AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Global Trade Analysis Project
https://www.gtap.agecon.purdue.edu/

This paper is from the
GTAP Annual Conference on Global Economic Analysis
https://www.gtap.agecon.purdue.edu/events/conferences/default.asp

A Global Database of Foreign Affiliate Sales

Tani Fukui*[1] and Csilla Lakatos†[2,1]

[1] *US International Trade Commission*
[2] *Center for Global Trade Analysis, Purdue University*

July 2012

Abstract

There is a severe lack of data describing foreign affiliate activity. To fill this gap, we produce a new dataset to further the literature on the behavior of multinational firms. Eurostat's Foreign Affiliate Statistics database, with a large number of sector-level, bilateral observations on foreign affiliate sales, provides a basis from which to extrapolate the relationship between various host and source country factors and the foreign affiliate activity produced by them. This paper exploits the detailed level of the data by introducing sector-specific variables that in turn permit out of sample predictions. Further, the large number of excess zeros in the Eurostat dataset presents added complexity and is addressed using techniques borrowed from the trade literature, which also experiences a "zeros" problem. The dataset produced in this paper also serves as an input into the GTAP-based FDI model of Lakatos and Fukui (2012). This model integrates the foreign affiliate sales dataset produced in this paper into a framework that permits the analysis of the behavior of foreign affiliates within the context of a general equilibrium model.

*Corresponding author. Tani Fukui (tani.fukui@usitc.gov) is an economist at the U.S. International Trade Commission from the Office of Economics. These views are strictly those of the authors and do not represent the opinions of the U.S. International Trade Commission or of any of its Commissioners.

†Csilla Lakatos is Research Economist at the Center for Global Trade Analysis at Purdue University and Visiting Fellow at the U.S. International Trade Commission.

1 Introduction

The examination of foreign affiliate operating activity is a relatively new branch of the literature, owing primarily to the paucity of data. Foreign direct investment (FDI) statistics are collected by numerous countries, but these do not provide a complete picture of the activities of multinational enterprises. In particular, FDI examines only the international transfer of funds rather than their operations. Without data on operations of multinationals, it is difficult to assess the effect of policy changes on foreign affiliate activity. As foreign affiliate activity grows in importance, this lack of data is slowly being addressed, and research is able to move forward. In particular, the establishment of Eurostat's Foreign Affiliate Statistics database (Eurostat, 2012) provides a much needed boost for this area of research. Eurostat provides a large amount of data on foreign affiliate activity, rather than data only on investment stocks or flows. In this paper we use the Eurostat dataset to estimate the behavior of foreign affiliate sales as a basis. It implements an econometric model consistent with the branch of the literature that originated in Markusen et al. (1996) and Markusen (1997) and that includes Bergstrand and Egger (2007) and Carr et al. (2001). Finally, we apply quadratic optimization techniques to compute the final database.

Blonigen (2005) provides a comprehensive review of the recent literature on FDI determinants. He concludes that the broad-based relationships between FDI and policies have been difficult to discern.[1] More importantly, he assesses that as FDI research progresses, it will continue to be thwarted in its search for overarching relationships, primarily because the reasons for which firms invest abroad are many and varied.

The economic literature on the drivers of FDI identifies two main types of investment rationales: market access (selling to consumers in the host market) and efficiency seeking (searching for low cost production sources). In addition, the proliferation of global supply chains has led to variations on each of these themes, so that goods (and to a lesser extent services) pass through multiple countries with final consumption sometimes taking place in one of the production countries, so that both efficiency seeking and market access motivate the foreign investment.

This heterogeneity can best be addressed by examining the matter at a more detailed level—honing in on particular sectors or countries, in which the investment rationale may be more uniform. As a result, the literature has increasingly gone the way of firm-level analysis, which permits the researcher to control more tightly by type of investment rationale. Despite

[1] The study examines investment stocks and flows as well as operations of multinationals.

this trend, we follow the literature in examining macro-level FDI statistics. However, in many cases, such as for the project we have taken on, it is necessary to make some assessment of overall macroeconomic behavior, although it may simply be a rough approximation of true firm behavior. Firm-specific effects cannot hope to provide approximations of macro-level activity, as well as a matter of practicality in attempting to estimate these effects for a large number of countries.

A problem presented by this dataset is the existence of a large number of missing values. This is a problem that has not been extensively addressed in the FDI literature. On the other hand, it has been addressed in the trade literature, which also has such problems. We integrate some approaches of that literature in our estimation strategy, in particular, the Poisson Pseudo Maximum Likelihood (PPML) proposed by Silva and Tenreyro (2006) and the zero inflated models discussed in De Benedictis and Taglioni (2011). Finally, there has been very little use of sector specific data in foreign affiliate data research, largely because it is not usually available. We take advantage of this extra dimension in the model to attempt to estimate sector-specific differences in foreign affiliate activity using sector specific data.

In addition to the zeros problem, there is also a large number of missing values in the database that prevents the immediate use of these data in a global model. This is due both to confidentiality or missing values (so that source-host-sector points are not available in many cases), and also to the constrained set of countries in the database. The database documents foreign investment into European countries. An important purpose of the database is to apply the database to a version of the Global Trade Analysis Project (GTAP) computable general equilibrium (CGE) model that explicitly models FDI. In order to apply the database to the newly developed FDI model for GTAP, it is necessary to extrapolate to all regions and sectors used in the GTAP model. The coefficients generated from the econometric analysis in this paper will be used as a starting point for the extrapolation.

This paper is one of two papers produced in tandem to provide a rich modeling tool for policy analysis. Our goal is to construct a set of tools to model the behavior of foreign affiliates. In order to properly model this, we need two elements: a set of databases and a model. We first construct a set of three databases that enable the breakdown of "domestic" elements of the economy into foreign and domestic elements – in particular, foreign capital stocks, value added, and foreign affiliate sales. Then we feed these databases into a modified version of the standard comparative static GTAP model. This is a data-driven general equilibrium model that models the global economy at a detailed regional and sectoral level, using 129 countries and regions and 57 sectors. The main focus of this model is the modeling

of international trade and in particular modeling the effects of trade policies on the economic welfare on countries. We modify this model to explicitly take into account the existence of foreign owned capital and foreign affiliate activity. The construction of the databases is detailed in this chapter; the construction of the model and its policy implications are in Lakatos and Fukui (2012).

To our knowledge, there has been only one prior attempt, in Hanslow et al. (2000), to construct a large scale, bilateral by sector, fully consistent database of foreign affiliate statistics. The purpose of that database was, as with ours, to use it within a version of the GTAP model modified to include FDI. There are a few key differences between their estimation attempt and ours is as follows. Hanslow et al. (2000) used ratios of foreign affiliates data—total assets to FDI capital and sales to asset ratios—by sector, extracted from U.S. BEA data, and applied those ratios to FDI stocks reported by CEPII. Similar ratios were used for value added. In our method, we broaden the set of underlying countries to include all European countries reported in the Eurostat database (the full list of countries is in the appendix) rather than relying solely on U.S. data. In addition we estimate the effects using a fully specified econometric model which does indeed display significant differences across both host and source countries, as well as across sectors. The use of econometrics within this context, therefore, is new. In addition, due to improvements in data collection by Eurostat, it has become possible to examine the cost structure of foreign affiliates using value added and employment costs. Therefore, rather than relying on calculations of value added based on pro rata allocations from sales, we are able to directly estimate the labor and capital shares of value added.

In the second section we provide a discussion on the use of foreign affiliate operating data rather than the more commonly used foreign direct investment data. The third section provides the econometric approach, including background literature, specification, data and the results. The fourth section describes the quadratic optimization procedure. A fifth section presents elements of the final database. The final section concludes.

2 FDI versus data on foreign affiliate activity

Recent publication of foreign affiliate statistics permit us to use operating data directly to examine the activity of foreign affiliates. Until recently, the lack of available data on the activities of foreign affiliates has often compelled researchers to use FDI stocks as a proxy for foreign affiliate operations data. Numerous econometric studies rely heavily on FDI stocks

and flows data to investigate various aspects of multinational corporation (MNC) activity and their impact on host/home countries. Similarly, prior CGE studies use FDI stocks/flows data to disaggregate domestic and foreign firms in the underlying data (in most cases, the GTAP database) or to infer structural information about the production characteristics of foreign firms as well as their sales patterns.

FDI and foreign affiliate activity (FAA) data reflect different facets of the role of multinationals in the world economy. FDI statistics are a measure of the monetary value of the movements of capital between investors and affiliates and they are a component of the capital account of a country's balance of payments. By contrast, FAA cover data regarding overall operations of foreign affiliates such as sales, production, and employment.

Apart from this, there are fundamental differences between FAA and FDI data that limit the comparability of these statistics. First, from a methodological point of view FAA data cover all affiliates that are foreign controlled (investors with more than 50% of the voting rights) while FDI data comprise all foreign interests that correspond to 10% or more of the voting power. Second, FAA data are assigned to the region or sector of the Ultimate Controlling Institution (i.e. parent company) while FDI statistics are based on the immediate counterparty country, i.e. the country of the immediate investor/recipient even if the capital is passing through a third country (Eurostat, 2007).

Finally, FDI stocks are a biased measure of FAA in that they over- or underestimate the activity of multinationals as a function of host country characteristics (Beugelsdijk et al., 2010). FDI statistics measure only movements of capital between direct investors and their affiliates, and not funds from unaffiliated persons. This can lead to an underestimation of foreign affiliate activity in countries with well-developed financial markets.

In addition, FDI in countries that are tax havens generate no actual productive activity - leading to an overestimation of the activity of foreign affiliates in these countries. Practically, this measn that FAA data are less likely to be influenced by the existence of tax havens. This is seen in the U.S. Bureau of Economic Affairs (BEA) data, where high levels of U.S. investment abroad are seen in known tax havens such as the British Virgin Islands, but foreign affilaite sales are substantially lower.

Sectors that are capital intensive, such as mining, should see an overestimation of foreign affiliate sales, while distribution sectors - sectors that generate a large number of sales relative to capital - should have their FAA values underestimated.

