Paper prepared for the 25th International Input-Output Conference
June 19-23, 2017, Atlantic City, New Jersey, USA

AN ICIO SPLIT ACCORDING TO DOMESTIC AND FOREIGN OWNERSHIP: THE OECD TiVA-MNE PROJECT

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

OECD

ABSTRACT: The development of Inter-Country Input-Output tables has allowed researchers and policymakers to measure trade in value-added terms and to identify the contribution of each country and industry to the value of final products. However, the picture is still incomplete. Many of the firms involved in global value chains are multinational enterprises (MNEs) who have a network of foreign affiliates resulting from their foreign investment. In this paper, we present ongoing work at the OECD aiming at adding an ownership dimension to the new WIOTs from the WIOD project (Timmer et al., 2016). The work first consists in the estimation of a full matrix of the output of foreign affiliates in 43 countries plus the 'rest of the world'. 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 ICIOs and the missing information is estimated by various statistical ways. In addition to this matrix describing world output by country, industry and country of ownership, a trade matrix is created for the exports of domestic-owned and foreign-owned firms, as well as a matrix of value-added ratios. Based on these three matrices, the WIOD ICIOs have been split according to domestic and foreign ownership within each country and industry. In order to create transactions among domestic- and foreign-owned firms, we use a methodology derived from the regional I-O literature. Adding an ownership dimension allows us to revisit TiVA statistics to show the contribution of foreign-owned firms to exports and domestic sales.

I. Introduction

The development of Inter-Country Input-Output (ICIO) tables has been a big step forward to understand global value chains. The global Leontief inverse has 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. Global value chains are not only composed of independent companies exporting and importing intermediate and final products. Many of the firms involved in global value chains 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 EORA[1], 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 profit) or through spillover effects.

The TiVA-MNE project was launched at OECD in 2016 with a double objective. First, in order to better understand the role of MNEs in global value chains, the project aims at creating 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 objective of the project is then 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 -at this stage- only split between domestic-owned firms 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 rely on 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 China. Fetzer 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.

[1] TiVA is a database from 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 (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.

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 country. 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).

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 global value chains 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 TiVA-MNE project 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 recently 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 41 industries, with still lines of zeroes and more aggregated figures for some countries where not all the industry detail is available.[2]

The paper is organised as follows. Section II presents some of the data challenges and the main steps in the creation of the split ICIO. Section III 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 IV summarises the methodology for

[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.

splitting the ICIO itself using the output, trade and value-added matrices by ownership. Section V concludes. Data sources are detailed in the Annex.

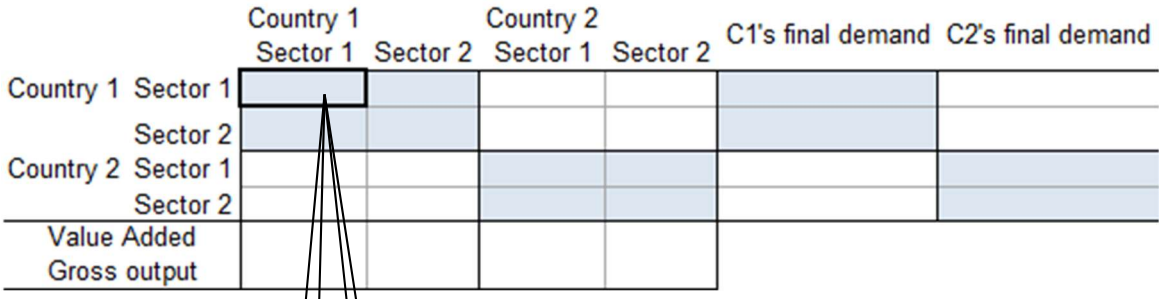
We do not present results in this paper but some results can be found in another paper where we also introduce a new conceptual framework that allows to trace value-added and double counting in domestic sales and exports of domestic-owned and foreign-owned firms (Miroudot and Ye, 2017).

