Cadestin, C. *et al.* (2018-02-27), "Multinational enterprises and global value chains: the OECD analytical AMNE database", *OECD Trade Policy Papers*, No. 211, OECD Publishing, Paris.
<http://dx.doi.org/10.1787/d9de288d-en>

OECD Trade Policy Papers No. 211

Multinational enterprises and global value chains: the OECD analytical AMNE database

Charles Cadestin,
Koen De Backer,
Isabelle Desnoyers-James,
Sébastien Miroudot,
Davide Rigo,
Ming Ye

OECD TRADE AND AGRICULTURE DIRECTORATE

This paper is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and the arguments employed herein do not necessarily reflect the official views of OECD countries.

The publication of this paper has been authorised by Ken Ash, Director of the Trade and Agriculture Directorate

Comments are welcome and should be sent to tad.contact@oecd.org.

© OECD (2018)

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

MULTINATIONAL ENTERPRISES AND GLOBAL VALUE CHAINS: THE OECD ANALYTICAL AMNE DATABASE

Charles Cadestin, Koen De Backer, Isabelle Desnoyers-James,
Sébastien Miroudot, Davide Rigo and Ming Ye (OECD)

In order to better understand the interdependencies between trade and investment in global value chains (GVCs), the OECD has developed a new dataset on the Activities of Multinational Enterprises (AMNE). This dataset starts from official AMNE statistics and combines the information with Inter-Country Input-Output (ICIO) tables to provide new insights on the trade-investment nexus in GVCs. This dataset allows the contribution of domestic firms, multinational enterprises (MNEs) and their foreign affiliates to global trade and production to be assessed. This paper details the methodology that was used to create the data, as well as the main assumptions and challenges in the work.

Key words: Multinational enterprises, global input-output, global value chains, trade in value-added, linkages

JEL codes: F14, F23, L16, L23

Acknowledgements

This paper was prepared by Charles Cadestin, Sébastien Miroudot and Ming Ye from the Trade and Agriculture Directorate and Koen De Backer, Isabelle Desnoyers-James and Davide Rigo from the Directorate for Science, Technology and Innovation. The report benefitted from discussions in the OECD Working Party of the Trade Committee, which agreed to make this paper more widely available through declassification on its responsibility on 15 December 2017. The Committee on Industry, Innovation and Entrepreneurship also discussed this paper and provided valuable feedback. The authors are grateful to Maria Borgia, John Drummond, Oscar Lemmers and Zhi Wang for useful comments and inputs.

TABLE OF CONTENTS

Executive Summary	4
Introduction	5
1. Splitting the ICIO according to ownership: Data challenges and main steps	6
2. Reconciling AMNE data with the ICIO framework: Output, value-added and trade	8
3. Splitting and balancing the ICIO according to ownership	19
4. Concluding remarks	21
Annex: Data sources	25

Tables

Table 1. Econometric results of the regression on foreign affiliate sales	13
Table 2. Estimation of total output of foreign-owned firms by country and by industry	15
Table 3. Estimation of output by country, industry and by country of ownership	16
Table 4. Comparison between the original AMNE database and analytical AMNE results (2014).....	22
Table A1. Classification of industries	25
Table A2. List of data sources for AMNE statistics.....	27

Figures

Figure 1. Illustration of ICIO splitting according to ownership.....	7
Figure 2. Adjustments on turnover data.....	10

EXECUTIVE SUMMARY

In order to better understand the interdependencies between trade and investment in global value chains (GVCs), the OECD has developed a new dataset on the Activities of Multinational Enterprises (AMNE). This dataset starts from official AMNE statistics and combine the information with Inter-Country Input-Output (ICIO) tables to provide new insights on the trade-investment nexus in GVCs. This dataset allows assessing the contribution of domestic firms, multinational enterprises (MNEs) and their foreign affiliates to global trade and production. It brings to MNE analysis the same conceptual shift introduced by TiVA (Trade in Value-Added statistics) in trade analysis.

While the ultimate goal of the project is to fully split a world input-output table according to the ownership of firms (domestic-owned versus foreign-owned), the project involves the construction of a full matrix of the output of foreign affiliates in 43 countries plus the rest of the world, as well as similar matrices for value-added, exports and imports. These matrices are already useful tools to carry out analysis on the role of MNEs in GVCs.

The methodology was first applied to the OECD ICIO build in the context of the TiVA project. The methodology was then used to create similar tables with the World Input-Output Tables (WIOTs) available up to 2014. This paper describes the methodology using the WIOD database, and which will be applied to continuing work with TiVA in the future.

The methodology includes two major steps. The first consists of the estimation of a bilateral matrix of output in all countries and industries according to the country of ownership of firms. The starting point is the OECD AMNE database complemented with additional national sources when available. The data are made consistent with output as measured in the ICIO and the missing information is estimated by various statistical means, in particular through FDI data and gravity regressions. The matrix is then balanced through an optimisation. In addition to this matrix describing world output by country, industry and country of ownership, a trade matrix is created for the exports and imports of domestic-owned and foreign-owned firms, as well as a matrix of value-added ratios. These matrices are first filled with the existing available information and then completed with estimates. Since the ICIO already provides the total figures, we make assumptions on differences between domestic-owned and foreign-owned firms in value-added ratios or export and import intensities.

The second step in the methodology is the splitting *per se* of the WIOD ICIO with all cells divided into the contribution of domestic-owned and foreign-owned firms. This is done in a single quadratic optimisation that uses as objectives the data from the three matrices previously constructed (output, value-added and trade by country, industry and ownership) and as a constraint the output, value-added and trade reported in the WIOD database. The resulting tables are fully balanced and have the same properties as the original WIOD tables.

Adding an ownership dimension allows us to revisit TiVA statistics to show the contribution of foreign-owned firms to exports and domestic sales. Following the same methodology, a third category of firms, the domestic MNEs –as opposed to domestic firms with no foreign affiliate– is added for a sub-set of countries for which data are available over the years 2011–2014.

Introduction

The development of Inter-Country Input-Output (ICIO) tables represents a significant step towards better understanding global value chains (GVCs). These tables have allowed researchers and policymakers to measure trade in value-added terms (Koopman et al., 2014) and to identify the contribution of each country and industry to the value of final products (Johnson and Noguera, 2012). However, the picture is still incomplete. GVCs are not only composed of independent companies exporting and importing intermediate and final products. Many of the firms involved are multinational enterprises (MNEs) who have a network of foreign affiliates resulting from their foreign investment.

As it is now, the main ICIOs, such as TiVA, WIOD or EORA¹, do not provide any information on the role played by foreign affiliates in GVCs. When ‘domestic value’ is added to exports, it can be the value added by domestic-owned firms but it can also be some value added by foreign-owned firms established in the country. The fact that foreign ownership is involved is not without implications. For example, in terms of income, it is likely that the activities of the foreign affiliates benefit the parent economy, either through direct transfers (e.g. repatriation of profits) or through spillover effects.

In order to provide new insights on the role and importance of MNEs in GVCs, the OECD has developed an analytical AMNE database (AMNE meaning Activities of Multinational Enterprises). This project has resulted in the creation of a full matrix of world output by country and by industry, split according to the ownership of firms. Three categories of firms are identified: foreign-owned firms, domestic MNEs (i.e. domestic firms with foreign affiliates) and other domestic firms (i.e. domestic firms without foreign affiliates). The second phase in the project was to fully split an ICIO according to this ownership dimension. Due to the challenges in distinguishing domestic MNEs from other domestic firms, the tables are split between the three categories only for the years 2011–2014 and for a sub-set of 16 countries. For other years and countries, there is a split only between domestic-owned and foreign-owned firms.

Splitting the ICIO according to ownership helps to account for the heterogeneity among firms with respect to their sourcing strategies. In current ICIOs, foreign-owned and domestic-owned firms –but also MNEs and non-MNEs– share the same production function and rely on the same mix of inputs. There are, however, important differences in the way these three types of firms use foreign and domestic inputs. For example, Koopman et al. (2008) find much lower ratios of domestic value-added in the exports of foreign-owned firms in the People’s Republic of China (hereafter “China”). Fetzner and Strassner (2015) highlight that there are differences in the composition of output among domestic firms, comparing those that are part of a multinational enterprise and those that are part of an enterprise entirely located in the United States.

The heterogeneity within industries is not a new issue in the input-output literature. In the past, the focus was mainly on differences across regions within countries. As pointed out by Miller and Blair (2009), it is a well-known fact that the production of electricity in Washington State by water power is based on a different mix of inputs as compared to the production by nuclear power elsewhere in the United States. Moreover, the production of new products generally requires an input mix that may differ from traditional products in the same sector.

There are numerous examples of input-output analyses that aimed at splitting input-output tables at the regional level, starting in the 1950s (Isard, 1951). This type of analysis has been extensively used in regional science research with an improvement in methodologies over years. For example, multi-regional input-output tables have been created for the United States (Polenske, 1980) or for China (Okamoto and Ihara, 2005).

1 . TiVA is a database from the OECD and WTO using inter-country input-output tables developed by the OECD (OECD, 2013). Data are available at <http://oe.cd/icio>. The World Input-Output Database (WIOD) is a project funded by the European Union and the Netherlands Organisation for Scientific Research (Timmer et al., 2015; Timmer et al., 2016) with data available at <http://www.wiod.org>. EORA was developed at the University of Sydney with funding from the Australian Research Council (Lenzen et al., 2013). Data can be found at <http://www.worldmrio.com>.

More recently, multi-regional input-output tables have been created at the global level with the ‘region’ corresponding to countries in a model of the world economy. This type of Inter-Country Input-Output tables was first developed for Asia by IDE-JETRO in the 1980s. In the past decade, global models have been developed in the context of the TiVA, WIOD and EORA projects, previously mentioned, as well as EXIOBASE. In a world of GVCs where production is split across countries, these tables aim at disentangling the value added by each country in the output of each industry.