In order to verify the significant differences between FAA and FDI, we construct a regression model that allows us to measure the extent to which foreign affiliate activity, in the

form of sales, and FDI statistics systematically vary across sectors and/or countries. Thus, we specify the following regression:

$$ln(FAS_{irst}) = \alpha_0 + \alpha_1 ln(FDI_{irst}) + \delta_i + \delta_r + \delta_s + \delta_t + \epsilon_{irst} \quad (1)$$

where FAS_{irst} describes foreign affiliates sales in sector i, in host country r of affiliates in country s in year t; FDI_{irst} represents FDI stocks in sector i, in host country r by country s and year t; finally $\delta_i, \delta_r, \delta_s$ and δ_t are sectoral, host, home country and year dummy variables, respectively. The data on foreign affiliate sales used in the estimation originate from OECD's Statistics on Measuring Globalisation and the data on foreign direct investment from OECD's database on International Direct Investment Statistics.

Table 1 reports regression estimates for the sample as a whole while estimates for the sectoral and host country dummy variables are plotted in Figure 1 and Figure 2 below. Although

Table 1: Regression estimates, FDI and FAS

	Log(FAS)
Log(FDI)	0.361***
	(0.000)
N	1705
adjR^2	0.895

Note: ***p<0.001

our results show a positive and significant relationship between FDI stocks and foreign affiliates sales, as depicted in Figure 1 and Figure 2 we also find important cross-country and cross-sectoral variation between the dependent and independent variables. For instance, FDI stocks tend overestimate affiliate sales in Slovakia, Greece, Slovenia and Finland, countries such as Japan, Germany, and France have overestimated FAS relative to other countries that have more developed financial institutions. With respect to sectoral mismatches, we find that FDI stocks data underestimates affiliate activity the most in Wholesale Trade, Sale and repair of motor vehicles, and motor vehicles. By contrast, it overestimates real estate activity, air transportation and mining and quarrying. This is expected: mining and quarring is capital intensive and frequently takes place in countries with relatively financial institutions. Wholesale and retail trade sectors, by contrast, have very high sales relative to capital investment.

These findings become particularly important with respect to existing CGE work that uses FDI stocks as a proxy to disaggregate the sales and other elements of foreign affiliate

activity as this method creates cross-country and cross-sectoral bias in the disaggregated data.

While FDI statistics might be considered to be an appropriate measure of the aggregate activity of foreign affiliates, our results show there are significant mismatches between sectoral and regional FAS and FDI statistics. In this context, our use of newly available foreign affiliate activity represents a substantial improvement over the use of FDI stocks as a proxy for the activities of foreign affiliates.

3 Econometric Estimation

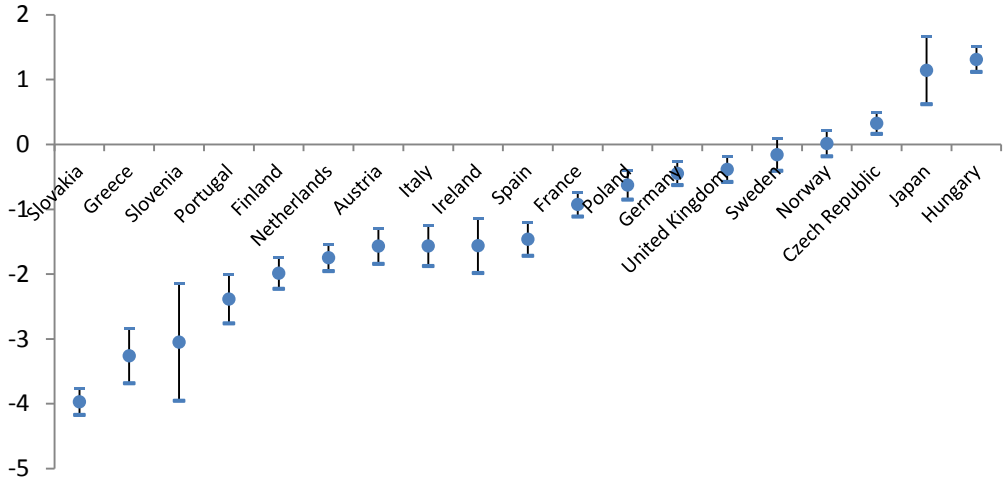
There is currently no global database of foreign affiliate sales. The closest such source available to us is the Eurostat database (Eurostat, 2012) which has detailed sectoral level foreign affiliate sales by source country for many European countries. In order to construct the required database, we first conduct an econometric analysis of the existing data to produce a set of coefficients that provide information about the relationship between various independent variables and foreign affiliate sales. These coefficients are then used to extrapolate to the full set of countries and sectors needed by the GTAP model.[2] Finally, the extrapolated dataset is merged with the known data: these data include the original Eurostat dataset as well as data from the OECD (OECD, 2010), the U.S. BEA (BEA, 2012), UNCTAD (UNCTAD, 2011) and the National Bureau of Statistics of China. Contradictory information among these data sources is resolved using an optimization procedure explained in detail in section 4.

There is a small but growing literature that has in recent years attempted to produce a well-formed model for the use of gravity-like models for FDI and foreign affiliate activity in the way that Anderson and Van Wincoop (2004) have pioneered for trade flows. The gravity model, frequently employed to explain trade flows, has also been employed to explain FDI. As with trade, the rationale for the gravity model began as a practical matter: the model "worked" in that it had a high degree of explanatory power, but the theoretical foundations were shaky or non-existent. In recent years, however, progress has been made in providing theoretical underpinnings to the model. These theories have naturally also produced modifications that are FDI-specific and warrant close attention.

The set of models described in Markusen and Maskus (2002) is one of few strands of

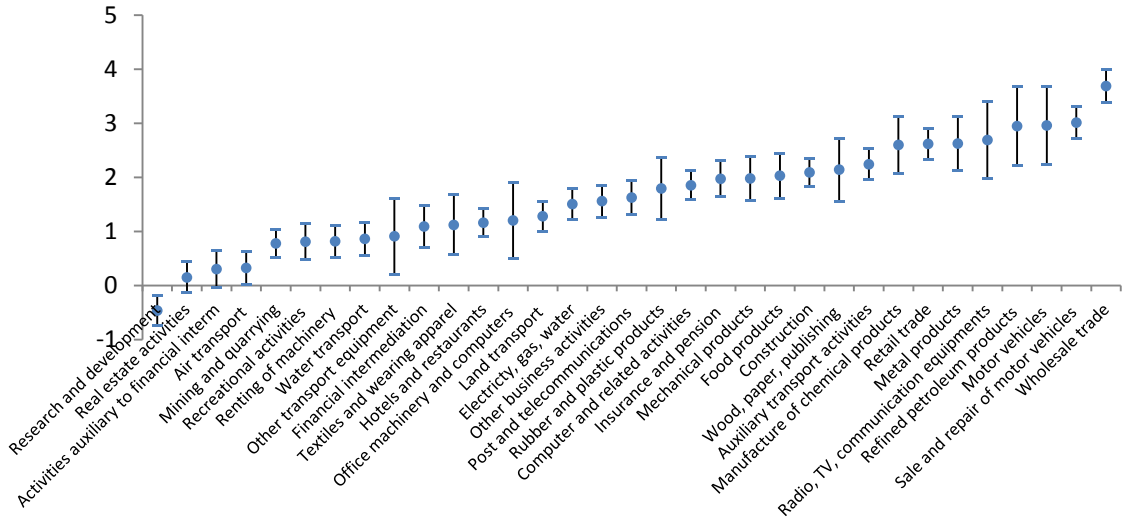
[2]Certain sectors are aggregated from the original GTAP model, including particularly the agriculture sectors.

Figure 1: Host-country variation - relative to the omitted dummy (USA)



Notes: Point estimates and 95% confidence interval plotted for each country

Figure 2: Cross-sectoral variation - relative to the omitted dummy (Agriculture)



Notes: Point estimates and 95% confidence interval plotted for each sector

literature to explicitly examine foreign affiliate sales rather than FDI. Kleinert and Toubal (2010) also present a model on foreign affiliate sales, lending further support to a gravity-type model. The original paper by Markusen discusses a 2 factor, 2 country, 2 good (2 x 2 x 2) knowledge capital model, whose main contribution is to delineate the difference between horizontal multinationals (those firms that establish subsidiaries abroad to sell in those markets) and vertical multinationals (those firms that establish subsidiaries abroad to reduce production costs).

In Carr et al. (2001), a horizontal and a vertical model are nested within the knowledge capital model in order to test whether one or the other is supported by the data. The results of these tests reject the vertical model, and cannot reject the horizontal. That is, at the aggregate level, the data demonstrate more horizontal than vertical characteristics. The data used are U.S.-associated values only (foreign affiliate sales), aggregated to the bilateral level. They do not have sector level data. Rather than an OLS model, they use WLS as well as a Tobit model. The main concern is heteroskedasticity because countries differ dramatically in size. The weights come from OLS residuals of the sum of GDP values. The Tobit regressions are conducted in order to address prevalence of zero values in the data.

Bergstrand and Egger (2007) uses an updated version of the model that advances this literature in a parallel way to the trade literature. This paper presents a 3 factor, 3 country, 2 good knowledge capital model that builds on Carr et al. (2001). The model in Bergstrand and Egger (2007) adds a third country: this permits the examination of third country effects on bilateral trade flows. That is, it attempts to examine whether the gravity relationships found in the trade literature also hold for foreign affiliate sales (and also for FDI). In particular, they attempt to examine essentially whether an Anderson and van Wincoop type effect is present, i.e. the multilateral resistance term. Most models in the FDI literature examine a two country model rather than a multi-country model which does not permit multilateral resistance terms.

In addition, they add a third factor (capital) that together with the third country produces complementarity between country size and the various trade variables (trade, foreign affiliate sales, and foreign direct investment). In the original 2x2x2 model of Carr et al. (2001), the national and multinational firms were mutually exclusive so that the existence of multinationals would mean that all single-country firms would cease to exist; this is counter to what is observed in the data.

Yeaple (2003) is a rare example in the literature of a paper that uses sector-specific data to distinguish FDI behavior. He uses U.S. BEA foreign affiliate sales data at the bilateral

and sectoral level. Yeaple uses the following sector specific information: transport costs (industry and host country specific), a measure of scale economies (industry specific), and a set of variables that reflect unit costs (industry and host country specific).[3]

3.1 Data and Econometric Specification

The model we use is based on a modified version of Bergstrand and Egger (2007) and Carr et al. (2001). These two papers use largely similar econometric specifications. We modify them in the following ways. First, based on the results presented in Bergstrand and Egger (2007)r, the FAS behaves similarly to FDI and so we replace the FDI with FAS. Second, we account for the sector specific nature of our data by replacing the GDP of host countries with the domestic production by sector. We follow Anderson and Van Wincoop (2004) in replacing GDP of the host country with domestic production.