II. Splitting the ICIO according to ownership: data challenges and main steps

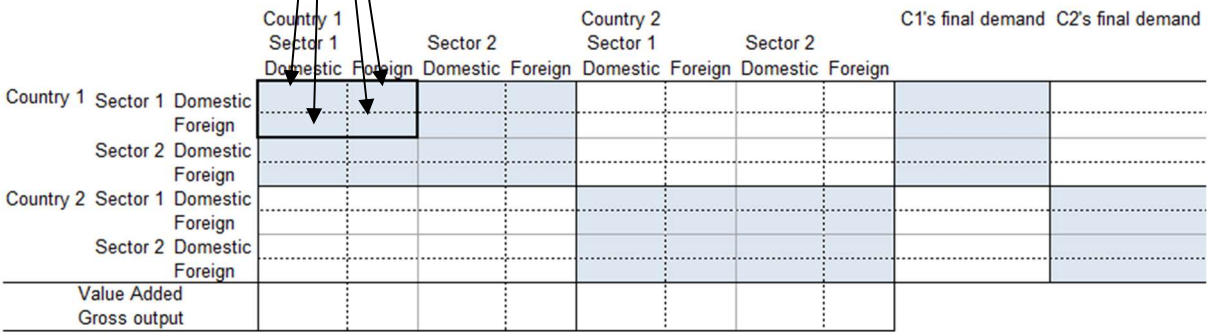
There are four main elements in the ICIO table: the intermediate consumption matrix, the final demand matrix, the value-added vector and the gross output vector. Cells across columns correspond to a country/sector's inputs; cells across lines correspond to the output of a country/sector. Figure 1 illustrates 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.

Figure 1 – Illustration of ICIO splitting according to ownership

Panel A



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 2 types of firms (domestic-owned and foreign-owned) the intermediate consumption matrix reports more than 24 million transactions, as opposed to "only" 6 million in the initial ICIO.

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.

Available information

In a perfect world, we would have directly information on all the transactions of domestic-owned and foreign-owned firms within national accounts and we could build the split 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 TiVA-MNE tables rely on two main sources of information: (1) the underlying World Input-Output Database (WIOD) -which is providing the whole structure of the ICIO when not distinguishing domestic-owned and foreign-owned firms and that we do not change-, and (2) a database internally constructed at the OECD from different data sources on the Activities of MultiNational Enterprises (AMNEs).

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 (see 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 41 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.

Our database on activities of multinational enterprises includes information on sales, output, value-added and exports of foreign-owned and domestic-owned firms in each country and industry. It is compiled by gathering information from various sources, mainly from the OECD AMNE statistics but also from the 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. The criteria 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.

Methodology

The methodology consists in two parts (Table 1). The first part is about building three balanced matrices of world output, value-added and trade (exports and imports) according to the ownership of firms. 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.

Table 1. Steps for the creation of the TiVA-MNE ICIO tables

III		Reconciling AMNE data with the ICIO framework: output, value-added and trade
III.a		Initial output matrix
III.b		Estimation of missing values
III.c		Balanced bilateral output matrix
III.d		Value-added and trade matrices
IV		Splitting and balancing of the ICIO according to ownership

III. Reconciling AMNE data with the ICIO framework: output, value-added and trade

In 2014, whereas the gross output of English manufacture of chemicals and pharmaceutical products is 67 billion USD in the AMNE database, it reaches 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 account:

- 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 is the 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: 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 sometimes 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. Unfortunately, we could not find any information on the type of unit surveyed in the AMNE database. 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 at the product level. Differences may appear with WIOD where trade is at the industry level (but the underlying WIOD data in this case provide figures by product and it can be reconciled).
- Lastly, most of the AMNE data that we have at 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.

In order to reconcile the AMNE database with the ICIO framework, there are several adjustments that have to be made.