In this work, the challenge of accounting for heterogeneity among producers or regions within countries has been already emphasised. Some studies have started to integrate regional characteristics into ICIO tables (Dietzenbacher et al., 2013; Los et al., 2017). The OECD ICIO tables account for the heterogeneity among exporting and non-exporting firms for two specific countries – China and Mexico – (in the case of China, exporters are also further split to account for processing exports).

The analytical AMNE database discussed in this paper aims at splitting the entire ICIO according to another dimension, the ownership of firms. Two existing ICIOs have been split in the course of the project: the OECD ICIO and the World Input-Output Tables (WIOTs) from the WIOD project. In this paper, we present the work done with the WIOD tables that were updated in December 2016 and for which data are available up to 2014 (Timmer et al., 2016). This database includes 43 countries plus the rest of the world, and 56 sectors from 2000 to 2014. We have kept all countries and all years but due to data limitations we do not have full results for the 56 industries. Our tables are initially split in 43 industries, with still lines of zeroes and more aggregated figures for some countries where not all the industry detail is available.²

This paper is organised as follows. Section 1 presents some of the data challenges and the main steps in the creation of the split ICIO. Section 2 explains how we have reconciled the statistics on activities of multinational enterprises (AMNE) with national accounts in the ICIO framework and estimated missing values. Section 3 summarises the methodology for splitting the ICIO itself using the output, trade and value-added matrices by ownership. Section 4 concludes. Data sources are detailed in the Annex.

This paper focuses on the methodology underpinning the development of the analytical AMNE database. It does not include the presentation of individual results. But some results can be found in two reports relying on the analytical AMNE database (Cadestin et al., 2018; Andrenelli et al., 2018).

1. Splitting the ICIO according to ownership: Data challenges and main steps

There are four main elements in the ICIO table: the intermediate consumption matrix, the final demand matrix, the value-added vector and the gross output vector. Cells across columns correspond to a country/sector’s inputs; cells across lines correspond to the output of a country/sector. Both rows and columns sum to the gross output in each country/sector. Figure 1 illustrates how each cell of the intermediate consumption matrix in the initial ICIO (Panel A) is divided into four cells corresponding to the inputs used by domestic-owned and foreign-owned firms (Panel B). The final demand matrix is split only across rows to reflect the final demand of products from domestic-owned and foreign-owned firms. The value-added and gross output vectors are split across columns to indicate the value-added and gross output of domestic-owned and foreign-owned firms in each country and sector. With the resulting ICIO, we can for instance calculate the input requirements of a foreign-owned firm operating in sector 2 and country 2 from a domestic-owned firm operating in sector 1 of country 1. When distinguishing domestic MNEs from other domestic-owned firms, each cell in Panel A is split into nine cells in Panel B.

This framework assumes that we know all the transactions among domestic-owned and foreign-owned firms across all countries and industries. For 44 countries, 56 sectors and two types of firms (domestic-owned and foreign-owned) the intermediate consumption matrix reports more than 24 million transactions, as

2. It should be noted that the original WIOD tables already have lines of zeroes and more aggregated figures for some non-EU countries where data are not fully available for the 56 industries.

opposed to “only” 6 million in the initial ICIO. It becomes almost 55 million when splitting the tables into three categories of ownership.

Last but not least, the consistency of the ICIO should be kept when splitting the table, with the sum of the intermediate consumption and value-added equal to output along columns and the sum of intermediate output and final demand equal to the same output along lines. The balancing of the split table also introduces additional challenges due to the overall size of the matrices involved.

Figure 1. Illustration of ICIO splitting according to ownership

Data sources

In a perfect world, we would have direct information on all the transactions of domestic-owned firms, domestic MNEs and foreign-owned firms within national accounts and we could split the ICIO by linking these data across countries with a methodology similar to the one used to construct the initial ICIO. But far from the perfect world, we can only recreate all the transactions by using various statistical methods and by starting from more aggregated data.

The analytical AMNE database is built on two main sources: (1) data on MNEs based and extended from the OECD database on the Activities of MultiNational Enterprises (AMNEs), and (2) the underlying World Input-Output Database (WIOD) which provide the whole structure of the ICIO when not distinguishing domestic-owned and foreign-owned firms and that we do not change.

The OECD AMNE database contains the official data collected and published by National Statistical Offices and Central Banks on activities of MNEs. The coverage of AMNE data has been increasing over time, with a particularly good coverage for the United States and most EU countries (as an EU regulation obliges countries to provide AMNE data to Eurostat). The OECD AMNE database contains data for 32 OECD countries plus Costa Rica and Lithuania, over more than 50 industries with the first year of reporting 1985. However, information is not equally available across countries, industries and years with data typically less available for earlier years and more disaggregated levels (e.g. bilateral at industry level).

The criterion for ownership is based on majority, i.e. an affiliate is regarded as foreign as soon as it has at least 50% of foreign ownership. For each variable, countries may report the “inward” activities of foreign enterprises in their territory – in such case the reporter is the country hosting the affiliate; otherwise countries report activities of their affiliates that are based abroad, i.e. the “outward” activities, and the reporter is the country of ownership. We trust more the results coming from inward AMNE statistics but we also use outward statistics to gather additional data.

In addition to the OECD AMNE database, information is derived from the OECD Trade by Enterprise Characteristics (TEC) database and data from National Statistics Offices. Given the scarcity of information, we collect any data that exists to our knowledge and work with different concepts (such as sales or turnover) in order to derive the required data for output, value-added and trade. Section 2 provides more detail on how these different data sources have been combined. The result is that the analytical AMNE database includes information on sales, output, value-added and exports of foreign-owned and domestic-owned firms in each country and industry.

As a second main information source, the WIOD database is a set of Inter-Country Input-Output tables with 43 countries plus the rest of the world and 56 sectors in the ISIC rev. 4 classification (Table A.1 in the Annex) and relies on the System of National Accounts 2008 (Timmer et al. 2015; Timmer et al., 2016). Due to data availability for sales of domestic-owned and foreign-owned firms, we collapse the WIOD tables into 43 sectors. The country and sector coverage is detailed in the Annex. WIOD tables are in basic prices meaning that taxes on products minus subsidies, trade margins and transport margins have been removed from the value of transactions. The estimated trade and transport margins are added respectively to wholesale and retail trade and transport sectors.

Methodology

The methodology consists of two parts. The first step is to construct three balanced matrices of world output, value-added and trade (exports and imports) according to the ownership of firms and to distinguish MNEs from non-MNEs among domestic firms when possible. It involves the reconciliation of AMNE data with national accounts, the estimation of missing values and balancing procedures. The second part is the splitting of the ICIO *per se*. It requires a methodology to infer all the detailed transactions of the split ICIO from the three matrices previously created and additional balancing to obtain a consistent final matrix. This methodology is described in more detail in the following sections.

2. Reconciling AMNE data with the ICIO framework: Output, value-added and trade

In 2014, the gross output of UK manufacture of chemicals and pharmaceutical products was 67 billion USD in the AMNE database, while it reached 95 million USD in WIOD tables. This example shows that even when AMNE data are available, they can drastically differ from WIOD which is built from the System of National Accounts (SNA). There are numerous statistical challenges due to methodological differences in the collection of data for AMNE statistics and national accounts:

- SNA data are in basic prices while AMNE data are in purchaser prices. Concretely, it means that the gross output of a manufacturer from the AMNE database contains the value of its production, but also taxes minus subsidies on the product, trade and transport margins. On the contrary, in the WIOD ICIO table, these taxes and margins are discounted and margins are respectively reallocated to wholesale and retail trade and transport sectors. It should not change the overall output at the country level but the allocation between industries.
- WIOD provides gross output whereas AMNE statistics mostly report the turnover or sales of firms. Turnover corresponds to the revenue of the firms while the gross output refers to production. In retail sectors, firms do not produce what they sell which leads to a much higher figure for turnover than for gross output. Also, firms that produce goods that are not sold increase their gross output but their production of unsold goods is part of their inventory; hence turnover is not affected.

- The sample of firms that is used for AMNE data may also differ from the sample for SNA data. The SNA takes into account all the firms and all the activities while AMNE databases are for some countries merely based on surveys of a sample of firms. In the case of the EU, even if the methodology has been harmonised by Eurostat, there are still differences across countries (see the Annex).
- The surveyed unit is also an important concern when it comes to ownership. For instance in the United States, Fetzer and Strassner (2015) compare the BEA MNE data with Statistics on Income as they are both at the firm level. An establishment is a single physical location with a main production activity while a firm can be an establishment or a set of establishments: it can encompass several sectors. Whether the firm or the establishment is surveyed therefore influences results at the sector level. A firm in the automotive industry which encompasses a wide array of establishments in different sectors registers its income statements for the automotive sector. If the survey is at the firm level, the figures of the establishments of the firms are not accounted in their respective sectors but in the firm's sector. The OECD AMNE database is for all countries based on the enterprise unit since 2005.
- The same problem arises for trade data when the unit is the product instead of the business entity that produces it. For example, our database contains export and imports observations from the Trade by Enterprise Characteristics (TEC) dataset which reconciles business and trade registers at the enterprise level. Differences may appear with WIOD where trade is at the industry level.
- Lastly, most of the AMNE data that we have on hand are at their most disaggregated level at 2 digits. Such level of disaggregation does not always permit a straightforward conversion to ISIC revision 4 as in WIOD.

Initial output matrix

In this section, we describe the data treatments that are made on the AMNE database for output values. For each data source, databases are imported and sectors are converted from their national classification to consistent ISIC revision 4. In some cases it involves splitting the data to estimate values for sub-sectors. For example, the 'Manufacture of pulp, paper and paper products; publishing and printing' in the NACE rev. 1 classification encompasses 'printing and reproduction of recorded media' but also 'publishing activities' in ISIC rev. 4. This sector needs to be split into two if we do not want to allocate all the sales of foreign affiliates to the paper manufacturing sector instead of the publishing service sector.