In addition, further adjustments were made for pragmatic reasons. The econometric model specified by Bergstrand and Egger (2007) does not include per capita variables. As a result, the extrapolation of the model is strongly influenced by the size of countries, to the point that the vast majority of sales are projected to be sourced from and hosted by the largest country, the United States. By contrast, the data show greater variation due to per capita GDP of both host and source countries. In order to correct for this, we add a GDP per capita variable for both the host and source. Under this specification, the regressions produce results that, after extrapolation, are less biased by the size of country.

$$\begin{aligned} FAS_{irst} =& \alpha_0 + \beta_1 ln(GDP_{st}) + \beta_2 ln(GDProw_{rst}) + \beta_3 ln(Production_{irt}) \\ &+ \beta_4 ln(GDP/capita_{rt}) + \beta_5 ln(GDP/capita_{st}) + \beta_6 ln(distance_{rs}) \\ &+ \beta_7 CommonLanguage_{rs} + \beta_8 TradeOpenness_{rt} + \beta_9 FDIrestrict_{ir} \\ &+ \beta_{10} ln[(S/U)_{rt}/(S/U)_{st}] + \sum \gamma_t + \epsilon_{irst} \end{aligned} \tag{2}$$

The subscript i refers to sectors, r refers to host; s refers to source, and t refers to time. The model includes a full set of time dummies, γ_t. All independent variables are listed in Table 2, along with the data source used and summary statistics. The dependent variable, foreign affiliate sales, is discussed in the following section in greater detail.

GDP_{st} is the GDP of the source country. There is considerable variation in the GDP variables, despite the fact that the countries are predominantly European countries, reflecting

[3] Anderson and Van Wincoop (2004) also presents a sector level model, although it is to explain trade flows rather than foreign affiliate activity.

that both large and small countries are included in the sample. These data are from World Bank World Development Indicators.

GDP RoW is the GDP of the rest of the world, i.e. GDP of the world less GDP of both source and host countries' GDP. The variation of this variable is quite small, as the size of countries is generally dwarfed by the size of global GDP. These data are also from WDI Online.

Rather than GDP of host, we use domestic production, by both domestically- and foreign-owned firms, of individual sectors, *Prod*. The rationale is that countries have a comparative advantage in certain sectors and develop strong multinational firms in those sectors with transferable skills that in turn invest abroad. Domestic production shares are also included as host country variables to capture the effect of a country that has a pronounced comparative advantage that is not transferable. This is most explicit in natural resources, but may also be a factor in manufacturing industries, where countries specialize in specific manufacturing sectors.

This variable also has a large standard deviation, reflecting both varying sizes of countries and of sectors. These data are from Eurostat and correspond to the same sectors provided in the foreign affiliate sales database.

GDP per capita of both the source and the host countries are used. As with the other GDP data, these are from WDI Online. There is slightly more variation across source countries than across host countries as host countries are uniformly developed European countries, while source countries include both developing and developed countries.

Distance is the distance between source and host capital cities. *Comlang* is a binary variable that takes the value of 1 if source and host share at least one language. It is usually 0, taking on the value 1 in only a handful of cases. Both this and the distance variable were obtained from CEPII.

Trade openness is a measure of aggregate trade restrictiveness set up by the host country. This index is obtained from the Fraser Institute's Economic Freedom of the World report, which uses primarily quantifiable measures on a range of topics to measure a country's economic freedom. The trade index, "Freedom to Trade Internationally", takes into account total revenues from tariffs, mean tariffs and the variance of tariffs across tariff lines. It is clear from the summary statistics that the openness observations are dominated by European countries that have extremely low trade barriers. As a result, the lowest level of trade openness reported is quite high (6.8 out of a possible 10), and the average, at 8.5, represents something substantially close to free trade. There is little variation in this variable.

The *FDI restrictiveness* index was obtained for G20 countries using Koyama and Golub (2006). This is a sector and host specific restrictiveness index, which takes into account foreign ownership and other national treatment aspects of investment. This is a one-off measure of restrictivenss collected roughly at the time of the study with no time series variation.

The variable ΔSK is the skill difference between two countries: the ratio of skilled to unskilled workers in the source country less the same ratio for the host country:

$$\Delta SK_{rst} = \frac{SK_{rst}}{UST_{rst}} - \frac{SK_{rst}}{UST_{rst}}$$

where SK is skilled labor, defined as subclassification 1, 2, or 3 (legislators, senior officials and managers; professionals; and technicians and associate professionals) by the ILO.[4] This is a negative number at the mean, so that the average source country in our sample has less skilled workers (relative to its stock of unskilled workers) than the average host country. This is expected because all host countries are developed countries in the EU while source countries include both developed and developing countries. Countries that are in the source list but not in the host country list include China, Russia and Turkey.

[4]ILO.org's LABORSTA database. Labor force survey data were used for all countries: http://laborsta.ilo.org/

Table 2: Independent Variables

	Years available	Source	Dimension	Units*	Mean	Median	Min	Max	StDev
Foreign affiliate sales	2003-2007	Eurostat (FATS)	sector, source, host, date	$ million	140	0	0	54,100	1,090
GDP, source	through 2009	World Bank	source, date	$ billion	830	233	5	14,000	1,920
GDP RoW	through 2009	World Bank	source, host, date	$ billion	44,800	45,000	24,500	55,700	6,370
GDP per capita, source	through 2009	World Bank	source, date	$	25,013	23,682	1,731	82,294	16,160
GDP per capita, host	through 2009	World Bank	host, date	$	22,230	18,424	2,555	56,894	14,386
Domestic production, host	2007	Eurostat	host, sector	$ million	14,700	1,780	0.4	584,000	49,200
Distance	n.a.	CEPII	source, host	km	3,314	1,727	161	19,539	4,215
Common language (ethno)	n.a.	CEPII	source, host	0 or 1	0.03	0	0	1	0.16
Economic Freedom: Trade	1995-2008	EFW	host, date	1 to 10	8.5	8.5	6.8	9.8	0.6
FDI restrictiveness	2010	OECD (2010)	sector, host	0 to 1	0.02	0	0	1	0.1
Skill difference	1989-2008	ILO	source, host, date	skill ratio	-0.006	0.001	-0.39	0.29	0.1

*Units are as reported here for ease of notation; for the regressions we use whole dollar values (rather than millions, etc.) for all values. Note: Summary statistics include only those observations that were ultimately included in regressions. There were a total of 41,083 observations with a complete set of independent variables, including those which had zero foreign affiliate sales.

3.2 Foreign Affiliate Sales Data

The primary data source that we use in our analysis is Eurostat's data on foreign affiliates.[5] This is our set of dependent variables. The dataset contains 41 source and 22 host countries (see appendix Tables 12 and 13 for a complete list of countries). The host countries are the reporting countries, and are all European. The source countries are mostly European, with some non-European developed countires and a handful of developing countries. The database provides "three dimensional" data: foreign affiliate sales by source country, host country, and sector. A total of 117 sectors and subsectors are covered in the original database, which includes sectors and their disaggregated subsectors. Only a relatively small subset of 21 sectors was selected—this is both because of lack of the corresponding sectoral data of an independent variables, domestic production, and to more closely match the targeted GTAP sectors. The database spans the years 2003 to 2007.

The dependent variable is foreign affiliate sales. This is the total sales reported by foreign affiliates and includes local sales as well as exports out of the host country. These are taken from the Foreign Affiliates Trade Statistics database produced by Eurostat. The database has a large number of gaps (see Table 3).

Table 3: Foreign Affiliate Sales Observations

Type	No. Observations	Share
Missing	76,703	48%
Zero	74,087	46%
Positive	10,325	6%
Total	161,115	

Source: Eurostat FATS database, 2003-2007

This is partly because the Eurostat database is very ambitious: the database aimed to collect data on 117 sectors and subsectors, but very few countries reported on more than a small fraction of these sectors. Approximately 48 percent of all possible observations are missing. In addition, 46 percent of the possible observations are zero values: these are

[5] Variable $fats\_g1a\_03$ under the category "Foreign control of enterprises - breakdown by economic activity and a selection of controlling countries". Accessed May 17, 2011. Data are originally in Euros and presented throughout this paper in US dollars. These data are from the inward FATS data collection, so that host countries are the reporting countries.

either smaller than the threshold set by Eurostat (500,000 Euros) or actually reported as zero. The presence of these zeros means that the econometric specification must be carefully determined, as discussed in the econometrics section.

At the level of disaggregation we use, Eurostat reports $4.3 trillion in foreign affiliate sales in 2007. In 2003, the sales are only $1.5 trillion. However, due to the missing values problem this does not necessarily imply a 30 percent annual growth rate, but rather that the data collection and coverage have expanded over these years.

According to the raw data, approximately two thirds of foreign affiliate sales reported in the dataset takes place in three countries – Germany, the United Kingdom, and Italy. Sector level data is also highly concentrated, with nearly 80 percent reported by two sectors: 46 percent by wholesale and retail trade, and 33 percent in manufacturing. These shares are of course influenced by reporting bias – if these countries or sectors are more likely to be able to report their affiliate sales, then they are overrepresented in these aggregate totals.

Out of the $4.3 trillion in sales, only $1.7 trillion worth of observations is used in the regressions. This is largely due to the relative paucity of data on domestic production of hosts.[6] Further summary statistics for foreign affiliate sales are noted in the appendix.

3.3 Estimation Strategy

The large number of zero cells in the dataset calls into question the conventional strategy used in the FDI literature. Much of the literature on FDI uses OLS to estimate the relationship between FDI and the dependent variables. The log transformation commonly used in the OLS specification does not permit an explanation for zeros. More problematically, OLS does not model the decision to enter (or not enter) a market as a separate process but rather simply models zeros as part of a linear function.

The trade literature has examined this problem extensively, as trade data also tends to have a large number of zeros. In our estimation procedure, we implement both OLS and several other methods borrowed from the trade literature, modified to include FDI-relevant variables.