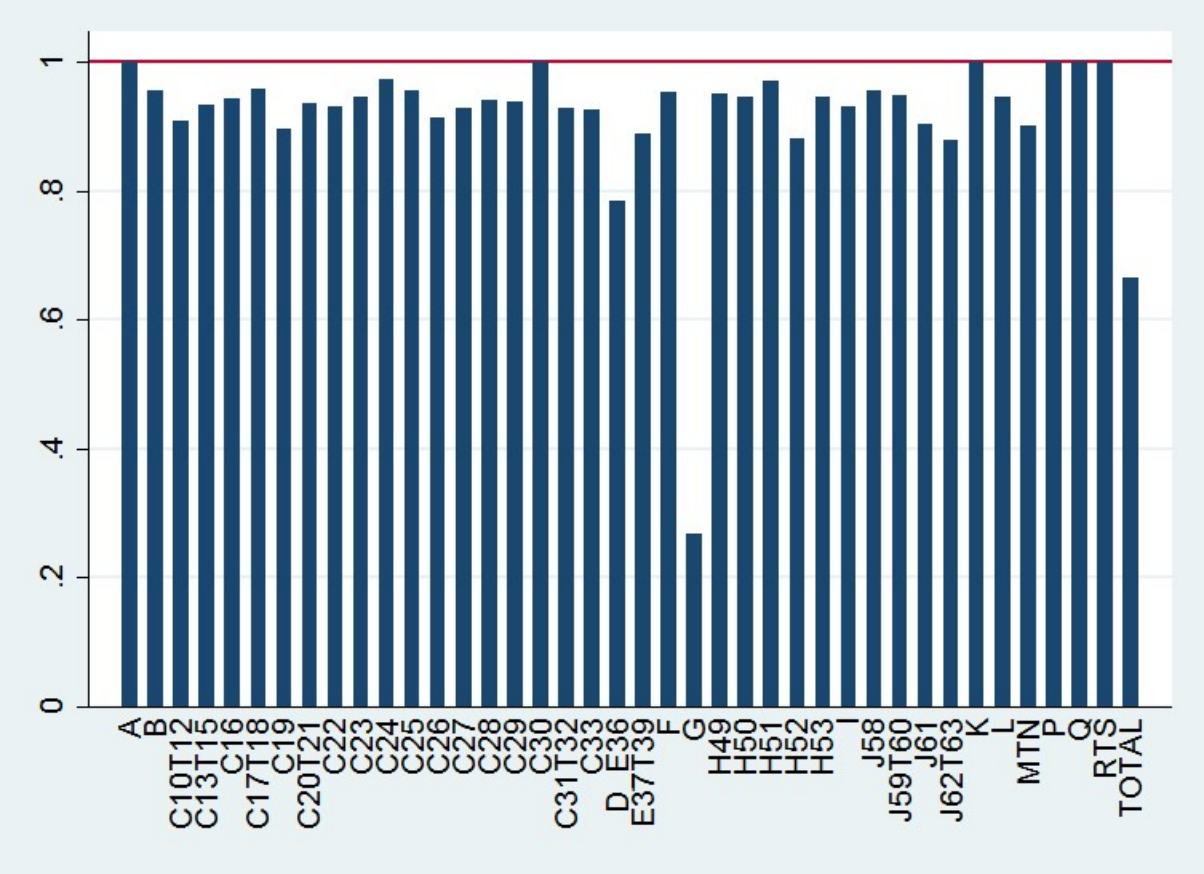
a) Initial output matrix

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. It involves in some cases splitting the data to estimate values for sub-sectors. For example, the 'paper, publishing and printing' sector 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 in converting turnover data to output data (when only the turnover or sales are available). Turnover corresponds to the revenue of firms while gross output is their production. For example, we have only turnover in the US data from the Bureau of Economic Analysis. We adjust these 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.

Figure 2 – Adjustments on turnover data



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 to do 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 imagine indicators that would include the use of the country of ownership in their calculations (using a bilateral matrix of value-added coefficients for example).