The second step consists of converting turnover data to output data (when only the turnover or sales are available). Turnover corresponds to the revenue of firms while gross output refers to production. When we have only turnover, we adjust the data to assess the equivalent output. The adjustment is based on data from countries for which we have both the turnover and output (such as EU countries in Eurostat data). The ratio of gross output to turnover is calculated and applied to the turnover values.

In Figure 2, the closer the value is to 1, the lower is the adjustment. The sector that is mostly adjusted is therefore the wholesale and retail trade sector where turnover includes the value of all the goods sold while gross output is based on the margin of the wholesaler or retailer. The total column highlights that adjusting for output reduces the overall AMNE turnover by about one third.

The output matrix is 'bilateral' as we keep the information from the AMNE database on the country of ownership in addition to the country where the firm is established. This bilateral dimension allows for further analysis by country of ownership but all foreign-owned firms are then collapsed when splitting the ICIO (independently of their country of ownership). It would not be manageable to work with a full ICIO split according to the country of ownership (as it would multiply each cell by the square of the number of countries in the matrix of intermediate consumption), but for analytical purposes it is possible to construct indicators that include the use of the country of ownership in their calculations (using a bilateral matrix of value-added coefficients for example).

Figure 2. Adjustments on turnover data

Note: Industry classification in Table A.1 in the Annex.

At the year-country-industry level, the number of observations we have for foreign-owned firms is 9 908 out of 28 380 potential data points for the full matrix. At the year-country-partner-industry level, the number of observations is 389 143 out of 1.2 million for the full matrix. The next step consists in estimating all these missing values.

Estimation of missing values

Many data points are missing. When there is no observation on the activities of MNEs, these data are estimated. This section sets out the different methodologies that were applied in order to construct the pre-optimisation matrix of output by year, host country, country of ownership and industry. Further explanations on when these different estimation methods are used are given in the next sub-section.

The gravity equation

The gravity dataset comprises 107 countries that represent 96% of world GDP³. The dependent variable is output. It is filled with the available figures at bilateral level: we consider any AMNE data point available and we use mirror statistics when the reporter's value is not available. We also use the adjusted turnover when output is not available. Then, in order to distinguish between “missing values” and zeros⁴, we consider that there is zero AMNE sales when there is zero investment. We use information coming from foreign direct investment (FDI) statistics to identify such cases. While FDI and AMNE data are different for several reasons,⁵ we assume that in case zero FDI is reported (as a stock) in a given sector for a specific parent

3. According to the World Development Indicators database.
4. Although the AMNE statistical guidelines ask for the identification of zeros, confidentiality issues may result in a missing value as opposed to what is simply zero.
5. Because of differences in concepts and statistics, there is now a consensus in the literature to regard FDI data as a biased measure of foreign affiliate activity (Beugelsdijk et al., 2010; Ali-Yrkkö and Leino, 2014; Blanchard and Acalin, 2016). First, FDI statistics provide information on cross-border capital flows which may be eventually sent to other countries without contributing at all to the local economy. This is especially the case for the so-called Special Purpose Entities (SPE) used as financial vehicles to shift profit and risks across countries. A large presence of these SPEs in a country typically results in high FDI flows reported for that country without the corresponding economic effects. These SPEs also explain why FDI inflows and outflows are strongly correlated for these countries. Recent initiatives have been taken to collect and present

country, no foreign affiliate of that home country was established in the home country and hence that sales are zero. We estimate the gravity coefficients based on this dataset.

Gravity models have a solid theoretical foundation (Anderson and Van Wincoop, 2004) and have produced some of the most robust empirical results in the trade literature. Although originally used to explain trade flows, gravity models have also been successfully used to estimate FDI flows and foreign affiliate sales.

The theoretical and empirical underpinning of the econometric extrapolation is the framework developed by Bergstrand and Egger (2007). Their model extends the knowledge-capital model pioneered by Markusen (2002), providing a theoretical framework for estimating gravity equations of aggregate bilateral FDI and sales of foreign affiliates. This framework lays out a tractable model that specifically identifies gravity variables as the sole determinants of FDI patterns and foreign affiliate sales.

In accounting for foreign affiliate sales we need to take into considerations the three broad types of MNE investments emerged from the literature: horizontal, vertical and export-platform FDI. However, within GVCs, networks of MNE subsidiaries are based on a mix of the three, with trade becoming complementary to FDI. Horizontal FDI arises when the parent company creates a plant in a foreign country producing the same product (Markusen, 2002). Vertical FDI arises when multinationals locate a subsidiary in a foreign country in order to exploit factor cost differentials (Markusen, 2002). Ekholm, Forslid and Markusen (2007) extend the literature to incorporate export-platform FDI when multinationals invest horizontally but with the objective of also serving third markets demand through exports.

The empirical literature has identified three main drivers: market size and market potential (i.e. host-market and third-market size); relative production costs between host and partner country; and relative market access costs (i.e. all the costs associated with exporting to a market versus setting up a foreign affiliate there).

We use host-country and partner-country GDP as a measure of expenditure and output in the location and investing country, respectively. Ideally we would like to include data on sectoral expenditure and output rather than GDP as such (Anderson and Yotov, 2010). However, this is not possible with our sample of countries.⁶ Also, relatively cheaper production costs lower the cost to set up a foreign affiliate, increasing FA sales. The difference between GDP per capita of host and controlling country is included in order to account for efficiency seeking FDI⁷.

If the costs for setting up a foreign affiliate is lower than those of exporting (e.g. variable or fixed trade costs), it would be more profitable for the multinational to invest in the location country. We thus include various empirical proxies for bilateral trade costs typical of the gravity literature⁸, such as the geographical distance between countries i and j (distance), a dummy variable that equals one for countries that share a common land border (contig), a dummy variable that equals one for country pairs that share a common official language (comlang), a dummy variable that equals one if countries i and j were once in a colonial relationship (colony), a dummy variable that equals one for country pairs that were colonized by the same power (comcol) and a dummy variable for shared legal origins (comleg).

FDI data without SPEs. Second, FDI only measures part of what foreign affiliates use to finance their activities and excludes the often-substantial amount of capital they raise from local sources. Third, as FDI is a financial input, hence excluding the contribution of labour, FDI stocks underestimate MNE activity in countries where labour is relatively more productive (Cadestin *et al.*, 2018).

6. It would be more correct to use a measure of sectorial output in order to account for countries comparative advantages in certain sectors which translate in investments abroad in those sectors. However, the WIOD database only provides output for 43 countries and further aggregation for the rest of the world.

7. Data on GDP and GDP per capita come from the World Bank Development Indicators.

8. These variables come from the CEPII database, see Mayer and Zignago (2011).

Finally, we augment this baseline specification to include variables accounting for investments costs. We thus account for host-country institutional and business environment adding a measure of political instability and violence (stability), and regulatory quality (regquality) from the World Bank Governance Indicators. These measures account for the costs to set up a foreign affiliate in the location country.

Host and partner fixed effects should be included in order to account for the multilateral resistances terms (Anderson and Van Wincoop, 2004). Given our dependent variable, the trade costs potentially vary by sector and time and so the multilateral resistance terms cannot be adequately captured by host- and partner-country fixed effects. Instead, to be consistent with the theoretical foundations of the model we would need sector, host-sector-year, partner-sector-year and year fixed effects (Yotov et al., 2016). A feasible alternative is to estimate the model using host, partner, sector and year fixed effects, assuming that the multilateral resistance terms do not change in the time period considered. The resulting econometric specification is the following:

$$y_{ijst} = \exp \left[\alpha_0 + \beta_1 \ln(GDP_{it}) + \beta_2 \ln(GDP_{jt}) + \beta_3 \left(\ln \left(\frac{GDP}{capita_{it}} \right) - \ln \left(\frac{GDP}{capita_{jt}} \right) \right) \right. \\ \left. + \beta_4 \ln(distance)_{ij} + \beta_5 comlang_{ij} + \beta_6 colony_{ij} + \beta_7 contig_{ij} + \beta_8 comcol_{ij} \right. \\ \left. + \beta_9 stability_{i,t} + \beta_{10} regquality_{i,t} + \delta_t + \delta_i + \delta_j + \delta_s \right] * \varepsilon_{ijst}$$

where i stands for location country (host), j for controlling country (parent), s for sector and t for time.

The first challenge of the econometric analysis is to deal with zeros. Our sample provides sector level foreign affiliate sales in 107 host and source countries, spanning the years from 2000 to 2014 with 10% of zero observations. The second challenge is the inconsistency of the OLS estimator in the presence of heteroskedasticity in the error terms and log-transformed dependent variable. We follow the trade literature and use the Poisson Pseudo Maximum Likelihood estimator (PPML), proposed by Santos Silva and Tenreyro (2006). Previous studies on the determinants of foreign affiliate sales (Bekkers and Girgzdyte, 2015; Fukui and Lakatos, 2012) have compared the performances of the PPML and zero inflated models, such as the ZIP (zero inflated Poisson) and ZINB (zero inflated negative binomial). The main arguments raised against the PPML are that it tends to under-predict the number of zeros and foreign affiliate sales generally exhibit over-dispersion, in contrast to the underlying assumption of the PPML which assumes the mean and variance to be equal.

However, the comparison with zero inflated models seems to be erroneous. First, simulation results in Santos Silva and Tenreyro (2011) show that the PPML estimator is well behaved even when the proportion of zeros is very large. Second, contrary to what was stated by previous papers, the PPML allows both for over- and under-dispersion and it is consistent as a pseudo-maximum likelihood estimator regardless of how the data are in fact distributed. The only improvement that could come from allowing for over-dispersion would be in terms of efficiency. However, for the efficiency gain to be real the exact nature of the over-dispersion would need to be known, which it usually is not.⁹ Third, the zero inflated estimators assume that the excess zeros are generated by a different process, which is not the case in our dataset (all zeros are actual zeros). Finally, the zero inflated estimators have an undesirable property: they are not scale invariant. Thus, results from a model with sales in thousands of US dollars as the dependent variable will be different from those obtained with sales in millions of US dollars.