In addition, two possible problems have been pointed out by other researchers. Silva and Tenreyro (2006) propose the use of Poisson Pseudo Maximum Likelihood (PPML). The original purpose of this method was to address the pervasive heteroskedasticity in the

[6] When the database is constructed, the original $4.3 trillion worth of observations are used to reconstruct it.

gravity equations rather than specifically addressing excess zeros. However, the Poisson distribution does permit zeros to occur, allowing an explanation of the prevalence of zeros. They demonstrate that Poisson performs well under certain heterogeneity conditions.

Some arguments have been raised against the use of the PPML model. First, it tends to under-predicts the number of zeros; second that data are generally over-dispersed, in contrast to the assumptions of the PPML, which assumes that the mean and variance are equal. These arguments have been put forth in Martin and Pham (2008) and De Benedictis and Taglioni (2011). The latter has proposed other methods such as the zero inflated models ZIP (zero inflated Poisson) and ZINB (zero inflated negative binomial). Zero-inflated models are models that combine a logit model with a Poisson type model. As a result, there are two possible ways in which these models can generate a zero: first, under the logit portion of the model, which predicts a binary go/no go decision; and second under the main part of the model which, conditional on a "go" decision of the logit model, predicts the value of that decision. ZIP and ZINB behave similarly with the one difference that the ZINB does not force equality between mean and variance. Both sufficiently high fixed and variable costs may generate zero foreign affiliate sales. It should be noted that the mere existence of overdispersion does not require the selection of ZINB over ZIP. ZIP, by virtue of its two processes, does not assume an over-dispersed set of data.

An added complication is that reported zeros in the Eurostat database do not all mean zero. They may be either small positive values (less than 500,000 Euros) or true zeros. There is no way of distinguishing the two cases given the currently available data.

We use the PPML specification in our final database because the fitted values are substantially closer to the original data than under alternate specifications.

3.4 Econometric Results

The results of the econometric estimation are presented in Table 4. Each of the four results in the table use the same set of independent variables. The first column in Table 4 uses OLS, the second uses PPML, the third uses ZIP and the fourth column use ZINB.

According to Bergstrand and Egger (2007), the expected sign of GDP source is positive.[7]

[7]Note that Bergstrand and Egger (2007) models FDI, FAS, and trade. These three variables generally behave similarly, although the FAS variable is not described in as great detail as FDI or trade, and is not tested against the data. One difference in predictions of variable behavior is in the effect of transport and investment costs: lower transport costs increase trade and increase FDI; higher investment costs decrease trade and increase FDI, and presumably FAS behaves similarly to FDI if only in the sign of their comovement.

In our estimation, this is the case only in the PPML version in in Table 4, and even there it is not robust. As a result GDP of the source country appears, in the OLS, ZIP and ZINB versions of the estimation, to be negatively correlated with foreign affiliate sales. However, the GDP of the rest of the world (GDP RoW) has a large negative coefficient, which is an expected result. Because this variable is inversely related to the GDP of the source country, the net effect of these two coefficients is such that GDP of the source country is positively correlated with foreign affiliate sales. That is, the positive effect of GDP source and host are captured in the highly negative coefficient of GDP RoW.

The expected sign of GDP RoW is negative. As noted above, this is indeed the case. Note that this coefficient is particularly large (and negative) for PPML. That is, the joint size of host and source country are a particularly strong driver of activity according to the PPML estimate but somewhat less so for the other estimates.

Domestic production, ln(Production), is expected to be positive. This is one of two variables that are sector-specific (the other being FDI restrictiveness, FDI Restrict). These variables are indeed positive and strongly significant for each of the cases.

As expected, GDP per capita coefficients are positive and highly significant for both source and host in each of the four versions in Table 4.

According to Bergstrand and Egger (2007), transportation costs should be positively related to foreign affiliate sales, i.e. as transportation costs increases, foreign affiliate sales increase. This implies a form of substitution between trade and foreign affiliate sales. The trade openness variable for the host countries is expected to have a negative coefficient, so that with greater trade openness, foreign affiliate sales are expected to be lower. This is the result we find in each of the four cases examined. Common language is expected to be positively related to foreign affiliate sales, so that countries that share a common language are likely to have higher affiliate sales in each other's markets. This variable is indeed positive in every case but PPML, where it is slightly negative and not significant. Distance is negative, as is the case in gravity equations. BE do not use it in their estimation; instead they use fixed effects by country pair; however it is used by Carr et al. (2001).

The FDI restrictiveness index is a measure of host-specific sector-level investment restrictiveness. The expected sign on this measure is negative: given a higher level of restrictiveness, foreign affiliate sales should be lower. The expected sign on the FDI restrictiveness is correspondingly negative. Our results follow both of these predictions.

The coefficient on the skilled/unskilled labor ratio is positive as expected, and implies that firms are more likely to invest in countries that are relatively less skilled labor intensive

Table 4: Econometric Results

	(1)	(2)	(3)	(4)
	OLS	PPML	ZIP	ZINB
Ln(GDP_{st})	-0.0162	0.0228	-0.177***	-0.100**
	(-0.52)	(0.77)	(-5.53)	(-2.67)
Ln(GDP RoW_{rst})	-11.24***	-15.93***	-12.53***	-9.747***
	(-20.39)	(-20.91)	(-17.37)	(-15.42)
Ln($Prod_{irt}$)	0.278***	0.531***	0.412***	0.252***
	(19.79)	(28.95)	(18.95)	(13.37)
Ln(GDP/cap_{st})	0.710***	1.538***	0.555***	0.511***
	(11.38)	(24.03)	(6.23)	(6.82)
Ln(GDP/cap_{rt})	1.293***	1.153***	1.097***	1.138***
	(32.4)	(16.57)	(14.88)	(18.32)
Ln($Distance_{rs}$)	-0.507***	-1.033***	-0.525***	-0.258***
	(-15.12)	(-22.09)	(-11.40)	(-5.80)
Comm $Lang_{rs}$	0.273***	-0.00414	0.0117	0.104
	(3.67)	(-0.04)	(0.12)	(1.50)
Trad $Open_{rt}$	-0.299***	-0.385***	-0.184	-0.189*
	(-6.04)	(-4.19)	(-1.93)	(-2.35)
FDI $Restrict_{ir}$	-0.428***	-0.550***	-0.762***	-0.584***
	(-3.32)	(-3.47)	(-5.11)	(-5.07)
Skill $Diff_{rst}$	2.708***	4.130***	3.082***	2.472***
	(8.56)	(6.83)	(5.71)	(6.36)
N	6,327	43,541	43,541	43,541
R^2	0.511	0.509		

Notes: Standard errors are robust for OLS, ZIP and ZINB; t statistics in parentheses; * p<0.05, ** p<0.01, *** p<0.001 . Dependent variable is ln(foreign affiliate sales) for the OLS specification and levels for the other three versions. Year dummies are not shown for brevity.

than themselves, or that a relatively large amount of unskilled labor is attractive to foreign investors.

In Bergstrand and Egger (2007), the estimated coefficients display results that are similar to ours.[8] They do not report regression results on FAS data, but rather only regression results on FDI data; however they analyze their model results with respect to both FDI and FAS and find that in most dimensions the two variables respond similarly to changes in model variables. In particular, the signs are the same with the exception of host country trade costs where their regressions produce the wrong sign. The coefficients from BE and from our regressions cannot be quantitatively compared because the two specifications use different measures for trade costs.

As another point of comparison, we examine Carr et al. (2001) which has similar analysis to ours. In their case, the model is only a 2 country, 2 factor model, but explicitly considers foreign affiliate sales rather than investment. All of the coefficient results are as predicted by their model.[9] There are some differences that make for difficulty in comparing their results with ours. Carr et al. (2001) uses the sum of the GDPs rather than source and host GDPs. Skill difference is positively related to foreign affiliate sales. Trade costs of host countries are positively related to foreign affiliate sales, and investment costs negatively related to foreign affiliate sales. Trade costs of source countries are negatively related to foreign affiliate sales. In addition to these variables, the model also includes GDP times skill difference and trade costs multiplied by skill difference, which act as quadratic terms and are negative as expected.

Sectoral production is available for 21 sectors, all but two of which are manufacturing sectors. The two remaining sectors are real estate, renting and business activities, and hotels and restaurants.

The two zero inflate models, ZIP and ZINB, each have an additional logistic portion of the model that is not displayed. In this portion of the model there are three variables that are meant to summarize the criteria under which a country may invest in a particular sector in another country. The three variables are the FDI restrictiveness index due to Koyama and Golub (2006), the measure of common language, and a measure of border contiguity. The latter is not part of the original model; it is drawn from CEPII's database and takes on

[8]The coefficients reported by Bergstrand and Egger (2007) are on FDI, not FAS. They do not report regression results on FAS data; however they analyze their model results with respect to both FDI and FAS and find that in most dimensions the two variables respond similarly to changes in model variables.

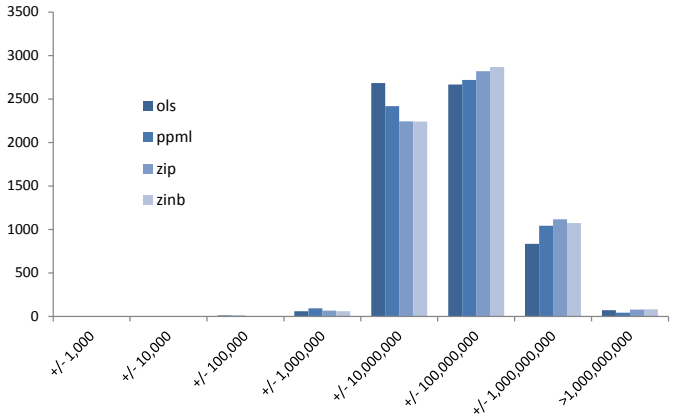
[9]The variables used by Carr et al. (2001) are: the sum of GDPs, the difference of GDPs squared, the skill difference, the interaction of skill difference and GDP difference, investment costs of host, trade costs of host, trade costs of host interacted with squared skill difference, trade cost of source and distance.

the value of one if two countries share a border and zero otherwise. The main portion of the model is very robust to the selection of the "inflate" variables.