At the year-country-industry level, the number of observations we have for foreign owned firms is 9,908 out of 27,060 potential data points for the full matrix. At the year-country-

partner-industry level, the number of observations is 333,000 out of 1.2 million for the full matrix. The next step consists in estimating all these missing values.

b) Estimation of missing values

Many data points are missing and we use several methods to fill the gaps. First, many "missing values" are in fact zeroes that we cannot well identify in the AMNE statistics (where we do not always have information on what is missing or confidential as opposed to what is simply zero and for which no value is reported). In order to identify these zeroes, we use a first methodology based on foreign direct investment (FDI) data. For various reasons, we cannot infer sales of foreign affiliates from the value of foreign direct investment[3]. However, when there is no FDI at all in a given sector for a specific parent country (and this across several years) we can safely assume that no foreign affiliate was established and that sales are zero. Alternatively, for values that are unambiguously missing and above zero, we use estimates from a gravity equation. The rest of this sub-section details the estimation method.

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 affiliates sales.

The theoretical and empirical underpinning of the econometric extrapolation is the framework developed by Bergstrand and Egger (2007). Their model extend 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 MNEs' 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)

[3] 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-Yrkko 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 risk 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 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.

extend the literature to incorporate export-platform FDI when multinationals invest horizontally but with the objective of also serving third markets demand through exports.

Foreign affiliate sales from the AMNE database are used as dependent variable.[4] For the gravity set, we consider any AMNE data point available and we use mirror when the reporter's value is not available, we also use the adjusted turnover when output is not available. Then, we consider that there is zero AMNE sales when there is no investment (as previously explained, we fill the missing values with FDI data when FDI data are equal to zero).

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).

First, 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.[5] Data on GDP come from the World Bank Development Indicators.

Second, if the costs for setting up a foreign affiliate is lower than those of exporting (e.g. variable or fixed trade costs), for the multinational would be more profitable to invest in the location country. We thus include various empirical proxies for bilateral trade costs typical of the gravity literature, such as the weighted 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) and a dummy variable that equals one for country pairs with a common legal origin (comleg). These variables come from the CEPII database (Mayer and Zignago, 2011).

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:

[4] When the AMNE inward sales are not available the outward sales are used as mirror data.

[5] It would be more correct to use a measure of sectoral output in order to account for countries comparative advantages in certain sectors which translate in investments abroad in those sectors. However, the STAN database does not provide output data for the full set of countries in our database.

$$y_{ijst} = \alpha_0 + \beta_1 \ln(\text{distance})_{ij} + \beta_2 \ln(GDP_{it}) + \beta_3 \ln(GDP_{jt}) + \beta_4 \text{contig}_{ij} + \beta_5 \text{comlang}_{ij} + \beta_6 \text{colony}_{ij} + \beta_7 \text{comleg} + \delta_i + \delta_j + \delta_s + \delta_t + \varepsilon_{ijst},$$

where *i* stands for location country, *j* for controlling country, *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 84 host and source countries, spanning the years from 2000 to 2014 with 5% of zero observations. 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 on what 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.[6] 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 on log of sales of foreign affiliates are presented in Table 2, using the PPML estimator. The coefficients are all significant and with the expected sign. The R-squared is at 85%. 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 consistent with the proliferation of global value chains, 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 bilateral trade and investments costs, such as having a border in common, a common language and being once in a colonial relationship, have all the expected positive signs.

[6] From http://personal.lse.ac.uk/tenreyro/LGW.html

Table 2 - Econometric results of the regression on foreign affiliate sales

	Output
log(Distance)	-0.511***
	(0.0210)
log(GDP) (host country)	0.567***
	(0.0775)
log(GDP) (country of ownership)	0.556***
	(0.141)
Contiguity dummy	0.194***
	(0.0444)
Common official language dummy	0.137***
	(0.0521)
Colonial relationship dummy	0.231***
	(0.0376)
Common legal origin dummy	0.391***
	(0.0275)
Observations	1,413,041
FE	country, partner, industry, year

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

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.[7]

c) Balanced bilateral output matrix

With the AMNE data and the estimates made in the previous section, we now have 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.

The next step consists in balancing this matrix and make it consistent with the vector of output found in the WIOD database. To do this, 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

[7] 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.

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 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.