Regression results of sales of foreign affiliates are presented in Table 1, using the PPML estimator. The coefficients are all significant and with the expected sign. The R-squared is at 93%. The coefficient on distance is negative suggesting a form of complementarity between trade and foreign affiliate sales. A possible explanation is that an increase in transportation costs, associated to higher distance between host and partner countries, makes vertical FDI less economical, decreasing foreign affiliate sales. This result is

9. From <http://personal.lse.ac.uk/tenreyro/LGW.html>.

consistent with the proliferation of GVCs, in which multinationals locate stages of their production processes in different countries searching for location specific advantages such as low costs for factors of production. In addition, the variables approximating for set-up costs, such as having a border in common, a common language and being once in a colonial relationship, have all the expected positive signs.

Using the coefficients from this estimation, we can fill the gaps in the output matrix for values that are not assumed to be zero. For the ‘rest of the world’ -for which we do not have AMNE data-, the values come from the sum of the predicted bilateral figures for a selection of countries that are not covered in WIOD.¹⁰

Table 1. Econometric results of the regression on foreign affiliate sales

	(1) OLS - bilateral	(2) PPML - bilateral	(3) PPML - industries
Distance (log)	-1.112*** (0.0216)	-0.423*** (0.0212)	-0.482*** (0.0245)
Host country's GDP (log)	-0.325 (0.282)	-0.0742 (0.321)	-1.842*** (0.538)
Ownership country's GDP (log)	2.207*** (0.277)	1.766*** (0.322)	3.195*** (0.547)
Difference of GDP per capita (log)	-1.378*** (0.288)	-1.078*** (0.322)	-2.382*** (0.504)
Contiguity	0.474*** (0.0534)	0.227** (0.0881)	0.198*** (0.0560)
Common language	0.0843 (0.0528)	0.0512 (0.0590)	0.0149 (0.0586)
Colonial relationship	0.731*** (0.0535)	0.204*** (0.0434)	0.250*** (0.0743)
Common legal origin	0.658*** (0.0318)	0.373*** (0.0295)	0.514*** (0.0399)
Regquality of the host country	0.289*** (0.0862)	0.123 (0.0915)	0.0592 (0.0909)
Stability of the host country	0.0935* (0.0515)	0.121** (0.0573)	-0.0589 (0.0938)
Constant	-10.69*** (1.558)	-15.14*** (1.709)	-18.27*** (2.127)
Observations	17,034	39,712	1,167,448
R-squared	0.761	0.927	0.791

Note: Clustered standard errors by country pair in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Other estimation methods

While the gravity equation is our main source of estimates, we use additional methods when we already have some data that can provide a good indication of what missing values are. For example, when FDI data are available, we take advantage of their correlation with the output of foreign affiliates through an Ordinary Least Square regression¹¹:

10 . Algeria, Argentina, Azerbaijan, Bangladesh, Belarus, Bosnia and Herzegovina, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, Guatemala, Hong Kong, China, Iceland, Iran, Israel, Kazakhstan, Kenya, Kuwait, Malaysia, Morocco, Myanmar, New Zealand, Nigeria, Oman, Pakistan, Peru, Philippines, Qatar, Saudi Arabia, Singapore, South Africa, Sri Lanka, Sudan, Thailand, Ukraine, Uruguay, Uzbekistan, Venezuela, Viet Nam.

11 . We have indicated before that FDI is not strongly correlated with the sales of foreign affiliates but the regression below introduces gross output and fixed effects to better predict the output of foreign affiliates based on FDI.

$$\log(X_{ijt}^{foreign-owned}) = \alpha + \beta_{FDI} * \log(FDI_{ijt}^{inward}) + \beta_X * \log(X_{ijt}) + \beta_C \cdot C_i + \beta_I \cdot I_j + \beta_Y \cdot Y_t + \epsilon_{ijt}$$

where X_{ijt} is gross output C_i , I_j and Y_t are the dummy variables for country, industry and year. The R-square of the model is 95% and the inward FDI stock is significant with a p-value under 1%. The exponential of the predicted values replace the missing output values.

There are a few observations where the predicted values are not within the WIOD boundaries (i.e. a foreign output between 0 and the total sectoral output); particularly for small countries such as Luxembourg where the output is small for some sectors. As a last resort, in order to estimate values that are consistent with WIOD data, an ordinary least square regression is used on the transformed ratio of foreign output on total sector output with the logit function. Such transformation allows us to estimate values that fit in the boundaries of the WIOD framework. We estimate this ratio on year, country and industry dummies. When there are no data that fit the limits imposed by the WIOD framework, the share of foreign output is regressed on the country and industry average:

$$\text{logit}\left(\frac{X_{ijt}^{foreign-owned}}{X_{ijt}}\right) = \alpha + \beta_C \cdot C_i + \beta_I \cdot I_j + \beta_Y \cdot Y_t + \epsilon_{ijt}$$

In this case, the R-square is lower at 67%.

Three time-series estimation methods are also used. The first is the linear interpolation: a missing value in the middle of two existing values is linearly interpolated. Second, for missing values that cannot be interpolated as they are not in the middle of two non-missing values, a moving average with the longest memory possible is used, with more weight for the most recent observations. Third, exponential smoothing is also used. This method is an adaptive forecasting algorithm that allows forecasting the missing observations based on the past values and an adaptive parameter. We let the algorithm decides the adaptive parameter by minimising the sum of the squared forecast errors. The next sub-section gives more details on when and why these different methods are used.

Balanced bilateral output matrix

With the AMNE data and the estimates made in the previous section, we can now build a full matrix of bilateral output by country and by industry for each year. Output is “bilateral” because the country of ownership is a dimension in the matrix. Cells where the country of ownership is the same as the country of production correspond to the output of domestic-owned firms (along a block diagonal where each block is the vector of industries in each country). The other elements in the matrix, where the country of ownership differs from the country of production, reflect the output of foreign-owned firms. The bilateral output matrix has about 1.2 million observations and is in the dimension: country (of output), country of ownership, industry and year.¹²

This matrix must be fully consistent with the WIOD database and between different aggregation levels. We use a quadratic optimisation where we minimise the square of the difference between the starting values and the values that fit the objective functions under the constraint of matching exactly the WIOD values for output by country and by industry for each year. We introduce two objective functions in this optimisation on the basis of the work done with the AMNE data. First, the sum of all the output of foreign firms by country and industry should be equal to total foreign output by country and industry as assessed on the basis of the AMNE data (and adjusted to be consistent with WIOD data). Second, the sum of all the output of firms by

12. Other efforts at creating a comprehensive matrix of sales of foreign affiliates in the country-partner-sector dimension include Fukui and Lakatos (2012) and Alviarez (2016). Ramondo *et al.* (2015) also create such matrix but not including the sector dimension. However, these three studies rely on the concept of sales (and not output) with no reconciliation with national accounts data.

country (of output) and by country of ownership should be equal to the total by country and by country of ownership coming from the adjusted AMNE data.

The values are obtained through the process described in Table 2 from the most straightforward and accurate estimation method to the least. In the first step (S1.1), we feed the matrix with the share of output of foreign-owned firms that we apply to the WIOD database total output. This is our most accurate estimation but possible only for 45% of the observations. The assumption is that what differs between AMNE statistics and WIOD data affects proportionally domestic-owned and foreign-owned firms. Also, for non-business industries such as public administration, activities of households and international organisations, we assign zero output for foreign-owned firms (11% of the observations), as these industries have by definition no foreign presence.

In a second step (S1.2), when the sectoral total is known for at least one year, we use the growth rate of the AMNE figure that we apply to the closest estimation from first step instead of using the value as it is recorded in the AMNE database. In doing so, we avoid some breaks in series that are due to the differences in accounting methodologies between the AMNE database and the WIOD database.

When none of the years are available but we still know the share of foreign-owned firms from the AMNE database, we directly take the gross output or the adjusted turnover of the AMNE statistics in the third step (S1.3). In this case, there is no attempt at correcting differences between the AMNE database and the WIOD framework. It means that the output is still in purchaser prices for example.

In a fourth step (S1.4), gaps are filled by means of a linear interpolation and moving average, the moving average being used for the first missing observation that is not interpolated. As in Ramondo et al. (2013), we take advantage of the similarities between FDI and AMNE data. In our database, the correlation coefficient between FDI and the estimated output from the previous steps is 0.77. The FDI database at the country/industry level contains values from the OECD and from the IMF Coordinated Direct Investment Survey (CDIS) database. The benchmark definition of FDI has evolved across years. From 2005 onward, the 4th benchmark definition takes into accounts Special Purposes Entities. However, in applying the OLS estimation to the FDI data as described in the previous sub-section, this step only provides 744 additional observations.

Table 2. Estimation of total output of foreign-owned firms by country and by industry

Step	Description	Number of observations	Percentage	Cumulative percentage
S 1.1	Share applied to WIOD output figures + ^public sectors with no foreign presence (ISIC rev.4 industries O, T and U)	12 642	45%	45%
S 1.2	Applying the growth rate of output/turnover with unknown sectoral total to previous step values	323	1%	46%
S 1.3	Total is unknown: gross output or adjusted turnover	1 651	6%	52%
S 1.4	Linear interpolation / moving average	1 945	7%	58%
S 1.5	Estimation based on FDI data	744	3%	61%
S 1.6	Applying the growth rate of the gravity model to previous step estimates	3 530	12%	73%
S 1.7	Raw gravity estimates	6 826	24%	97%
S 1.8	Exponential smoothing	114	0%	98%
S 1.9	Regression on the country and industry average	605	2%	100%
		28 380		

The remaining steps are pure estimations. In step S1.6, the same method as in step 1.2 is used but the growth rate also comes from the gravity estimates (with the output of foreign-owned firms calculated as the sum of the bilateral relationships). This gravity growth rate is applied to the closest known observation. However, this is not always feasible; in the absence of any observation, the gravity estimates are directly used in step S1.7.