The data are substantially overdispersed as seen in Table 5. This indicates against the use of PPML, as the Poisson distribution assumption of equal mean and variance is not met. In terms of mean values, the PPML fitted values come very close to the mean value of the data (the PPML result's mean value is matched exactly to that of the data). PPML underpredicts the mean value at 87 percent of that of the data, conditional on positive values. ZIP and ZINB mean fitted values underpredict the mean data values both unconditionally and conditional on positive values. ZIP underpredicts at 41 percent of the mean data value with zeros and at 38 percent without zeros; ZINB underpredicts at 40 percent with zeros and 37 percent without. From the perspective of dispersion, none of the estimation methods manages to capture the high level of dispersion of the data. PPML again does the best job, capturing 69 percent of the dispersion (conditional on positive values). ZIP and ZINB fail to capture most of the dispersion with standard deviations only 2=32 and 37 percent that of the data respectively.

We find very different zeros for data, PPML and ZIP/ZINB. ZIP and ZINB produce the same number of predicted zeros as the logit regression is the same for both. OLS is not displayed as it predicts no zeros. Clearly PPML produces far too few zeros. The ZIP/ZINB values are targeted to the data by selecting the cutoff point that produces the share of zeros observed in the data. There is no theoretical reason to choose a particular cutoff value.

Figure 3: Residuals compared across versions



We perform several tests of the econometric specifications to formalize the preceding analysis. Examining the (negative) log likelihoods generated by PPML, ZIP and ZINB indicates that ZINB is the most preferred out of the three, given that its log likelihood is

Table 5: Examining the Dispersion of Data and Fitted Values

Foreign affiliate sales	Mean ($ mil)	StdDev	CoefVar
Data			
with zeros	136	1,070	7.9
without zeros	936	2,680	2.9
size difference (without/with)	6.9	2.5	
OLS			
without zeros	405	688	1.7
percent of Data results	43%	26%	
PPML			
with zeros	136	766	5.6
percent of Data results	100%	72%	
without zeros*	810	1,860	2.3
percent of Data results	87%	69%	
size difference (without/with)	6	2.4	
ZIP fitted values			
with zeros	56	369	6.6
percent of Data results	41%	34%	
without zeros	351	868	2.5
percent of Data results	38%	32%	
size difference (without/with)	6.3	2.4	
ZINB fitted values			
with zeros	54	319	5.9
percent of Data results	40%	30%	
without zeros	342	736	2.2
percent of Data results	37%	27%	
size difference (without/with)	6.3	2.3	

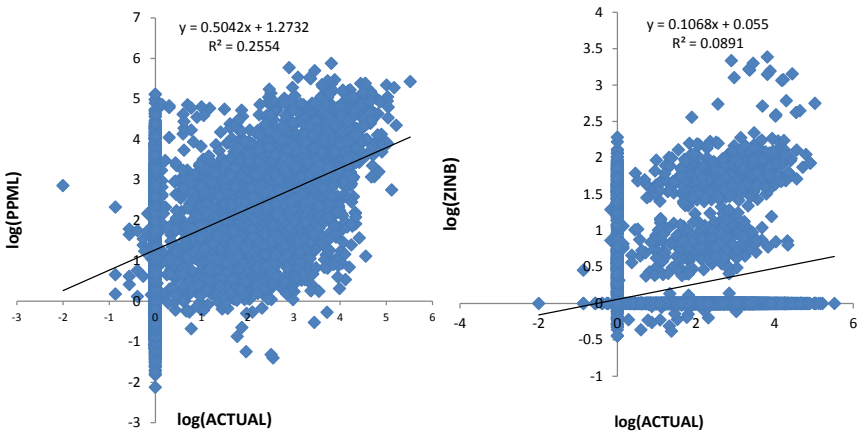
* Estimates less than $500,000 are rounded to zero

Table 6: Zeros

Source	Positive Values	Zeros
Data	14.50%	85.50%
PPML	78.90%	21.10%
ZIP/ZINB	15.90%	84.10%

the smallest.[10] Additionally we compute a more specific test to examine whether the ZIP or ZINB proves to be a better fit. The likelihood ratio test for over dispersion between ZIP and ZINB examines whether the estimated mean and variance are equal (as in ZIP) or substantially different (as in ZINB). See Cameron and Trivedi (1998). The LR test yields a result that strongly rejects the null hypothesis that the mean equals the variance.

Figure 4: Comparison of actual values versus PPML and ZINB



The results obtained using the theoretical models present certain problems. The variables used in the logistic portion of the zero inflated regressions – the so-called "inflate" variables – present some difficulty in terms of operationalizing the extrapolation of data based on the coefficients produced by the regressions. The regressions described above were based on a set of inflate variables that are known to act as barriers FDI – lack of common language, contiguous borders, and policies that restrict FDI. Although these variables produced estimates that in at least some behave substantially better than either OLS or PPML estimations, a

[10]We can also examine the consistent Akaike information criterion (CAIC), which in our case presents essentially identical results, as the main difference between the two – adjustments for number of observations and number of parameters – are similar across our models.

close examination of the logistic portion of the model reveals some peculiarities. The zero inflated methodologies produce thresholds that do not vary sufficiently by country – common language and contiguous borders take the value of one in a minority of the cases. The major variation is across sectors. The clear solution is to add variables that are country specific such as GDP or per capita GDP; however such variables tend to overwhelm the FDI restrictiveness in importance and economic significance; as a result the opposite problem is seen where each country will either receive investment in all of its sectors or receive no investment at all. As a result, despite the promising behavior of the zero inflated models, we proceed with the PPML version of the model.

4 Quadratic Optimization

Subsequent to filling in the missing values using econometric extrapolation, final consistency of the database is obtained using a quadratic optimization technique[11] that allows us to incorporate and reconcile information from different sources: econometric estimates, OECD (OECD, 2010), EUROSTAT (Eurostat, 2012), BEA (BEA, 2012) and the National Bureau of Statistics of China. This approach parallels that of Gouel et al. (2012).

The objective is to minimize the difference between initial estimates and final values subject to adding up constraints. Thus, for a given sector i, host country r and source

[11] Quadratic optimization has several numerical advantages in implementing very large models relative to cross-entropy minimization techniques (Canning and Wang, 2005).

country s and reliability weight w, the quadratic optimization is implemented as follows:

$$
\begin{aligned}
&Minimize \\
&\sum_{irs} w_{irs}(FATS^1_{irs} - FATS^0_{irs})^2 + \sum_{rs} w_{rs}(FATS^1_{rs} - FATS^0_{rs})^2 \\
&+\sum_{ir} w_{ir}(FATS^1_{ir} - FATS^0_{ir})^2 + \sum_{is} w_{is}(FATS^1_{is} - FATS^0_{is})^2 \\
&+\sum_{r} w_{r}(FATS^1_{r} - FATS^0_{r})^2 + \sum_{s} w_{s}(FATS^1_{s} - FATS^0_{s})^2 + w(FATS^1 - FATS^0)^2 \\
&subjectto \\
&\sum_{irs} FATS^1_{irs} = FATS^{UNCTAD}_{AGG} \\
&\sum_{i} FATS^1_{irs} = FATS^{EURO+OECD+BEA}_{rs} \\
&\sum_{s} FATS^1_{irs} = FATS^{EURO+OECD+BEA+NBSChina}_{ir} \\
&\sum_{r} FATS^1_{irs} = FATS^{EURO+OECD+BEA}_{is} \\
&\sum_{is} FATS^1_{irs} = FATS^{EURO+OECD+BEA}_{r} \\
&\sum_{ir} FATS^1_{irs} = FATS^{EURO+OECD+BEA}_{s}
\end{aligned}
\tag{3}
$$

where the $FATS^0$ variables denotes the initial sector/host/source specific foreign affiliates sales data constructed using the econometric estimates and the raw data collected from OECD, EUROSTAT, BEA and China's NBS. $FATS^1$ denotes the final values resulting from the optimization. Apart from the three-dimensional data we enrich the dataset with information about host and sectoral totals. The constraints of the optimization are aimed to target these aggregate values such as the total global activity of foreign affiliates $FATS^{UNCTAD}_{AGG}$, sector/host specific totals $FATS^{EURO+OECD+BEA+NBSChina}_{ir}$, sector/source specific totals $FATS^{EURO+OECD+BEA}_{is}$, bilateral totals $FATS^{EURO+OECD+BEA}_{rs}$ and host and source specific totals $FATS^{EURO+OECD+BEA}_{r}$ and $FATS^{EURO+OECD+BEA}_{s}$.

Certain data were fixed, and therefore were not optimized as above but taken as reported. These were the U.S. BEA data at the three dimensional level (source, host and sector), Eurostat and OECD data at the three dimensional level (including all zero values), the UNCTAD global total for foreign affiliate sales, and China inward data at the three dimensional level. Wherever the two dimensional data at the source-host-bilateral level, were available from the OECD, Eurostat or BEA, these were also used as targets. For these data, we placed

upper and lower bounds for feasibility reasons, i.e. in certain cases the bilateral values are inconsistent with the data available for the three dimensional values.

Initial values for foreign affiliate sales for host countries, where there is no information available, are derived from the proportions of each country's inward FDI stocks data to global FDI stocks. A similar process is undertaken for foreign affiliate sales by source. Initial values for host country foreign affiliate sales by sector are derived from the share of a country's GDP that is in that sector (derived from GTAP data) multiplied by the initial values for the host country's foreign affiliate sales.

Most recent data are used when possible. For the global value of foreign affiliate sales, the 2005-2007 average provided by UNCTAD was used.[12] For Eurostat data, the average of 2005-2007 was used in order to be able to take into account data from countries that did not report in 2007. For the U.S. BEA and China, 2007 data were used. Outward foreign affiliate data used are an average of 2007-2009.

Reliability weights are chosen to reflect our confidence in the correctness of the underlying data. Higher weights increase the penalty for deviating from the initial values and so are used with data in which we have greater confidence; correspondingly, lower weights are used for less certain data. Note that when all weights are equal to one the solution of this model is the constrained least square estimator. The weights on the extrapolated estimates were the lowest, at 10, implying the greatest level of uncertainty. Two dimensional data for sector and host are given weights of 10,000, and the weights for the one dimensional host and source data are 100,000.

