where revenues from the 'undertaxed payments rule' would accrue. The potential recipients of revenues from the 'income inclusion rule', which would accrue to the jurisdiction of ultimate parent of the MNE, have been identified directly based on the information in the profit matrix.

[6] A potential formulaic substance-based carve-out would imply that the amount of low-taxed profit subject to the minimum tax would be reduced in relation to the level of economic activity of the MNE in the jurisdiction where this profit is located. The amount of economic activity could be measured based on criteria including tangible assets depreciation and payroll (see Chapter 3).

7 See discussion in Box 3 of OECD (2020[3]) and also https://www.oecd.org/tax/tax-policy/anonymised-and-aggregated-cbcr-statistics-disclaimer.pdf.

8 For a detailed discussion on ORBIS coverage and representativeness, see Bajgar et al. (2020[26]).

9 See more details on FDI data and methodology, see http://www.oecd.org/corporate/mne/statistics.htm.

10 In a relatively small number of cells, data are missing in the panel on 'sub-groups with positive profits', but not missing in the panel focusing on all sub-groups. For these cells, data from the latter panel were used instead of the former, except in the few cases where the reported amount of profit in the panel focusing on all sub-groups was negative. In these cases, CbCR data have been considered missing for the purpose of building the matrices.

11 Out of the 26 jurisdictions of ultimate parent included in the OECD publication of 2016 CbCR data, data from one ultimate parent jurisdiction (China) has not been used in the analysis in this chapter. This is because Chinese CbCR data for 2016 are based only on a subsample of 82 CbCRs, while it is estimated that a significantly larger number of CbCRs were filed in China for the fiscal year 2016.

12 In 2016, 1 101 US MNE groups reported CbCR, against 1 575 in 2017. The total turnover of these groups was respectively USD 16.3 trillion and USD 21.6 trillion.

13 See discussion in Box 3 of OECD (2020[3]) and also https://www.oecd.org/tax/tax-policy/anonymised-and-aggregated-cbcr-statistics-disclaimer.pdf.

14 This calculation was based on Table II.B 1-2 from the BEA data on the activity of MNEs, which provides data on the balance sheet of foreign affiliates of US MNEs and in particular separates inventories from property, plant and equipment. Inventories are found to represent on average 24% of the total of inventories, property, plant and equipment of US MNEs in 2016 and in 2017. As a result, the CbCR data in the tangible assets matrix has been scaled down by 24%, except for US MNEs where the adjustment was based on the exact share of inventories of US MNEs in each market jurisdiction, when available in the BEA data.

15 The quality of ORBIS coverage has been assessed based on benchmarking of key variables against CbCR data and aggregate numbers from the Analytical AMNE database. For 22 jurisdictions, ORBIS is used both for domestic-owned and foreign-owned MNE entities. For two jurisdictions, it is used only for foreign-owned entities. See detailed lists in Annex 5.A.

16 This selection has been done in the following way. For each MNE group, the total profit of the group in a given jurisdiction has been computed by summing the profit of all entities belonging to the group in that jurisdiction, based on ORBIS data. All entities from MNE groups in a loss position in a given jurisdiction have been eliminated from the final ORBIS dataset used as a data source for the four matrices.

17 A possible future refinement would be to apply a EUR 750 million global revenue threshold to entities belonging to MNE groups where the necessary information on global revenues at the group-wide level is available in ORBIS.

18 For example, if the value of tangible assets is missing for 10% of MNE entities (weighted by turnover) in a jurisdiction, the total estimate of tangible assets of MNE entities in this jurisdiction has been scaled up by 1/(1-10%)=11.1%.

19 This approach was inspired by Tørsløv et al. (2018[1]) who apply the ratio of gross operating surplus to labour compensation in FATS to the compensation of employees in national accounts. The main limitation

of the approach is that the ratio of payroll to turnover may vary across industries, which cannot be accounted for with the available data.

20 An alternative approach to the whole extrapolation methodology would have been to use data on FDI income instead of FDI positions as a starting point. While this would a priori have seemed a more direct approach, it would have posed significant challenges for extrapolation (FDI income being more volatile than FDI positions) and, more importantly, it would have made it difficult to identify ultimate investors, because the available data on FDI by ultimate investor focus only on FDI positions and not on FDI income.

21 The recent literature using pseudo maximum-likelihood methods in gravity models has often used a method based on a Poisson distribution (PPML) as first described by Santos Silva and Tenreyro (2006[20]). Head and Mayer (2014[21]) compare different methods trying to achieve the same goals (in particular a better handling of zeroes in the variable of interest) and suggest that methods based on the Poisson distribution (PPML) and the Gamma distribution (GPML) perform best, and that there is no obvious reason to prefer one method over the other since their relative performance will depend on the structure of the error term. Without any strong *a priori* on this structure, the GPML method has been chosen for the extrapolation of turnover in this chapter because it provided a better fit with the available benchmarks (i.e. alternative data sources such as CbCR data and US BEA data) among the type of jurisdictions for which extrapolations play an important role (e.g. low income jurisdictions). The two methodologies yield broadly similar coefficients in the estimation, and broadly similar aggregates: for instance, global turnover in the matrix is 0.6% higher after using GPML compared to PPML.