While the first objective function focuses on the output of foreign-owned firms, it is also an objective in terms of the output of domestic-owned firms as this output is just the difference between total output (which is fixed under the constraint of WIOD data) and the output of foreign-owned firms. From the AMNE data, we have to assess how output is split between domestic-owned firms and foreign-owned firms even if no information is available on total output or the output of domestic-owned firms. The difficulty in this exercise is that, as previously emphasised, output by country and by industry in AMNE statistics and in national accounts are not always consistent. For countries where we have both data, we find sometimes large discrepancies. Therefore, when we only have the output of foreign-owned firms, it is not clear that subtracting this AMNE total from the WIOD output figure is a good approximation for the output of domestic-owned firms.

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

	Estimation method	Number of cases
A	Inward adjusted output	9,878
B	Inward adjusted turnover	11,020
C	Inward output	11,274
D	Inward turnover	11,957
E	Inward adjusted output from aggregate sector	12,843
F	Interpolation & moving average	15,149
G	Sum of bilateral flows	26,815
H & I	Regression & zeros	27,720

As a consequence, we use a variety of methods to create the matrix for the first objective function (total foreign output by country and by industry). The values are obtained through the process described in Table 2 with A being the most straightforward and accurate estimation and other methods being used and described as B, C, D, etc. in descending order in terms of the quality of the estimate.

In the case of methods A and B, we have domestic and foreign values for output (or turnover) and we can apply these shares to the WIOD output figures (with turnover converted to output in the case of B). Since we know how different is the output by industry measured in AMNE statistics and in WIOD, we can rescale the AMNE figure and work with a share. The only assumption is that what differs between AMNE statistics and WIOD affects proportionally domestic and foreign-owned firms.

For C and D, we do not have information on total output in the AMNE statistics (i.e. on both the output of domestic-owned and foreign-owned firms). In the case of C, we use only the output of foreign-owned firms without any adjustment to match the WIOD output (since we

have no information allowing us to rescale). In the case of D, we use the turnover value adjusted to reflect the output (but not rescaled either).

In this first objective matrix, all data are by country and by industry. There are also gaps in the AMNE statistics for specific industries. Method E corresponds to cases where we can apply a methodology similar to A (or B, C and D in some cases) but where we have to work with data at a more aggregated level in terms of industries. We also have gaps in the AMNE data when it comes to specific years. In this case (F), we use interpolation techniques or a moving average (to extrapolate or when the number of years with data does not allow a robust interpolation) on the share of output by foreign-owned firms that we apply to the WIOD output.

Lastly, when no information at all is available for the output of foreign firms in a given country and industry, we use the sum of bilateral flows (method G). This sum comes from the inward values of the reporter, from mirror values of outward reporters and from the gravity dataset. In a few cases (and generally for small countries such as Luxemburg), values from the sum of bilateral flows are larger than the WIOD output. In this case, we estimate the values with a regression on the ratio of industries and country fixed effects (method H). We have also included there cases where the output from WIOD is equal to zero or where the industry is the public administrative sector with no foreign-ownership by definition (case I).

For the second objective matrix with bilateral totals by country (of output) and country of ownership, 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.

Table 3. Estimation of total output by country and by country of ownership

	Estimation method	Number of cases
A	Inward output	8,425
B	Inward turnover	10,753
C	Outward output & turnover	15,312
D	Inward output & turnover from aggregate sector	15,312
E	Bilateral flow from gravity equation	13,068

Bilateral totals are either available from the inward AMNE statistics (A and B) of from outward statistics (C). And we also have to deal with missing industries (case D). When no information is available for a given country pair, gravity estimates are used (case E). 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.

Table 2 and Table 3 have established a hierarchy among estimates with additional trust or accuracy in the methods listed first. This information is used to build a confidence index which 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 when it comes to their deviation from initial values in the objective function. The minimisation process will therefore change less these values 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.