5 Final Database

The final database has 128 countries and 31 sectors. The extrapolated dataset estimates that 35.3 percent of global foreign affiliate sales are in manufacturing, while 58.1 percent are in services (with the remaining 6.5 percent in agriculture and natural resources). See Table 7. Verifying the validity of these results is difficult because a sectoral breakdown of original source data is particularly scarce and is not available at a global level. When compared with the Eurostat data, the global extrapolated results show a relatively higher weight for manufacturing and for natural resources than does the Eurostat data. This seems reasonable given that many developing countries are likely to be overweighted in the mining sector and that services (particularly financial services) more likely to take place in European Union

[12] UNCTAD (2011) World Investment Report. Table 1.5.

countries than in many other countries outside the EU. However, the extent to which the rather substantial difference between the two is a true reflection of sectoral divisions cannot be determined without new data sources.

Table 7: Final Database versus original data input ($ billions)

	Eurostat Database		Extrapolation	
Sector	Value	Share	Value	Share
Agriculture & Natural Resources	52	1.20%	1,390	6.50%
Manufacturing	1,440	33.10%	7,523	35.30%
Services	2,832	65.50%	12,379	58.10%
Total	4,324		21,293	

Host countries exhibit a mix of foreign affiliate sales by sector. In Table 8 the 129 countries of the database are grouped into eight regions, with Australia and New Zealand grouped together, East Asia (including Japan, S. Korea, and Taiwan) in another group, the ten ASEAN countries in a third, the EU as a fourth region and the United States, India and China each treated separately.

There is a realistic amount of heterogeneity across sectors, which is consistent with perceptions of variations in invested sectors by countries. Overall, foreign affiliate sales in manufacturing accounts for 35.3 percent of global foreign affiliate sales. According to the final database, China has a higher share of foreign affiliate sales in the manufacturing sector than any other country ($474 billion out of a total $531 billion foreign affiliate sales in China, or 89.3 percent); this is in line with the prior notion of China as a manufacturing giant. Similarly, Australia and New Zealand have very high mining revenues as a share of its total foreign affiliate sales ($49 billion out of a total $671 billion, or 7.3 percent), well above the 5.0 percent global average for mining, which is consistent with Australia's large extractive sector. India is shown as having a particularly small foreign owned wholesale and retail trade sector, which is to be expected given its lack of reform in retail services, particularly in single brand retail. The data for the United States largely reflect the data provided by the U.S. BEA. East Asia is a relatively major host of transport services, with 15.3 percent of its foreign affiliate sales attributed to that sector (relative to a global average of 5.3 percent).

The database also exhibits expected patterns of investment behavior from the source perspective. The United States has a large share of foreign affiliate sales in manufacturing

Table 8: Final database values of host countries by sector, average 2005-2007 ($ bn)

Host	Agric	Mining	Manuf	Distrib	Transp	OthServ	Total	Share
United States	2	372	1,203	1,126	138	775	3,616	17%
China	1	3	474	43	3	7	531	2%
India	12	5	68	11	13	49	157	1%
East Asia	20	54	493	449	267	462	1,745	8%
ASEAN	32	41	441	174	80	280	1,047	5%
Aus/NZ	20	49	201	67	41	293	671	3%
EU	48	80	2,680	2,677	245	2,106	7,838	37%
RoW	201	450	1,963	765	330	1,978	5,687	27%
Total	**336**	**1,054**	**7,523**	**5,311**	**1,119**	**5,949**	**21,293**	**100%**
Share	1.60%	5.00%	35.30%	24.90%	5.30%	27.90%	100.00%	

Table 9: Final database values of source countries by sector, average 2005-2007 ($ bn)

Source	Agric	Mining	Manuf	Distrib	Transp	OthServ	Total	Share
United States	98	364	2,263	1,247	241	1,307	5,522	26%
China	0.1	32.7	35	21	10	52	151	1%
India	0.1	0.1	1.6	19	0.3	32.5	54	0%
E Asia	27	79	773	753	131	514	2,277	11%
ASEAN	0.7	30	81	159	37	131	440	2%
Aus/NZ	9	47	228	42	64	183	573	3%
EU	120	331	2,919	2,294	444	2,631	8,741	41%
RoW	79	170	1,221	773	190	1,097	3,532	17%
Total	**336**	**1,054**	**7,523**	**5,311**	**1,118**	**5,949**	**21,293**	**100%**
Share	1.60%	5.00%	35.30%	24.90%	5.30%	27.90%	100.00%	

abroad: it has a slightly greater share of foreign affiliate sales coming from its investments in manufacturing (41.0 percent) abroad than the global average (35.3 percent). In line with the known outward foreign investment by Chinese firms, 21.6 percent of China's foreign affiliate sales abroad come from the mining sector. (By contrast, it is a virtually negligible host for mining firms). Fully 60.0 percent of India's foreign affiliate sales abroad are in "other services", which includes business services, a particular strength of India.

The largest source countries, as estimated by the final database, are largely in line with expectations. The United States, Germany, and the United Kingdom are all well-known as significant sources of foreign direct investment capital. Hong Kong, France, Canada and Australia also appear among both the top ten host and source countries. The Netherlands, Switzerland and Japan round out the remaining top ten sources, while Italy, Russia and China round out the remaining top ten hosts.

Certain countries are known to exhibit asymmetries; in particular, Japan is known as a source of FDI but not as a host of FDI. Consistent with this, Japan is ranked eighth among sources of FDI while it is 24th among hosts of FDI. Conversely, developing countries such as China and India generally attract FDI but are not themselves major sources of FDI. In our database, China ranks as the tenth largest host of foreign affiliate sales activity but 25th as a source while India is the 31st largest host of foreign affiliate sales activity and the 34th largest source.

According to FDI statistics, sources of foreign investment tend to be more concentrated than hosts: wealthy developed countries dominate the ranks of sources of investment, while their investments are scattered broadly across all countries. Our database fits this stylized fact well. The top ten source countries together account for approximately 72.8 percent of global foreign affiliate sales, or $15.5 trillion in sales. The top ten hosts are less concentrated, comprising only 54.1 percent of total sales, or about $11.5 trillion. Detailed tables for source, host, and sector totals are presented in appendix Tables 17, 18, and 19.

Finally, we compare our inward foreign affiliate sales data with inward FDI stocks obtained from UNCTAD's World Investment Report data. Comparing foreign affiliate sales with FDI is problematic as they are substantially different objects. Certain issues may weaken the correlation between shares of each of the two variables, such as the relative capital intensity of the investment in particular regions of the world. For example, we might find countries that have large banking sectors such as Luxembourg to be the host to a much higher proportion of global FDI relative to its global share of foreign affiliate sales. Age of the capital installed may also matter; countries that have experienced very recent invest-

ments of capital may have not yet realized the full potential in terms of foreign affiliate sales. Finally, countries in which foreign investment is generally made via joint venture or other forms of partial ownership will see high foreign affiliate sales relative to their investment (capital stocks for only their partial ownership is reported, whereas total sales by the affiliate are reported) while a country where 100 percent investments are common would see lower foreign affiliate sales relative to FDI.

These caveats aside, there generally is a positive association between FDI and foreign affiliate sales, and a comparison with FDI may offer some hints as to the appropriateness of the new dataset. The two sets of data, UNCTAD and our foreign affiliate database, are compared for the eight regions in Table 10. The shares exhibit a close correspondence. The largest region, the EU, comprises 36.8 percent of global foreign affiliate sales as host, while also accounting for 39.6 percent of inward FDI stocks. The United Sates, is second largest for each of the measures. East Asia (composed of Japan, Hong Kong, Taiwan and South Korea) is a both a moderately-sized host of foreign affiliate activity and of inward stocks. Smaller hosts of foreign affiliate activity – ASEAN, Australia and New Zealand, China and India – each have similarly small shares of inward FDI stocks. As a result, the share of the rest of the world are also similar between FDI and foreign affiliate sales.

Table 10: Comparison of final database with inward FDI stocks, average 2005-2007 ($ bn)

Host	Inward FDI Stocks		Foreign affiliate sales	
	Value	*share*	Value	*share*
United States	3,551.30	18.70%	3,616.20	17.00%
China	327.1	1.70%	530.9	2.50%
India	105.8	0.60%	157.4	0.70%
East Asia	1,360.00	7.20%	1,745.40	8.20%
ASEAN	654.6	3.50%	1,047.40	4.90%
Aus/NZ	453.5	2.40%	671	3.20%
EU	7,515.80	39.60%	7,838.10	36.80%
RoW	4,989.20	26.30%	5,686.50	26.70%
Total	18,957.30		21,293.00	

Source: Inward FDI Stocks taken from UNCTAD's World Investment Report, Annex table 3. http://www.unctad.org/Templates/Page.asp?intItemID=5823&lang=1 Accessed 2/29/2012.

6 Conclusion

The purpose of this study has been to bring as much data as is currently available to bear on the problem of constructing a large global database of foreign affiliate sales. The newer methods of handling zeros proved to be substantially better at handling the Eurostat dataset than prior methods. In this sense, we present empirical evidence to suggest that future work with foreign affiliate sales and indeed foreign direct investment should be performed using models that take into account the information that the zeros in the dataset provide. However, as a practical matter for extrapolating values from the coefficients, there remains considerable work to be done. Obtaining probabilities from the logistic regression that produce realistic patterns proved elusive. As a result, PPML remained the most useful technique for both addressing zeros and providing plausible numbers for extrapolation.

Future work will be done on the estimation, in particular to attempt to identify relevant variables that can render zero inflated models operational. Additionally, there is a great lack of data that hinders the construction of the database. Although there is an increasing amount of data on the investment side there is not a sufficiently strong correlation between the two to permit their interchangeability. There is a great need to improve the availability of data on the foreign affiliate side.

A Appendix

Table 11: Foreign affiliate sales: data sources

Source	Database	Years used	Variable	Last accessed
BEA	Financial and Operating Data for U.S. MNCs	2007		3/8/2012
Eurostat	Foreign Affiliate Statistics	2004-2007	fats_g1a_03; fats_out	11/15/2011
NBS China	Aggregated firm survey data	2007	v209	
OECD	Statistics on Measuring Globalization	2004-2007	fats_in3_serv; fats_out3_serv	11/16/2011
UNCTAD	2011 World Investment Report	2005-2007	affiliate sales Table I.5	

Table 12: Source Countries in Eurostat database

Australia	France	Liechtenstein	Slovakia
Austria	Germany	Lithuania	Slovenia
Belgium	Greece	Luxembourg	Spain
Bulgaria	Hong Kong	Malta	Swede
Canada	Hungary	Netherlands	Switzerland
China (incl. HK)	Iceland	New Zealand	Turkey
Cyprus	Ireland	Norway	United Kingdom
Czech Republic	Israel	Poland	United States
Denmark	Italy	Portugal	
Estonia	Japan	Romania	
Finland	Latvia	Russia	

Source: Eurostat. Note that Liechtenstein and Luxembourg are excluded from the regression analysis.