Still some gravity estimates may not fit into the matrix. It is the case for 719 observations where the estimated output of foreign-owned firms is higher than WIOD output. A first attempt at solving this case is the exponential smoothing methodology that allows extrapolating the values based on past data. As a last resort, the missing values are estimated through a regression of the share of foreign-owned firms in output that includes country, industry and year fixed effects as described in the previous sub-section. By mathematical construction, such estimation is between 0 and total output, however, this is based on the country's results in other industries as well as the global industry's average.

For the second objective matrix with bilateral totals by country (of output) and country of ownership and for the most disaggregated matrix at the country-partner-industry level, we use a similar process. There are however fewer cases, as illustrated in Table 3, since no adjustment has to be made to match WIOD data, as they do not have this country of ownership dimension.

Bilateral values are either available from the inward or outward AMNE statistics (step S2.1). If not, we extrapolate the values (S2.2) using the observations available for the country/partner pair. When no information is available for a given country pair, gravity estimates are used (S2.3). In this case, we calculate the remaining output to be allocated to foreign firms as the difference between the total foreign output and the sum of output already allocated to some partners. Then, we calculate the share of each country of ownership among missing countries of ownership, and we apply this share to the remaining output. In the starting matrix, we assign a value for each year, country, partner and industry, which is tantamount to filling a matrix of 1.2 million observations.

Tables 2 and 3 have established a hierarchy among estimates with additional trust or accuracy in the methods listed first. For each step, we calculate the average deviation between the estimated value and the first step value. Based on this standard deviation, a confidence index is calculated. This confidence index is an important indicator for the construction of the balanced matrix at the bilateral level: it is used as a weight in the quadratic optimisation to influence the results. The values in which we have more confidence have a higher weight in the objective function. The minimisation process will therefore avoid changing these values too much as compared to the ones we trust less and that have a lower weight. The confidence index is calculated according to the average deviation of the corresponding value to the value for which we are the most confident in. For the figures estimated with the gravity equation, the confidence index is calculated according to the number of observations that are available for host countries, sectors and partner countries.

Table 3. Estimation of output by country, industry and by country of ownership

Step	Description	Number of observations	Percentage	Cumulative percentage
S 2.1	AMNE output, adjusted turnover and public sectors	516 845	41%	41%
S 2.2	Extrapolation	93 413	7%	49%
S 2.3	Gravity shares applied to remaining country-partner pairs	638 462	51%	100%
		1 248 720		

Once we have the starting bilateral output matrix, the two objective matrices, the WIOD constraints and the confidence index, we run the following quadratic optimisation:

$$\begin{aligned} \min o(\hat{v}) = & \sum_{i,j,k} \gamma_{i,j,k} * (v_{i,j,k}^{starting} - \hat{v}_{i,j,k})^2 + 10 * \sum_{i,j} \gamma_{i,j} * \left(v_{i,j}^{bil.obj.} - \sum_k \hat{v}_{i,j,k} \right)^2 + 100 \\ & * \sum_{i,k} \gamma_{i,k} * \left(v_{i,k}^{ind.obj.} - \sum_j \hat{v}_{i,j,k} \right)^2 \\ & s. t. \sum_j \hat{v}_{i,j,k} = output_{i,k}^{WIOD} \end{aligned}$$

where γ are the respective confidence indices, v are the values from the starting and objective matrices and $\hat{v}$ are the 1.2 million values estimated through minimisation of the above objective function. The resulting matrix is a balanced bilateral output matrix by country, country of ownership and industry that perfectly matches the WIOD output data (for all years in the dataset).

Value-added and trade matrices

Once the bilateral output matrix is created, the following step consists in producing value-added and trade matrices (exports and imports) also split according to ownership and consistent with WIOD data. For these matrices, we use the same methodology which consists in applying the difference between foreign-owned firms and domestic-owned firms in the AMNE database to the WIOD framework. But it involves again dealing with differences between AMNE totals and WIOD figures.

For reasons similar to what we explained about output, value-added figures in AMNE data are not equal to the value-added data in WIOD. In particular, value-added is in purchaser prices in AMNE data and therefore different from basic prices due to taxes minus subsidies. Furthermore, in the case of trade, there is no correction for merchanting in the AMNE database. Merchanting appears when a company buys for direct re-sell abroad without adding any input to the product; it inflates trade flows. In order to create the value-added and trade matrices, we need both to estimate missing values and to reconcile AMNE data with WIOD figures. The methodology for the value-added matrix is presented below, but it is the exact same one used for the trade matrix.

First, value-added at the sectoral level is split between domestic-owned and foreign-owned value-added: $v = v_d + v_f$. Second, value-added can be expressed as the value-added per output unit times output:

$$v = \frac{v_d}{x_d} \cdot x_d + \frac{v_f}{x_f} \cdot x_f$$

where v corresponds to value-added at the sectoral level, v_d is domestic-owned value-added, x_d is domestic output and the subscript f applies to foreign firms for each variable.

Now let define parameter p as the premium ratio between foreign-owned firms value-added intensity and domestic-owned firms value-added intensity:

$$p = \frac{\frac{v_f}{x_f}}{\frac{v_d}{x_d}}$$

Integrating p into the equation leads to:

$$x_d + p \cdot x_f = v * \left(\frac{x_d}{v_d} \right)$$

Domestic value-added can be estimated as:

$$v_d = \frac{v}{1 + p \cdot \left(\frac{x_f}{x_d}\right)}$$

Then foreign value-added is $v_f = v - v_d$.

Such methodology was selected because it fulfils the two objectives. First, it reconciles the WIOD value-added by industry with the information from AMNE data. Domestic value-added plus foreign value-added is equal to the WIOD total. Second, the methodology facilitates the estimation of missing values: the only information that is required from the AMNE database is the “premium” ratio p which represents the difference in the value-added to output ratio between foreign-owned firms and domestic-owned firms. When p is missing, we use the average value of the premium at the closest level available or for comparable countries and industries.

However, with such methodology, the estimation of value-added can potentially provide values that are higher than output. It is the case when $p \leq \frac{v-x_d}{x_f}$, leading to values for domestic value-added higher than domestic output, or when $p \geq \frac{x_d}{v-x_f}$ (in this case the foreign value-added is higher than the foreign output). When it happens, we chose the closest value of p that fits into the constraint of value-added being lower than output.

For trade, the same methodology is employed, based on differences in export-intensity and import-intensity among domestic-owned and foreign-owned firms. The resulting matrices are exports and imports, by country, industry and type of ownership (domestic or foreign).

Distinguishing domestic MNEs from other domestic-owned firms

We now have a database that breaks down the sectoral output of 43 countries plus the rest of the world. It tells us how much of the total production of a sector is made by foreign-owned firms. In concrete terms, German firms produced 7,753 million USD in the French manufacture of chemicals and pharmaceutical products in 2014, it represented 6% of the total production of this sector – meanwhile 56% was domestic. However, there is still a missing part in the MNE puzzle and not the least. This is the distinction between the domestic MNEs and other domestic-owned firms that are not MNEs. What share of this 56% of domestic-owned production is attributable to purely domestic-oriented firms and to domestic MNEs? In this section, we explain the methodology to further split the domestic-owned production into domestic MNEs and other domestic-owned firms not involved in international investment.

We only dispose of enough data for 16 countries¹³ and still not all sectors are covered; plus we hardly have any information on the value added. The Trade by Enterprise Characteristics (TEC) database provides figures on trade flows of domestic MNEs and other domestic-owned firms, but only for 2011–2014. We therefore limit our sample to countries available and to the time coverage 2011–2014, corresponding to a total of 2 752 observations. First, we use the raw data when available and populate about 42% of the database. In the absence of raw data, we use FDI statistics in order to fill the zeros: when outward FDI is equal to zero, we set the output of domestic MNEs to zero¹⁴; it only adds 12 observations. Remaining values are estimates, first

13 . Austria, Belgium, Germany, Finland, France, United Kingdom, Hungary, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Poland, Portugal, Sweden and United States.

14 . It is safe to use FDI data in this case since the limitations on the comparison between AMNE and FDI statistics do not apply. First, we work in the country and industry dimension and not with bilateral data. Second, foreign investments are reported by the parent firm thus identifying the relevant industry.

according to the country's output¹⁵. Lastly, we regress the share of domestic MNE output on country, industry and year levels that we apply to WIOD output, as in the 9th step of the construction of the starting matrix (S1.9).

3. Splitting and balancing the ICIO according to ownership

We now have three matrices (output, value added and trade) that include information on domestic-owned and foreign-owned firms, the WIOD ICIO and a database on the breakdown of the domestic output between MNEs and non-MNEs. The next step consists in splitting the WIOD ICIO along the ownership dimension. It is presented as one step as it is done through a single optimisation.

The basic idea is to use the sector-ownership level gross output that we created in the previous steps to determine the relative proportion of domestic and foreign value within each sector as starting values. We also use the value-added, exports and imports data by country, sector and ownership that determine the balancing conditions. The methodology predicts values through a quadratic programming model that fits the WIOD ICIO data with values the closest as possible to the AMNE matrices of gross output, value-added, exports and imports. In this section, we outline the methodology using a simple example.

Let us define an ICIO composed of G countries and n sectors. Z_{ij} is an $n \times n$ matrix and its elements indicate the delivery of intermediate inputs from country i to country j , the special case $i = j$ therefore corresponds to domestic deliveries. Let define V_i a vector of dimension $1 \times n$ whose elements indicate the value-added in country i and Y_{ij} a matrix of dimension $n \times n$ that denotes final goods produced in country i and consumed in country j .

We also define X_i^{D*} and X_i^{F*} as country i 's gross output for respectively domestic-owned and foreign-owned in the gross output matrix. We have $X_i^{D*} + X_i^{F*} = X_i$ where X_i is the vector of gross output for country i . Moreover, we define the vector of output ratios by domestic-owned firms as $\sigma_i^D = X_i^{D*} / X_i$ and vector of output ratios by foreign-owned firms as $\sigma_i^F = X_i^{F*} / X_i$.