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:

$$\min_{\hat{v}} o(\hat{v}) = \sum_{i,j,k} \gamma_{i,j,k} * \left(v_{i,j,k}^{starting} - \hat{v}_{i,j,k}\right)^2 + 10 * \sum_{i,j} \gamma_{i,j} * \left(v_{i,j}^{bil.obj.} - \sum_{k} \hat{v}_{i,j,k}\right)^2 + 100 * \sum_{i,k} \gamma_{i,k} * \left(v_{i,k}^{ind.obj.} - \sum_{j} \hat{v}_{i,j,k}\right)^2$$

$$s.t. \sum_{j} \hat{v}_{i,j,k} = output_{i,k}^{WIOD}$$

where γ are the respective confidence indices, v are the values from the starting and objective matrices and $\hat{v}$ are the 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).

d) Value-added and trade matrices

Once the bilateral output matrix is created, the following step consists in producing value-added and trade vectors (exports and imports) also split according to ownership and consistent with WIOD data.

For these three matrices, we use the same methodology which consists in applying the difference between foreign firms and domestic 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 and foreign value-added.

$$v = v_d + v_f$$

Second, value-added can be expressed as the value-added per output unit times output.

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

Where v corresponds to value-added at sectoral level, v_d is domestic value-added, x_d is domestic output and the subscript f applies to foreign firms for each variable. Now let's 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:

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

Domestic value-added can be estimated as:

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

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).

IV. Splitting and balancing the ICIO according to ownership

We now have four matrices (output, value added, imports and exports) that include information on domestic-owned and foreign-owned firms and the WIOD ICIO. 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 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 4 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:

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

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, $\mathrm{Y}_{\mathrm{ij}}^{\mathrm{D}}$ and $\mathrm{Y}_{\mathrm{ij}}^{\mathrm{F}}$, 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:

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

$$\mathrm{Y}_{\mathrm{ij}}^{\mathrm{D}} + \mathrm{Y}_{\mathrm{ij}}^{\mathrm{F}} = \mathrm{Y}_{\mathrm{ij}}$$

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

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

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

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:

$$Min\ S=\sum_{i,j}\frac{(Z_{ij}^{DD}-Z0_{ij}^{DD})^2}{Z0_{ij}^{DD}}+\sum_{i,j}\frac{(Z_{ij}^{DF}-Z0_{ij}^{DF})^2}{Z0_{ij}^{DF}}+\sum_{i,j}\frac{(Z_{ij}^{FD}-Z0_{ij}^{FD})^2}{Z0_{ij}^{FD}}+\sum_{i,j}\frac{(Z_{ij}^{FF}-Z0_{ij}^{FF})^2}{Z0_{ij}^{FF}}$$
$$+\sum_{i,j}\frac{(Y_{ij}^{D}-Y0_{ij}^{D})^2}{Y0_{ij}^{D}}+\sum_{i,j}\frac{(Y_{ij}^{F}-Y0_{ij}^{F})^2}{Y0_{ij}^{F}}+100*(\sum_{i}\frac{(V_{i}^{D}-V0_{i}^{D})^2}{V0_{i}^{D}}+\sum_{i}\frac{(V_{i}^{F}-V0_{i}^{F})^2}{V0_{i}^{F}})$$
$$+100*(\sum_{i}\frac{(E_{i}^{D}-E_{i}^{D\bullet})^2}{E_{i}^{D\bullet}}+\sum_{i}\frac{(E_{i}^{F}-E_{i}^{F\bullet})^2}{E_{i}^{F\bullet}}+\sum_{i}\frac{(M_{i}^{D}-M_{i}^{D\bullet})^2}{M_{i}^{D\bullet}}+\sum_{i}\frac{(M_{i}^{F}-M_{i}^{F\bullet})^2}{M_{i}^{F\bullet}})$$

where $E_i^{D^*}$, $E_i^{F^*}$, $M_i^{D^*}$ and $M_i^{D^*}$ 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.

V. Concluding remarks

This paper has explained the different steps involved in the creation of an ICIO split according to the ownership of firms. We are still experimenting with the methodology and at this stage we do not present detailed results as we plan to introduce further refinements in the data before proceeding to some analytical work. Some preliminary calculations on the share of domestic-owned and foreign-owned value-added in domestic sales and exports are presented in Miroudot and Ye (2017), as well as a methodology to trace value-added and double counting in domestic sales (to complement the existing literature focusing on value-added in gross exports).

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.

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.

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.