Table 13: Host Countries in Eurostat database

Austria	Finland	Lithuania	Slovenia
Bulgaria	France	Netherlands	Spain
Cyprus	Germany	Poland	Sweden
Czech Republic	Hungary	Portugal	United Kingdom
Denmark	Italy	Romania	
Estonia	Latvia	Slovakia	

Source: Eurostat

Table 14: Sectors covered in the econometric estimation

Manufacturing Sectors
·Food products, beverages and tobacco*
·Textiles*
·Wearing apparel; dressing; dyeing of fur*
·Leather and leather products*
·Wood and wood products*
·Pulp, paper and paper products; publishing and printing*
·Coke, refined petroleum products and nuclear fuel
·Chemicals, chemical products and man-made fibers*
·Rubber and plastic products*
·Other non-metallic mineral products*
·Basic metals*
·Fabricated metal products, except machinery and equipment*
· Machinery and equipment n.e.c.*
·Office machinery and computers*
·Electrical machinery and apparatus n.e.c.*
·Radio, television and communication equipment and apparatus*
·Medical, precision and optical instruments, watches and clocks*
·Motor vehicles, trailers and semi-trailers*
·Other transport equipment*
·Manufacturing n.e.c.*
Services and Other Sectors
·Mining and quarrying
·Electricity, gas, steam and hot water supply
·Collection, purification and distribution of water
·Construction
·Sale, maintenance and repair of motor vehicles and motorcycles
·Wholesale trade and commission trade, except of motor vehicles
·Retail trade, except of motor vehicles and motorcycles; repair
·Hotels and restaurants*
·Transport, storage and communication
·Financial intermediation, except insurance and pension funding
·Insurance and pension funding, except compulsory social security
·Activities auxiliary to financial intermediation
·Real estate, renting and business activities*

Source: Eurostat. Note that * denotes sectors included in the regression analysis.

Table 15: Eurostat data on foreign affiliate sales by source country, 2007.

Source Country	in $ billions	share
United States	589	34.60%
Netherlands	190	11.20%
Germany	187	11.00%
France	183	10.80%
Switzerland	136	8.00%
United Kingdom	102	6.00%
Sweden	48	2.80%
Italy	40	2.30%
Finland	38	2.20%
Austria	36	2.10%
Japan	32	1.90%
Denmark	29	1.70%
Belgium	26	1.50%
Norway	18	1.00%
Spain	16	1.00%
Ireland	16	0.90%
Canada	4	0.30%
Russian Federation	2	0.10%
Cyprus	2	0.10%
Czech Republic	2	0.10%
Israel	1	0.10%
Greece	1	0.10%
Australia	1	0.10%
Portugal	1	0.00%
Turkey	0	0.00%
Iceland	0	0.00%
Hungary	0	0.00%
Estonia	0	0.00%
Hong Kong	0	0.00%
Slovenia	0	0.00%
Poland	0	0.00%
Romania	0	0.00%

Malta	0	0.00%
Lithuania	0	0.00%
Romania	0	0.00%
Slovakia	0	0.00%
Hong Kong	0	0.00%
Bulgaria	0	0.00%
Latvia	-	0.00%
New Zealand	-	0.00%
Latvia	-	0.00%
Total	1,702	100.00%

Table 16: Eurostat data on foreign affiliate sales by host country, 2007.

Host country	in $ billions	share
Germany	579	34%
United Kingdom	329	19%
Italy	239	14%
Netherlands	108	6%
Poland	97	6%
Sweden	97	6%
Austria	73	4%
Hungary	65	4%
Finland	27	2%
Denmark	25	1%
Portugal	24	1%
Romania	20	1%
Slovakia	14	1%
Estonia	3	0%
Latvia	2	0%
Slovenia	1	0%
Cyprus	0	0%
Total	1,702	100%

Table 17: Final database results by source country, average 2005-2007

Rank	Source	Sales (USD m)	Share	Rank	Source	Sales (USD m)	Share
1	United States	5,522,695	25.90%	61	Ukraine	4,613	0.00%
2	Germany	2,071,352	9.70%	62	Kazakhstan	4,341	0.00%
3	United Kingdom	1,335,498	6.30%	63	Indonesia	4,254	0.00%
4	France	1,314,500	6.20%	64	Lithuania	3,323	0.00%
5	Netherlands	1,152,668	5.40%	65	Peru	3,219	0.00%
6	Hong Kong	1,069,122	5.00%	66	Egypt	2,587	0.00%
7	Switzerland	880,871	4.10%	67	Oman	2,470	0.00%
8	Japan	835,746	3.90%	68	Slovakia	2,398	0.00%
9	Canada	768,995	3.60%	69	Iran	2,297	0.00%
10	Australia	545,764	2.60%	70	Pakistan	2,095	0.00%
11	Italy	542,532	2.50%	71	Morocco	2,047	0.00%
12	Russia	491,028	2.30%	72	Botswana	1,926	0.00%
13	Caribbean	378,487	1.80%	73	Rest of East Asia	1,781	0.00%
14	Slovenia	336,429	1.60%	74	Nigeria	1,629	0.00%
15	Singapore	334,969	1.60%	75	Rest North America	1,554	0.00%
16	Spain	308,431	1.40%	76	Rest Southeast Asia	1,304	0.00%
17	Denmark	307,883	1.40%	77	Latvia	986	0.00%
18	Taiwan	262,368	1.20%	78	Bulgaria	931	0.00%
19	Ireland	251,055	1.20%	79	Central Africa	885	0.00%
20	Sweden	243,331	1.10%	80	South Central Africa	785	0.00%
21	Brazil	223,020	1.00%	81	Rest Oceania	705	0.00%
22	Finland	210,243	1.00%	82	El Salvador	651	0.00%

23	Belgium	200,208	0.90%	83	Costa Rica	631	0.00%
24	Austria	155,795	0.70%	84	Guatemala	575	0.00%
25	China	151,551	0.70%	85	Cambodia	556	0.00%
26	Luxembourg	148,079	0.70%	86	Ecuador	546	0.00%
27	Korea	108,906	0.50%	87	Cameroon	506	0.00%
28	South Africa	103,391	0.50%	88	Mauritius	489	0.00%
29	Norway	102,921	0.50%	89	Zimbabwe	489	0.00%
30	Malaysia	78,072	0.40%	90	Uruguay	478	0.00%
31	Israel	75,154	0.40%	91	Rest Eastern Africa	431	0.00%
32	Mexico	74,184	0.30%	92	Sri Lanka	415	0.00%
33	Portugal	58,894	0.30%	93	Paraguay	382	0.00%
34	India	54,141	0.30%	94	Kenya	335	0.00%
35	Chile	52,946	0.20%	95	Senegal	335	0.00%
36	Argentina	51,411	0.20%	96	Malta	333	0.00%
37	Panama	44,620	0.20%	97	Romania	321	0.00%
38	Arab Emirates	43,508	0.20%	98	Nicaragua	223	0.00%
39	Rest of EFTA	32,932	0.20%	99	Zambia	222	0.00%
40	New Zealand	27,186	0.10%	100	Bangladesh	208	0.00%
41	Venezuela	25,140	0.10%	101	Bolivia	182	0.00%
42	Poland	23,589	0.10%	102	Tunisia	173	0.00%
43	Kuwait	21,028	0.10%	103	Rest S African Customs	148	0.00%
44	Colombia	19,990	0.10%	104	Rest of Europe	142	0.00%
45	Turkey	19,680	0.10%	105	Laos	106	0.00%
46	Cyprus	19,546	0.10%	106	Kyrgyzstan	102	0.00%
47	Hungary	17,577	0.10%	107	Rest of Central America	91	0.00%

48	Croatia	15,622	0.10%	108	Albania	89	0.00%
49	Saudi Arabia	15,028	0.10%	109	Mongolia	86	0.00%
50	Czech Republic	14,662	0.10%	110	Georgia	77	0.00%
51	Thailand	14,368	0.10%	111	Cote d'Ivoire	62	0.00%
52	Greece	13,116	0.10%	112	Rest Eastern Europe	61	0.00%
53	Rest of Western Asia	12,655	0.10%	113	Belarus	53	0.00%
54	Bahrain	12,615	0.10%	114	Honduras	48	0.00%
55	Azerbaijan	8,539	0.00%	115	Namibia	32	0.00%
56	Estonia	7,832	0.00%	116	Armenia	27	0.00%
57	Rest North Africa	6,680	0.00%	117	Malawi	24	0.00%
58	Philippines	6,579	0.00%	118	Madagascar	13	0.00%
59	Rest Western Africa	6,565	0.00%	119	Rest of South America	3	0.00%
60	Qatar	5,531	0.00%		World Total	21,293,003	99.70%

Table 18: Final database results by host country, average 2005-2007

Rank	Host	Sales (USD m)	Share	Rank	Host	Sales (USD m)	Share
1	United States	3,616,247	17.00%	65	Rest North Africa	32,364	0.20%
2	United Kingdom	1,544,237	7.30%	66	Bulgaria	30,156	0.10%
3	Germany	1,390,611	6.50%	67	South Central Africa	27,671	0.10%
4	Hong Kong	1,147,514	5.40%	68	Rest Eastern Africa	27,665	0.10%
5	France	978,663	4.60%	69	Panama	26,764	0.10%
6	Canada	666,334	3.10%	70	Azerbaijan	25,559	0.10%
7	Italy	558,931	2.60%	71	Qatar	23,738	0.10%
8	Russia	548,654	2.60%	72	Bahrain	23,201	0.10%
9	Australia	543,758	2.60%	73	Rest Western Africa	22,624	0.10%
10	China	530,903	2.50%	74	Ecuador	21,702	0.10%
11	Switzerland	491,125	2.30%	75	Rest EFTA	20,596	0.10%
12	Singapore	477,870	2.20%	76	Slovenia	20,031	0.10%
13	Mexico	473,075	2.20%	77	Rest Europe	18,762	0.10%
14	Brazil	458,975	2.20%	78	Rest East Asia	17,761	0.10%
15	Netherlands	428,128	2.00%	79	Lithuania	16,012	0.10%
16	Spain	398,643	1.90%	80	Costa Rica	15,031	0.10%
17	Ireland	374,849	1.80%	81	Latvia	14,997	0.10%
18	Caribbean	368,211	1.70%	82	Malta	13,740	0.10%
19	Belgium	305,921	1.40%	83	Zambia	13,627	0.10%
20	Poland	258,524	1.20%	84	Oman	13,603	0.10%
21	Sweden	255,504	1.20%	85	Rest of Oceania	11,822	0.10%
22	Korea	245,227	1.20%	86	Estonia	11,819	0.10%