Z_{ij} is split into four matrices using the proportionality assumption: Z_{ij}^{DD} , Z_{ij}^{DF} , Z_{ij}^{FD} and Z_{ij}^{FF} . This split is for the initial values in the optimisation, the coefficients will then change in the optimisation to reflect the constraints and the objectives. At the end, we obtain different production functions and a different mix of inputs for domestic- and foreign-owned firms both as suppliers of inputs and purchasers of inputs.

Z_{ij}^{DD} is the matrix of intermediate inputs supplied by domestic-owned firms to domestic-owned firms. Z_{ij}^{DF} is a matrix of intermediate inputs supplied by domestic-owned firms to foreign-owned firms; and so forth for Z_{ij}^{FD} and Z_{ij}^{FF} . The starting values of the four Z matrices are calculated as follows (with the hat notation used for the diagonal matrix of the vector):

$$Z0_{ij}^{DD} = \hat{\sigma}_i^D Z_{ij} \hat{\sigma}_j^D, Z0_{ij}^{DF} = \hat{\sigma}_i^D Z_{ij} \hat{\sigma}_j^F, Z0_{ij}^{FD} = \hat{\sigma}_i^F Z_{ij} \hat{\sigma}_j^D \text{ and } Z0_{ij}^{FF} = \hat{\sigma}_i^F Z_{ij} \hat{\sigma}_j^F$$

We also split the Y_{ij} matrix into two matrices: Y_{ij}^D and Y_{ij}^F where Y_{ij}^D is the final demand for the output of domestic-owned firms and Y_{ij}^F is the final demand for the output of foreign-owned firms. The starting values of these two matrices are calculated as follows:

$$Y0_{ij}^D = \hat{\sigma}_i^D Y_{ij} \text{ and } Y0_{ij}^F = \hat{\sigma}_i^F Y_{ij}$$

15. $\log(X_{i,j,t}^{MNEs}) = \alpha + \beta_X * \log(X_{i,j,t}) + \beta_{C,I} * D_{C,I} + \epsilon_{i,j,t}$ where $D_{C,I}$ stands for the combination of country industry fixed effects.

V_i is split into two vectors: V_i^D and V_i^F . V_i^D is the value-added vector for country i 's domestic-owned firms and V_i^F is the value-added vector for country i 's foreign-owned firms. The starting values of these two vectors are extracted from the value-added matrix created in the previous steps.

$$V0_i^D = V_i^{D*} \text{ and } V0_i^F = V_i^{F*}$$

To obtain the unobservable I-O coefficients, we need to estimate the new intermediate input blocks in the ICIO table: Z_{ij}^{DD} , Z_{ij}^{DF} , Z_{ij}^{FD} and Z_{ij}^{FF} , the new final demand blocks, Y_{ij}^D and Y_{ij}^F , as well as the new value-added vectors, V_i^D and V_i^F . Each block should satisfy these constraints: 1) the sum of the split new blocks should be equal to the original matrices/vectors in the WIOD tables; 2) the new ICIO should be balanced, i.e. the sum of each row and sum of each column should be equal to output. These constraints can be written as follows:

$$Z_{ij}^{DD} + Z_{ij}^{DF} + Z_{ij}^{FD} + Z_{ij}^{FF} = Z_{ij}$$

$$Y_{ij}^D + Y_{ij}^F = Y_{ij}$$

$$V_i^D + V_i^F = V_i$$

$$\sum_j Z_{ij}^{D*} + \sum_j Y_{ij}^D = \sum_j Z_{ji}^{*D} + V_i^D = X_i^{D\bullet}$$

$$\sum_j Z_{ij}^{F*} + \sum_j Y_{ij}^F = \sum_j Z_{ji}^{*F} + V_i^F = X_i^{F\bullet}$$

where the notation $*$ corresponds to the set $\{D, F\}$ that identifies the domestic and foreign blocks in the split ICIO tables.

Additional constraints are needed to split the exports and imports data in a way consistent with the matrices created with AMNE database. These constraints are:

$$E_i^D = \sum_j Z_{ij}^{D*} + \sum_j Y_{ij}^D, j \neq i \quad E_i^F = \sum_j Z_{ij}^{F*} + \sum_j Y_{ij}^F, j \neq i$$

$$M_i^D = \sum_j Z_{ji}^{*D} + \sum_j Y_{ji}^D, j \neq i \quad M_i^F = \sum_j Z_{ji}^{*F} + \sum_j Y_{ji}^F, j \neq i$$

Using the above notations, the objective function in the optimisation is specified as:

$$\begin{aligned} \text{Min } S = & \sum_{i,j} \frac{(Z_{ij}^{DD} - Z0_{ij}^{DD})^2}{Z0_{ij}^{DD}} + \sum_{i,j} \frac{(Z_{ij}^{DF} - Z0_{ij}^{DF})^2}{Z0_{ij}^{DF}} + \sum_{i,j} \frac{(Z_{ij}^{FD} - Z0_{ij}^{FD})^2}{Z0_{ij}^{FD}} + \sum_{i,j} \frac{(Z_{ij}^{FF} - Z0_{ij}^{FF})^2}{Z0_{ij}^{FF}} \\ & + \sum_{i,j} \frac{(Y_{ij}^D - Y0_{ij}^D)^2}{Y0_{ij}^D} + \sum_{i,j} \frac{(Y_{ij}^F - Y0_{ij}^F)^2}{Y0_{ij}^F} + 100 * \left(\sum_i \frac{(V_i^D - V0_i^D)^2}{V0_i^D} + \sum_i \frac{(V_i^F - V0_i^F)^2}{V0_i^F} \right) \\ & + 100 * \left(\sum_i \frac{(E_i^D - E_i^{D\bullet})^2}{E_i^{D\bullet}} + \sum_i \frac{(E_i^F - E_i^{F\bullet})^2}{E_i^{F\bullet}} + \sum_i \frac{(M_i^D - M_i^{D\bullet})^2}{M_i^{D\bullet}} + \sum_i \frac{(M_i^F - M_i^{F\bullet})^2}{M_i^{F\bullet}} \right) \end{aligned}$$

where E_i^{D*} , E_i^{F*} , M_i^{D*} and M_i^{F*} are the exports and imports values from the AMNE matrices created in the previous steps.

This process allows to fully split the WIOD table on the basis of domestic and foreign ownership with at the end balanced tables that have exactly the same figures as in the initial tables when not distinguishing the foreign-owned and domestic owned firms. We follow the same methodology with a further split for domestic MNEs for the years 2011–2014.

4. Concluding remarks

The construction of the database involves a wide range of statistical treatments from the data collection to the quadratic optimisation. To summarize, taking stock of the data, converting industries in a unified framework, dealing with differences of concepts and data coverages between the AMNE statistics and the system of national accounts, estimating the missing values, populating the matrix in a consistent way, optimising it and finally splitting the ICIO. What do all these tasks imply for the AMNE figures?

The following table illustrates how values are transformed in the process, comparing the data before and after the different treatments. The left side of the table is in million USD and the right side is in shares. The table shows the foreign output at country level, the adjusted turnover (as described in section 2), the sum of the sectoral foreign output (with missing values) and the sum of the corrected sectoral foreign output. The corrected figures correspond to the estimation of output from turnover data when the output is not available. We also show these values before (section 2) and after the optimisation (section 2) in the last two columns. A difference between the turnover and the output is expected but another salient feature of the data is the difference between the sum of foreign output at the sectoral level (third column) and the total foreign output at the country level (first column).

Our estimates follow the published data at the sectoral level. The perimeter of the total figure diverges across countries. For example, for Eurostat countries, the total excludes agriculture, insurance and non-business sectors which account for about 10% of the economy. For the United States, the total encompasses all economic activities. Furthermore, we are not able to say whether the difference between the total and the sectors are due to missing values or the values at sectoral level are under or over estimated. As our figures follow the sectoral allocation, the correlation coefficient between the post-optimisation share and the sum of sectoral level is 99% while it is 90% for the total sector.

Differences between the pre- and post- optimisation figures highlight how the values have evolved due to several reconciliations. The first is the reconciliation between the AMNE data and the WIOD framework: the foreign plus the domestic output should be equal to the WIOD output. The second reconciliation is that in a given sector, the sum of the partners should be equal to the sum of foreign output; last, the third reconciliation is that the sum of sectors for a country/partner pair should be equal to its total. However, the evolution of the figures takes into account the confidence that we have in these different data: the less confident we are, the more the value deviates from its starting point. For Mexico, we see that the optimisation process has dramatically changed the values. The only data we have for Mexico are partners' data. It means that we are more confident in these bilateral data than in the sectoral foreign output.

There are several ways in which the work can be improved. First, there are on-going efforts at the level of national statistical offices and within international organisations involved in statistics to provide more data on the activities of foreign-owned firms in the context of national accounts. As illustrated by Fetzer and Strassner (2015) for the United States, the ownership dimension could be more systematically incorporated in national accounts. Access to such data could solve the two main issues we have encountered in this work: (i) the lack of information on the output and value-added of foreign-owned and domestic-owned firms and (ii) the discrepancies between output as measured in AMNE statistics and in national accounts.

It may take time for countries to release more information on MNEs within their national accounts and for the work we propose to be based on fewer assumptions and estimated data and more on official statistics. There are however existing data sources that could further be used for this work, such as firm-level data.

Within the project, we are already working with the ORBIS dataset from Bureau Van Dijk and some firm-level data for China. National datasets can provide more information on the output of domestic-owned and foreign-owned firms. The Trade by Enterprise Characteristics (TEC) database that we use for our trade matrix is an example of data compiled from firm-level information that allow the identification of domestic and foreign ownership.