BEA (2014), *U.S. International Economic Accounts: Concepts and Methods*, Bureau of Economic Analysis, U.S. Department of Commerce.

Bekkers, E. and I. Girgzdyte (2015), "Sectoral Determinants of Foreign Affiliate Sales Employing European Data", mimeo.

Bergstand, 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.

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".

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.

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 EU (Timmer *et al.*, 2016). The 2016 update covers 43 countries[8] 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 TiVA-MNE project.

Table A.1. Classification of industries

Code	*Label*	*TiVA-MNE*
A01	Crop and animal production, hunting and related service activities	A
A02	Forestry and logging	A
A03	Fishing and aquaculture	A
B	Mining and quarrying	B
C10-C12	Manufacture of food products, beverages and tobacco products	C10-C12
C13-C15	Manufacture of textiles, wearing apparel and leather products	C13-C15
C16	Manufacture of wood and of products of wood and cork, except furniture; etc.	C16
C17	Manufacture of paper and paper products	C17
C18	Printing and reproduction of recorded media	C18
C19	Manufacture of coke and refined petroleum products	C19
C20	Manufacture of chemicals and chemical products	C20
C21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	C21
C22	Manufacture of rubber and plastic products	C22
C23	Manufacture of other non-metallic mineral products	C23
C24	Manufacture of basic metals	C24
C25	Manufacture of fabricated metal products, except machinery and equipment	C25
C26	Manufacture of computer, electronic and optical products	C26
C27	Manufacture of electrical equipment	C27
C28	Manufacture of machinery and equipment n.e.c.	C28
C29	Manufacture of motor vehicles, trailers and semi-trailers	C29
C30	Manufacture of other transport equipment	C30
C31_C32	Manufacture of furniture; other manufacturing	C31_C32

[8] 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, Russia, Slovak Republic, Slovenia, Sweden, Turkey, Chinese Taipei and United States.

C33	Repair and installation of machinery and equipment	C33
D	Electricity, gas, steam and air conditioning supply	D_E36
E36	Water collection, treatment and supply	D_E36
E37-E39	Sewerage; waste collection, treatment and disposal activities; materials recovery; etc.	E37-E39
F	Construction	F
G45	Wholesale and retail trade and repair of motor vehicles and motorcycles	G
G46	Wholesale trade, except of motor vehicles and motorcycles	G
G47	Retail trade, except of motor vehicles and motorcycles	G
H49	Land transport and transport via pipelines	H49
H50	Water transport	H50
H51	Air transport	H51
H52	Warehousing and support activities for transportation	H52
H53	Postal and courier activities	H53
I	Accommodation and food service activities	I
J58	Publishing activities	J58
J59_J60	Motion picture, video and television programme production, sound recording and music publishing activities; etc.	J59_J60
J61	Telecommunications	J61
J62_J63	Computer programming, consultancy and related activities; information service activities	J62_J63
K64	Financial service activities, except insurance and pension funding	K
K65	Insurance, reinsurance and pension funding, except compulsory social security	K
K66	Activities auxiliary to financial services and insurance activities	K
L	Real estate activities	L
M69_M70	Legal and accounting activities; activities of head offices; management consultancy activities	M_N
M71	Architectural and engineering activities; technical testing and analysis	M_N
M72	Scientific research and development	M_N
M73	Advertising and market research	M_N
M74_M75	Other professional, scientific and technical activities; veterinary activities	M_N
N	Rental and leasing activities, Employment activities, Travel services, security and services to buildings	M_N
O	Public administration and defence; compulsory social security	O_U
P	Education	P
Q	Human health and social work activities	Q
R-S	Creative, Arts, Sports, Recreation and entertainment activities and all other personal service activities	R_S
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	T
U	Activities of extra-territorial organisations and bodies	U

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's 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.[9]

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. AMNE statistics are provided on an establishment basis (BEA, 2014). Unfortunately, no further indications are given for other countries.

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, 2017). The variables from the AMNE data therefore come from the

9 This information comes from the Orbis and Factset databases.

balanced sheets and the incomes 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 some 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 A.2. 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