23	Austria	240,128	1.10%	87	Rest Former Soviet U	11,706	0.10%
24	Japan	233,265	1.10%	88	Tanzania	11,219	0.10%
25	Turkey	229,334	1.10%	89	Bolivia	11,100	0.10%
26	Norway	213,015	1.00%	90	El Salvador	10,402	0.00%
27	South Africa	198,385	0.90%	91	Cote d'Ivoire	10,230	0.00%
28	Czech Republic	186,458	0.90%	92	Uruguay	9,375	0.00%
29	Chile	181,737	0.90%	93	Bangladesh	8,641	0.00%
30	Thailand	165,678	0.80%	94	Guatemala	8,471	0.00%
31	India	157,370	0.70%	95	Georgia	8,080	0.00%
32	Hungary	144,532	0.70%	96	Honduras	7,724	0.00%
33	Denmark	136,398	0.60%	97	Cameroon	7,522	0.00%
34	Luxembourg	132,294	0.60%	98	Ethiopia	6,996	0.00%
35	Argentina	130,960	0.60%	99	Belarus	6,871	0.00%
36	New Zealand	127,251	0.60%	100	Cambodia	6,618	0.00%
37	Indonesia	125,723	0.60%	101	Namibia	6,509	0.00%
38	Malaysia	124,494	0.60%	102	Sri Lanka	6,373	0.00%
39	Saudi Arabia	112,857	0.50%	103	Ghana	6,191	0.00%
40	Israel	107,235	0.50%	104	Mozambique	6,181	0.00%
41	Taiwan	101,657	0.50%	105	Nicaragua	5,968	0.00%
42	Colombia	99,190	0.50%	106	Uganda	5,836	0.00%
43	Venezuela	95,729	0.40%	107	Armenia	3,979	0.00%
44	Rest Western Asia	88,790	0.40%	108	Paraguay	3,818	0.00%
45	Greece	88,539	0.40%	109	Rest North America	3,678	0.00%
46	Arab Emirates	87,553	0.40%	110	Albania	3,493	0.00%
47	Egypt	84,682	0.40%	111	Rest S African Customs	3,155	0.00%

48	Finland	84,500	0.40%	112	Zimbabwe	3,076	0.00%
49	Romania	82,919	0.40%	113	Kenya	2,985	0.00%
50	Portugal	81,092	0.40%	114	Rest Eastern Europe	2,950	0.00%
51	Viet Nam	75,118	0.40%	115	Rest South Asia	2,702	0.00%
52	Kazakhstan	73,779	0.30%	116	Rest South America	2,541	0.00%
53	Nigeria	67,938	0.30%	117	Malawi	2,219	0.00%
54	Morocco	63,920	0.30%	118	Mongolia	2,195	0.00%
55	Croatia	62,258	0.30%	119	Mauritius	2,121	0.00%
56	Slovakia	58,410	0.30%	120	Cyprus	2,035	0.00%
57	Ukraine	56,111	0.30%	121	Madagascar	1,981	0.00%
58	Tunisia	46,428	0.20%	122	Laos	1,962	0.00%
59	Peru	45,223	0.20%	123	Botswana	1,902	0.00%
60	Iran	37,986	0.20%	124	Kuwait	1,692	0.00%
61	Philippines	37,474	0.20%	125	Rest Central America	1,579	0.00%
62	Pakistan	35,439	0.20%	126	Kyrgyzstan	1,400	0.00%
63	Central Africa	32,549	0.20%	127	Senegal	1,198	0.00%
64	Rest Southeast Asia	32,507	0.20%	128	Nepal	268	0.00%
					World Total	21,293,000	100.00%

Table 20: Final sectoral coverage - GTAP sector correspondence

Sector	**GTAP sectors**	**Description**
agr	pdr wht gro v_f osd c_b pfb ocr ctl oap rmk wol frs fsh	Agriculture
coaoilgas	coa oil gas	Mining
foodb_t	omn cmt omt vol mil pcr sgr ofd b_t	Food
tex	tex	Textiles
wap	wap	Wearing apparel
lea	lea	Leather products
lum	lum	Wood products
ppp	ppp	Paper products, publishing
p_c	p_c	Petroleum, coal products
crp	crp	Chemical, rubber, plastic products
nmm	nmm	Mineral products nec
i_snfm	i_s nfm	Ferrous and other metals
fmp	fmp	Metal products
ome	ome	Machinery and equipment nec
ele	ele	Electronic equipment
mvh	mvh	Motor vehicles and parts
otn	otn	Transport equipment nec
omf	omf	Manufactures nec
elygdt	ely gdt	Electicity and gas
wtr	wtr	Water
cns	cns	Construction
trd	trd	Trade
otp	otp	Transport nec
wtp	wtp	Water transport
atp	atp	Air transport
cmn	cmn	Communication
ofi	ofi	Financial services nec
isr	isr	Insurance
obs	obs	Business services nec
ros	ros	Recreational and other services
osg	osg	Public Administration, Defense, Education

Table 19: Final database results by sector, average 2005-2007.

Sector	**Sales (USD m)**	**Share**
Trade	5,311,393	24.90%
Chemical, rubber, plastic products	1,225,926	5.80%
Business services nec	1,160,428	5.40%
Coal, Oil, Gas	1,054,046	5.00%
Motor vehicles and parts	999,120	4.70%
Financial services nec	993,111	4.70%
Public admin, defense, education, health	954,294	4.50%
Construction	937,703	4.40%
Electronic equipment	900,583	4.20%
Machinery and equipment nec	765,874	3.60%
Food, Bev, Tobacco	749,512	3.50%
Transport nec	733,228	3.40%
Communication	658,454	3.10%
Petroleum, coal products	573,803	2.70%
Insurance	514,168	2.40%
Ferrous and other metal	449,888	2.10%
Utilities	381,244	1.80%
Agriculture	336,325	1.60%
Paper products, publishing	295,220	1.40%
Mineral products nec	272,483	1.30%
Metal products	257,491	1.20%
Recreational and other	228,996	1.10%
Transport equipment nec	214,919	1.00%
Air transport	213,086	1.00%
Manufactures nec	190,953	0.90%
Textiles	183,111	0.90%
Wood products	180,773	0.80%
Water transport	172,534	0.80%
Wearing apparel	155,956	0.70%
Water	120,951	0.60%
Leather products	107,429	0.50%
World Total	21,293,000	100.00%

References

Anderson, J. and E. Van Wincoop (2004). Trade Costs. *Journal of Economic Literature 42*, 691–751.

BEA (2012). Financial and Operating Data for U.S. Multinational Companies. *Direct Investment and Multinational Companies (database)*.

Bergstrand, J. 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), 278–308.

Beugelsdijk, S., J. 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), 1444–1459.

Blonigen, B. (2005). A Review of the Empirical Literature on FDI Determinants. *Atlantic Economic Journal 33*(4), 383–403.

Cameron, A. and P. Trivedi (1998). *Regression Analysis of Count Data*, Volume 30. Cambridge University Press.

Canning, P. and Z. Wang (2005). A Flexible Mathematical Programming Model to Estimate Interregional Input–Output Accounts. *Journal of Regional Science 45*(3), 539–563.

Carr, D., J. Markusen, and K. Maskus (2001). Estimating the Knowledge-Capital Model of the Multinational Enterprise. *The American Economic Review 91*(3), 693–708.

De Benedictis, L. and D. Taglioni (2011). The Gravity Model in International Trade. *Chapter 4 of The Trade Impact of the European Union Preferential Policies: an Analysis Through Gravity Models. Ed. Luca De Benedictis and Luca Salvatici.*

Eurostat (2007). Recommendations manual on the production of foreign affiliates statistics (fats). *Eurostat Methodologies and Working Papers.*

Eurostat (2012). Foreign Controlled EU Enterprises - Inward and Outward FATS. *Foreign affiliates statistics - FATS (database)*.

Gouel, C., H. Guimbard, and D. Laborde (2012). A Foreign Direct Investment Database for Global CGE Models. *CEPII Working Paper No 2012-08.*

Hanslow, K., T. Phamduc, and G. Verikios (2000). The Structure of the FTAP Model. *Economic Analysis 27*, 30.

Kleinert, J. and F. Toubal (2010). Gravity for FDI. *Review of International Economics 18*(1), 1–13.

Koyama, T. and S. Golub (2006). OECD's FDI Regulatory Restrictiveness Index: Revision and Extension to More Economies. *OECD Working Papers on International Investment*.

Markusen, J. (1997). Trade Versus Investment Liberalization. *NBER Working Paper No. 6231*.

Markusen, J. and K. Maskus (2002). Discriminating Among Alternative Theories of the Multinational Enterprise. *Review of International Economics 10*(4), 694–707.

Markusen, J., A. Venables, D. Konan, and K. Zhang (1996). A unified Treatment of Horizontal Direct Investment, Vertical Direct Investment, and the Pattern of Trade in Goods and Services. Technical report, National Bureau of Economic Research.

Martin, W. and C. Pham (2008). Estimating the gravity equation when zero trade flows are frequent. *MPRA Paper 9453*.

OECD (2010). Activity of Multinationals in Services: Inward and Outward Activity. *OECD Statistics on Measuring Globalisation (database)*.

Silva, J. and S. Tenreyro (2006). The Log of Gravity. *The Review of Economics and Statistics 88*(4), 641–658.

UNCTAD (2011). World Investment Report: Non-Equity Modes of International Production and Development. *United Nations*.

Yeaple, S. (2003). The Role of Skill Endowments in the Structure of US Outward Foreign Direct Investment. *Review of Economics and Statistics 85*(3), 726–734.