Lastly, it is not clear to what extent the technical coefficients obtained from the optimisation are correctly reflecting the production functions of the different types of firms. Another step in the project will be to perform a sensitivity analysis and look more closely at how the different assumptions we made impact the stability of coefficients. Comparison with actual data (for countries that have created I-O information split according to the ownership of firms) can also help to assess the robustness of the methodology

Table 4. Comparison between the original AMNE database and analytical AMNE results (2014)

Gross output of foreign affiliates (in million USD)							Share of foreign affiliates in total output				
	Original AMNE data	After adjustment for turnover (section 2.i)	Based on the sum of sectoral output in original AMNE data	Based on the sum of sectoral output after adjustment for turnover (section 2.i)	Based on the sum of sectoral pre-optimisation output (section 2.iii)	Based on the sum of sectoral post-optimisation output (section 2.iv)	Original AMNE data	Based on the sum of sectoral output in original AMNE data	After adjustment for turnover (section 2.i)	Based on the sum of sectoral pre-optimisation output (section 2.iii)	Based on the sum of sectoral post-optimisation output (section 2.iv)
Australia					480,408	509,076				17.6	18.7
Austria	302,003	211,163	172,624	172,624	190,337	198,252	34.9	19.9	28.5	23.5	24.5
Belgium			248,310	248,310	307,682	317,345				27.7	28.6
Bulgaria			25,553	25,553	36,065	36,443				29.4	29.7
Brazil					490,126	548,307				11.9	13.4
Canada	995,108	973,055	808,072	542,698	707,775	744,054				21.8	22.9
Switzerland					337,303	402,353				24.1	28.8
China					2,271,076	2,300,713				7.2	7.2
Czech Republic	262,829	189,238	181,889	181,889	211,927	215,876	46.7	32.3	44.9	43.0	43.8
Germany	1,805,746	1,262,593	1,345,088	1,374,308	1,447,598	1,455,493	22.7	16.9	24.8	20.5	20.6
Denmark	142,681	99,763	83,217	83,217	103,343	116,134	23.2	13.5	19.4	16.8	18.9
Spain	620,708	537,188	343,826	343,826	557,429	560,758	27.2	15.1	17.4	21.7	21.8
Estonia	12,146	12,146	10,963	10,963	19,762	19,794	44.8	40.5	40.5	36.3	36.3
Finland	124,463	48,581	67,981	67,981	81,678	91,525	23.1	12.6	32.3	15.9	17.8
France	959,959	703,175	595,046	595,046	722,772	777,413	20.4	12.6	17.2	14.4	15.5
United Kingdom	2,136,938	1,309,476	1,117,737	1,117,737	1,399,958	1,456,976	35.2	18.4	30.1	26.5	27.6
Greece			16,733	16,733	32,065	37,228				8.5	9.9
Croatia			12,724	12,723	20,273	21,992				20.8	22.6
Hungary	184,329	128,884	110,352	110,352	136,007	139,034	52.8	31.6	45.2	47.8	48.9
Indonesia					145,990	164,435				8.5	9.6
India	271,842	190,074	177,462	146,988	169,103	174,683				4.2	4.4
Ireland					222,905	232,629				43.8	45.7
Italy	694,793	573,632	468,867	468,867	521,968	610,030	17.8	12.0	14.5	12.8	15.0
Japan	394,175	275,611	358,992	267,473	317,025	359,437				3.7	4.1
Korea					196,296	204,743				5.8	6.0
Lithuania	21,261	21,261	15,025	15,025	19,593	21,176	21.9	15.5	22.1	22.9	24.7
Luxembourg	108,632	75,956	39,404	39,404	89,969	96,793	52.1	18.9	27.0	42.4	45.7
Latvia	27,286	19,078	9,777	9,777	16,911	17,622	39.3	14.1	20.2	26.1	27.2
Mexico					295,031	354,235				13.8	16.6
Malta			638	638	2,729	4,041				9.4	14.0
Netherlands	671,576	408,346	342,924	342,924	429,579	445,148	35.3	18.0	29.6	25.7	26.6
Norway	236,713	165,512	174,687	174,687	186,164	202,047	27.8	20.5	29.4	22.3	24.2
Poland	345,880	241,842	237,576	237,576	351,244	343,275	37.2	25.6	36.6	31.8	31.1
Portugal	90,008	62,935	57,400	57,400	78,492	84,286	22.1	14.1	20.2	18.9	20.3
Romania			109,164	104,767	142,374	143,158				35.7	35.9
Rest of the world					2,092,479	2,064,524				8.2	8.1
Russia					170,256	207,973				5.0	6.2
Slovak Republic	111,900	78,241	78,781	78,781	100,859	102,493	51.8	36.5	52.2	44.0	44.7
Slovenia	31,450	21,990	15,446	15,446	19,689	21,392	29.1	14.3	20.4	20.9	22.7
Sweden	231,143	231,143	202,792	202,792	244,186	257,204	29.7	26.0	26.0	24.0	25.3
Turkey	137,716	137,716	120,389	120,389	166,957	160,161	13.8	12.1	12.1	11.2	10.7
Chinese Taipei					145,835	148,120				11.9	12.1
United States	4,134,287	2,890,727	4,141,778	2,951,413	2,967,356	3,177,397				9.6	10.3

References

- Ali-Yrkkö, J. and T. Leino (2014), “How Does Foreign Direct Investment Measure Real Investment by Foreign-Owned Companies? Firm-level analysis”, ETLA reports, No. 27, The Research Institute of the Finnish Economy, May.
- Alvarez (2016), “Multinational Production and Comparative Advantage”, Working Paper, Sauder School of Business, University of British Columbia.
- Anderson, J. E. and E. van Wincoop (2004), “Trade Costs”, *Journal of Economic Literature*, 42(3), pp. 691-751, <http://dx.doi.org/10.1257/0022051042177649>.
- Anderson, J. E., and Y. V. Yotov (2010), “The Changing Incidence of Geography”, *American Economic Review*, 100(5), pp. 2157-2186, <http://dx.doi.org/10.1257/aer.100.5.2157>.
- Andrenelli, A., C. Cadestin, K. de Backer, S. Miroudot, D. Rigo and M. Ye (2018), “Multinational production and trade in services”, *OECD Trade Policy Paper*, forthcoming.
- BEA (2014), *U.S. International Economic Accounts: Concepts and Methods*, Bureau of Economic Analysis, U.S. Department of Commerce.
- Bekkers, E. and I. Girgzyte (2015), “Sectoral Determinants of Foreign Affiliate Sales Employing European Data”, mimeo.
- Bergstrand, J.H. and P. Egger (2007), “A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises”, *Journal of International Economics*, 73(2), pp. 278-308, <https://doi.org/10.1016/j.jinteco.2007.03.004>.
- Beugelsdijk, S., J.-F. Hennart, A. Slangen and R. Smeets (2010), “Why and How FDI Stocks Are a Biased Measure of MNE Affiliate Activity”, *Journal of International Business Studies*, 41(9), pp. 1444-1459, <http://dx.doi.org/10.1057/jibs.2010.29>.
- Blanchard, O. and J. Acalin (2016), “What Does Measured FDI Actually Measure?”, Policy Brief PB16-17, Peterson Institute for International Economics, Washington DC.
- Cadestin, C., K. de Backer, I. Desnoyers-James, S. Miroudot, D. Rigo and M. Ye (2018), “Multinational Enterprises and Global Value Chains: New Insights on the Trade-Investment Nexus”, *STI Working Paper*, forthcoming.
- Dietzenbacher E., J. M. J. Guilhoto and D. Imori (2013), “The Role of Brazilian Regions in the Global Value Chain”, FEA/USP Working Paper Series, No. 2013-15.
- Ekholm, K., R. Forslid and J. Markusen (2007), “Export-Platform Foreign Direct Investment”, *Journal of the European Economic Association*, 2007, 5(4), pp. 776-795, <http://dx.doi.org/10.1162/JEEA.2007.5.4.776>.
- Eurostat (2012), *Foreign Affiliates Statistics (FATS) Recommendations Manual*, Publications Office of the European Union, Luxembourg.
- Fetzer, J. J. and H. Strassner (2015), “Identifying Heterogeneity in the Production Components of Globally Engaged Business Enterprises in the United States”, Bureau of Economic Analysis, Washington DC.
- Fukui, T. and C. Lakatos (2012), “A Global Database of Foreign Affiliate Sales”, Office of Economics Working Paper, No. 2012-08A, U.S. International Trade Commission, Washington DC.
- Isard, W. (1951), “Interregional and Regional Input-Output Analysis: A Model of a Space-Economy”. *Review of Economics and Statistics*, 33(4), pp. 318-328, <http://dx.doi.org/10.2307/1926459>.
- Johnson, R. and G. Noguera (2012), “Accounting for intermediates: Production sharing and trade in value added”, *Journal of International Economics*, 86(2), pp. 224-236, <http://dx.doi.org/10.1016/j.jinteco.2011.10.003>.