22 AMNE data and Analytical AMNE data were not included in this regression because AMNE does not include diagonal terms and the diagonal terms in Analytical AMNE are generally based on imputations.

23 An additional difficulty is that turnover can be distorted by profit shifting behaviour, in which case it may not give a good indication of the level of economic activity, and, in turn, of tangible assets in a jurisdiction. This is a limitation of the approach that is difficult to fully address with the available data.

24 In jurisdictions with less than ten MNE groups in ORBIS consolidated account data, the data is considered to be insufficiently representative and no adjustment is made (i.e. 'Delta 2'=0). Similarly, in jurisdictions present in less than three ultimate parent jurisdictions in CbCR data, it is assumed that 'Delta 1'=0. For the purpose of computing 'Delta 2', turnover in ORBIS consolidated account data is rescaled to take into account the fact that consolidated turnover does not include intra-group sales, while these sales are included in the turnover matrix. This rescaling is done based on the global ratio of unrelated to total revenue in CbCR data, which is 69%.

25 The ratio of payroll to turnover computed on the aggregate inward and outward AMNE data (13.5%) is broadly similar to the average in ORBIS. Differences might be explained by the restricted sectoral coverage in AMNE (which often excludes the financial sectors), the different geographical coverage in the two sources, and the fact that the ORBIS ratio is computed on subgroups with positive profits while this distinction is unavailable in AMNE.

26 More precisely, the adjustment is based on the share of depreciation in gross operating surplus by market jurisdiction, as computed by Tørsløv et al. (2018[1]). In jurisdictions where no data are available in Tørsløv et al. (2018[1]), the ratio for the "Rest of the world" is used.

27 In practice, when the total of a column in the profit matrix for a given jurisdiction of ultimate parent was less than 50% of the total consolidated profit of MNE groups from this jurisdiction of ultimate parent, as observed in the ORBIS consolidated account dataset, the diagonal cell for jurisdiction in the matrix (when based on extrapolation) has been adjusted upwards to cap the difference at 50%. In addition, when the

diagonal cell in the profit matrix for a jurisdiction was above 100% of the total consolidated profit of MNE groups from this jurisdiction of ultimate parent, as observed in the ORBIS consolidated account dataset, the diagonal cell has been adjusted downwards, to 100% of this total. Similar adjustments were made in the turnover matrix. These adjustments mainly focus on smaller jurisdictions, where data quality issues may be more frequent and where the matrices rely more on extrapolations than in larger jurisdictions. In the turnover matrix, this adjustment affects 48 jurisdictions (out of 222) and the total turnover in the cells affected represents 0.5% of the total turnover in the turnover matrix. In the profit matrix, this adjustment affects 51 jurisdictions, and the total profit in the cells affected by the adjustment represents 4.4% of the total profit in the profit matrix.

[28] No similar adjustment is made in the tangible assets and payroll matrices since the imputed values in those matrices rely on the turnover matrix, which has already been adjusted.

[29] One limitation of the matrices for the analysis of profit shifting is that they focus only on MNE sub-groups with positive profits, and therefore do not reflect potential loss-shifting behaviour.

[30] The authors only report an exact number for 2017, but their Figure 6 suggests a level of global FDI positions slightly below USD 35 trillion.

[31] Austria, Canada, Switzerland, Czech Republic, Germany, Estonia, Finland, France, Hungary, Iceland, Italy, Lithuania, Poland, Turkey, and the United States.

[32] For example, Borga and Calandro (2018[12]) define "pass-through capital [as] capital that flows into one economy and that is subsequently invested in another economy".

[33] In a recent paper, Coppola et al. (2020[27]) show that similar patterns occur when considering the issuance of securities. For instance, they find that classifying securities by the ultimate issuer (the parent) instead of the immediate issuer (an affiliate) increases substantially the level of portfolio investment from developed economies to emerging economies.

[34] SPEs are entities that have little or no employment, physical presence, or operations in a jurisdiction but that do provide important services to the MNE, such as holding assets and liabilities or raising capital (see https://www.oecd.org/daf/inv/How-MNEs-channel-investments.pdf).

[35] Here, "outward" FDI is not used as the "reporting" principle under which the data is reported since the FDI data by ultimate investor is only available in the reports that some jurisdictions make about their inward FDI. The term "outward" is used here to specify the direction of the flow: for instance, the inward FDI reported by France from the United States is an outward flow from the United States to France.

[36] For instance, OECD data shows that in 2016, France was the ultimate investor for USD 42 billion of FDI positions in France, or 6.1% of its global inward FDI; while the United States was the ultimate investor for USD 73 billion FDI positions in the United States, or 1.9% of its global inward FDI.

[37] Among the 15 jurisdictions reporting FDI data by ultimate investor, Austria does not report an aggregate inward FDI position in the OECD data.

[38] Jurisdiction-level results are similar when considering data from the IMF balance of payment statistics for FDI income and the IMF CDIS dataset for FDI positions.

From:

Tax Challenges Arising from Digitalisation – Economic Impact Assessment

Inclusive Framework on BEPS

Access the complete publication at:
https://doi.org/10.1787/0e3cc2d4-en

Please cite this chapter as:

OECD (2020), “Construction of the data “matrices” underlying the impact assessment”, in *Tax Challenges Arising from Digitalisation – Economic Impact Assessment: Inclusive Framework on BEPS*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/823ced63-en