- Koopman, R., Z. Wang and S.-J. Wei (2008), “How much of Chinese exports is really made in China? Assessing domestic value-added when processing trade is pervasive”, NBER Working Paper Series, No. 14109, National Bureau of Economic Research, June.
- Koopman, R., Z. Wang and S.-J. Wei (2014), “Tracing value-added and double counting in gross exports”, *American Economic Review*, 104(2), pp. 459-494, <http://dx.doi.org/10.1257/aer.104.2.459>.
- Lenzen, M., D. Moran, K. Kanemoto and A. Geschke (2013), “Building Eora: A Global Multi-Regional Input-Output Database at High Country and Sector Resolution”, *Economic Systems Research*, 25(1), pp. 20-49, <http://dx.doi.org/10.1080/09535314.2013.769938>.
- Leontief, W. (1936), “Quantitative Input and Output Relations in the Economic System of the United States”, *Review of Economic and Statistics*, 18(3), pp. 105-125, <https://dx.doi.org/10.2307/1927837>.
- Los, B., M. Lankhuizen and J. Goddard (2017), “New Measures of Regional Competitiveness in a Globalizing world” in P. McCann, F. Can Oort and J. Goddard (eds), *The Empirical and Institutional Dimensions of Smart Specialisation*, Routledge.
- Markusen, J. R. (2002), *Multinational Firms and the Theory of International Trade*, MIT Press, Cambridge, MA.
- Mayer, T. and S. Zignago (2011), “Notes on CEPII’s distances measures: The GeoDist database”, CEPII Working Paper No. 2011-25, December.
- Miller, R. E. and P. D. Blair (2009), *Input-Output Analysis: Foundations and Extensions*, Second Edition, Cambridge University Press, Cambridge.
- Miroudot, S. and M. Ye (2017), “Tracing value-added and double counting in sales of foreign affiliates and domestic-owned companies”, paper prepared for the 25th International Input-Output Conference, Atlantic City, New Jersey, USA.
- OECD (2013), *Interconnected Economies: Benefitting from Global Value Chains*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264189560-en>.
- OECD (2017), “Inward activity of multinationals by industrial sector database”, Available at: http://stats.oecd.org/Index.aspx?DataSetCode=AMNE_IN.
- Okamoto, N. and T. Ihara (eds) (2005), *Spatial Structure and Regional Development in China. An Interregional Input-Output Analysis approach*, Palgrave Macmillan, Basingstoke, UK.
- Polenske, K. R. (1980), *The U.S. Multiregional Input-Output Accounts and Model*, Lexington Books (D. C. Heath and Co.), Lexington, MA.
- Ramondo, N., Rodríguez-Clare, A. and F. Tintelnot (2015), “Multinational Production: Data and Stylized Facts.”, *American Economic Review*, 105(5), pp. 530-536, <https://doi.org/10.1257/aer.p20151046>.
- Santos Silva, J.M.C and S. Tenreyro (2006), “The log of Gravity”, *Review of Economics and Statistics*, 88(4), pp. 641-658, <http://dx.doi.org/10.1162/rest.88.4.641>.
- Santos Silva, J.M.C. and S. Tenreyro (2011), “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator”, *Economics Letters*, 112(2), pp. 220-222, <https://doi.org/10.1016/j.econlet.2011.05.008>.
- Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer and G. J. de Vries (2015), “An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production”, *Review of International Economics*, 23, pp. 575-605, <http://dx.doi.org/10.1111/roie.12178>.
- Timmer, M. P., B. Los, R. Stehrer and G. J. de Vries (2016), “An Anatomy of the Global Trade Slowdown based on the WIOD 2016 Release”, GGDC research memorandum number 162, December.
- Yotov, Y., R. Piermartini, J.-A. Monteiro and M. Larch (2016), *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, WTO and UNCTAD, WTO Publications.

Annex

Data Sources

WIOD

The World Input-output Database (WIOD) is a set of Inter-Country Input-Output (ICIO) tables built by a consortium of institutions led by the Groningen Growth and Development Centre (GGDC) and funded by the European Union (Timmer et al., 2016). The 2016 update covers 43 countries¹⁶ plus the ‘rest of the world’ and 56 industries with annual tables from 2000 to 2014. It is constructed following the SNA 2008 framework with industries in ISIC Rev. 4. Table A.1 lists the 56 industries and indicates how they have been aggregated into 41 for the analytical AMNE database.

Table A1. Classification of industries (ISIC Rev. 4)

Label	Code
Agriculture, forestry and fishing	A
Mining and quarrying	B
Manufacture of food products; beverages; tobacco products	C10T12
Manufacture of textiles; wearing apparel; leather and related products	C13T15
Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	C16
Manufacture of paper and paper products; Printing and reproduction of recorded media	C17T18
Manufacture of coke and refined petroleum products	C19
Manufacture of chemicals and chemical products; basic pharmaceutical products and pharmaceutical preparations	C20T21
Manufacture of rubber and plastics products	C22
Manufacture of other non-metallic mineral products	C23
Manufacture of basic metals	C24
Manufacture of fabricated metal products, except machinery and equipment	C25
Manufacture of computer, electronic and optical products	C26
Manufacture of electrical equipment	C27
Manufacture of machinery and equipment n.e.c.	C28
Manufacture of motor vehicles, trailers and semi-trailers	C29
Manufacture of other transport equipment	C30
Manufacture of furniture; other manufacturing	C31T32
Repair and installation of machinery and equipment	C33
Electricity, gas, steam and air conditioning supply; water collection, treatment and supply	D_E36
Sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services	E37T39
Construction	F
Wholesale and retail trade; repair of motor vehicles and motorcycles	G
Land transport and transport via pipelines	H49

16 . Australia, Austria, Belgium, Bulgaria, Brazil, Canada, Switzerland, China, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Croatia, Hungary, Indonesia, India, Ireland, Italy, Japan, Korea, Lithuania, Luxembourg, Latvia, Mexico, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Slovak Republic, Slovenia, Sweden, Turkey, Chinese Taipei and United States.

Label	Code
Water transport	H50
Air transport	H51
Warehousing and support activities for transportation	H52
Postal and courier activities	H53
Accommodation and food service activities	I
Publishing activities	J58
Motion picture, video and television programme production, sound recording and music publishing activities; programming and broadcasting activities	J59T60
Telecommunications	J61
Computer programming, consultancy and related activities; information service activities	J62T63
Financial and insurance activities	K
Real estate activities	L
Professional, scientific and technical activities	M
Administrative and support service activities	N
Public administration and defence; compulsory social security	O
Education	P
Human health and social work activities	Q
Arts, entertainment and recreation; other service activities	RTS
Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	T
Activities of extraterritorial organizations and bodies	U
Total	TOTAL

AMNE database

Statistics on Activities of Multinational Enterprises encompass items that permit to evaluate the activities of enterprises evolving in a multinational environment. For this project, we are interested in five types of variables from AMNE statistics: gross output, turnover, value-added, exports and imports.

The legal unit of interest is the enterprise which can be further decomposed into branches. In the Eurostat definition, an enterprise is “the smallest combination of legal units that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources”. A branch corresponds to a legal unit depending on a controlling legal unit. For each branch, the Ultimate Controlling Institutional unit and its territory are identified. This UCI unit is the legal unit that has a direct or indirect control over the affiliates, i.e. when the UCI unit owns more than 50% of the affiliate or it has control over an affiliate controlling this affiliate (indirect control). If the UCI unit is on the domestic territory, then the branch is domestic, otherwise the branch is foreign controlled. In the Eurostat framework, the identification of the UCI is not sufficient, and whether the UCI is solely a *Global Group Head* or a *Global Decision Centre* or both should be determined (Eurostat, 2012).

The Global Group Head is not necessarily the entity in which the strategic decisions are taken. Let take the example of the semiconductor company STMicroelectronics. STMicroelectronics NV, domiciled in Switzerland, is the Global Ultimate Owner (GUO) of the group which has affiliates around the world. The French and Italian governments are its most important shareholders: they both own 50% of STMicroelectronics Holding domiciled in the Netherlands which itself owns 27.5% of STMicroelectronics. In this case, the Global Decision Centre is in Switzerland although the Global Group Head is in the Netherlands.¹⁷

17. This information comes from the Orbis and Factset databases.

When data are broken down by industry, the identification of the institutional unit (the establishment or the branch) becomes more important. For example, a car manufacturer may also propose repair services through an affiliate that should not be classified in the automotive industry but in the service sector ‘repair of motor vehicles’. For the United States, the BEA explicitly identifies the primary industry of the controlling unit, as well as the primary industry of its affiliates, according to the type of products with the highest share in sales (BEA, 2014). Unfortunately, this information is not always available for all countries. The EU FATS Regulation recommends that AMNE data be compiled on an enterprise basis (but sometimes the UCI can be a natural person).

The methodology to compile statistics depends on whether the AMNE statistics are *inward* or *outward*. The *inward* statistics correspond to the foreign-owned firms having a production activity in the reporter country – the hosting economy. On the contrary, *outward* statistics correspond to the reporter’s affiliates operating abroad: in this case, the reporter country is the country of ownership. Inward statistics generally come from Structural Business Surveys (SBS) with additional information from other data sources in order to evaluate the ownership status of the firm (OECD, 2017a). The variables from the AMNE data therefore come from the balanced sheets and the income statements reported as in the Generally Accepted Accounting Principle (GAAP). For *outward* statistics, further surveys are required. For example, in France, a survey is conducted on a sample of around 2 500 firms that fulfil threshold criteria. This is the reason why we rely more on the inward statistics than on the outward statistics that are generally based on a sub-sample of firms. There are country notes indicating precisely the source for each country in the OECD AMNE database (OECD, 2017). Table A.2 lists the other sources we have used in addition to the OECD data.

Table A2. List of data sources for AMNE statistics

Database	Industry classification	Year coverage	Flow
OECD Activities of MultiNational Enterprises	ISIC rev. 4	2008 - 2014	Inward and outward
	ISIC rev. 3	2000 - 2007	Inward and outward
U.S. Department of Commerce Bureau of Economic Analysis	NAICS	2000 - 2014	Inward and outward
Eurostat Foreign Affiliates statistics - FATS	NACE rev.1	2000 - 2007	Inward and outward
	NACE rev.2	2008 - 2014	Inward and outward
Compilation of firm-level data for Chinese input-output tables (data provided by Dr. Wang Zhi)	Chinese I-O classification	2002 - 2007	Inward
Statistics Canada CANSIM tables 03760151 and 03760152	NAICS	2010 - 2013	Inward
Trade by Enterprise Characteristics (TEC)	ISIC rev. 4	2011 - 2014	Inward
Australian Bureau of Statistics (ABS) Economic Activity of Foreign Owned Businesses in Australia	SISCA	2000 - 2001	Inward
Research Institute of Economy, Trade & Industry (RIETI) Foreign Direct Investment Database	RIETI classification	2000 - 2006	Outward
Orbis (firm-level data)	NACE rev.2	2007 - 2014	Inward