CEPII
CENTRE
D'ÉTUDES PROSPECTIVES
ET D'INFORMATIONS
INTERNATIONALES

No 2012 – 08
April

DOCUMENT DE TRAVAIL

A Foreign Direct Investment database for global CGE models

Christophe Gouel
Houssein Guimbard
David Laborde

TABLE OF CONTENTS

Non-technical summary 3
Abstract 4
Résumé non technique 5
Résumé court 6
1. Introduction 7
2. Construction of the database 8
2.1. Sources 9
2.1.1. Original sources 9
2.1.2. Some issues concerning FDI statistics 10
2.2. Methodology 12
2.2.1. Econometrics 13
2.2.2. Matrix balancing 17
2.2.3. Differences with original values 19
3. Conclusion 20
References 22
APPENDIX A: Coverage of the global sources 24
APPENDIX B: Sectoral and geographical breakdown 25
APPENDIX C: Sensitivity Analysis of econometrics estimation 28
List of working papers released by CEPII 31

A FOREIGN DIRECT INVESTMENT DATABASE FOR GLOBAL CGE MODELS

Christophe Gouel[1]
Houssein Guimbard[2]
David Laborde[3]

NON-TECHNICAL SUMMARY

Various international institutions gather FDI data and make them available to the public. IMF and UNCTAD do this at the world level. They provide for each country their inward and outward flows and stocks. OECD and Eurostat, through a joint questionnaire, provide aggregated bilateral data as well as sectoral-host country of OECD and European countries. But all this valuable information is far from a balanced database usable in Computable General Equilibrium (CGE) models. Indeed, these databases are both mutually and internally inconsistent. Except the Eurostat dataset, they are also missing one dimension needed (investor, host, sector) or contain many cells with unreported data.

We propose and apply a method to construct a balanced tri-dimensional (investor, host, sector) FDI database for 2004. The methodology is twofold. Firstly, we estimate all the missing values. Eurostat provides us with a good coverage of European FDI, but because of confidential data and missing values, this database remains partially filled. We complete European database and also data for other countries with estimates obtained from gravity-based regressions.

Secondly, we balance the database while imposing constraints. The database must respect, at least with some slacks, information brought by the various sources. We impose the database to match with aggregate values from IMF and UNCTAD, and with bilateral and sectoral inward values from OECD, WiiW (The Vienna Institute for International Economic Studies) and data from Statistical Yearbook of China (other constraints could also be added in further releases). The matrix-balancing is done by minimising the discrepancies between our prior information (both original sources and estimates) and final values while verifying constraints, based on aggregated information.

[1] World Bank, Washington and CEPII, Paris.
[2] CEPII, Paris, Correspondance : houssein.guimbard (at) cepii.fr.
[3] International Food Policy Research Institute, Washington.

The purpose of this exercise is thus to deliver a ready-to-use database for CGE modellers aiming at conducing exercises involving

ABSTRACT

We describe the methodology used to construct a global database of foreign direct investments in three dimensions (investor country, host country and sector) for 2004.

Based on Eurostat data, we estimate theoretical investments for all countries. Then we constrain our estimates subject to existing data with lower dimensions (1 or 2) during the balancing of the matrix, using a quadratic optimization.

This database is intended for use for CGE modeling studies.

JEL Classification: C 68, C 82, F 21

Key Words: Computable general equilibrium models, Foreign Direct investment, Databases

UNE BASE D'INVESTISSEMENTS DIRECTS A L'ETRANGER

POUR LES MODELES CALCULABLES MONDIAUX D'EQUILIBRE GENERAL

Christophe Gouel
Houssein Guimbard
David Laborde

RESUME NON TECHNIQUE

Plusieurs organisations internationales collectent auprès de leurs membres les statistiques d'investissment direct à l'éranger (IDE) que les banques centrales rassemblent *via* des questionnaires ou des déclarations bancaires. Le contenu de ces données, fournies en flux et en stocks, diffère d'une base à l'autre. Le FMI et la CNUCED construisent, chacun, une base mondiale d'IDE : pour chaque pays est fourni l'IDE en provenance (flux entrant) ou à destination (flux sortant) du reste du monde. L'OCDE et Eurostat construisent leurs propres bases à partir d'enquêtes communes. La base OCDE contient des données à deux dimensions : l'une, bilatérale (IDE entrant et sortant de chacun des pays de l'OCDE avec chaque pays du monde), l'autre sectorielle (IDE entrant dans chacun des pays de l'OCDE par secteur). La base Eurostat est la seule à diffuser, pour les pays européens, des données en trois dimensions : pays investisseur-pays hôte dans la nomenclature sectorielle d'Eurostat.

Ces différentes bases de données ne sont pas directement utilisables par les modélisateurs. Celle d'Eurostat est la seule à offrir le triplet dimensionnel intéressant, mais sa couverture géographique est limitée. Par ailleurs, il existe des incohérences importantes entre les différentes bases, mais aussi, parfois, à l'intérieur d'un même ensemble de données. Nous proposons ici une méthode systématique et documentée qui permet de tirer parti de toute l'information disponible et de construire, pour une année (2004), une base de données mondiale d'IDE en trois dimensions. Les méthodes employées pour calculer les stocks d'IDE diffèrent d'une base à l'autre et il nous a semblé inutile de chercher à réconcilier les données de flux et de stocks, étant donnée la méthodologie spécifique utilisée pour construire la base de stocks. Nous appliquons donc notre méthode parallèlement sur les deux ensembles de données.

La procédure se déroule en deux étapes. Dans un premier temps, en partant de la base Eurostat en trois dimensions, nous estimons économétriquement, dans un cadre gravitationnel, des IDE « théoriques » là où les données sont manquantes. Dans un second temps, nous nous calons sur l'information disponible dans les autres bases de données. Un programme d'optimisation quadratique sous contraintes permet de minimiser les écarts entre les données estimées et les données réelles. Les contraintes obligent la matrice finale (stocks ou flux) à être cohérente avec les données réelles, au niveau global (FMI, CNUCED), bilatéral et sectoriel-unilatéral (OCDE). Pour compléter les données des organisations internationales, nous avons utilisé les données du WiiW (The Vienna Institute for International Economic Studies) pour les pays d'Europe de l'Est et celles fournies pour la Chine par le Statistical Yearbook of China. D'autres ensembles de données-source peuvent être utilisés pour améliorer l'estimation économétrique des données ou l'équilibrage des matrices.

Nous proposons ainsi une base de données directement utilisable par les modélisateurs intéressés par la problématique des investissements directs à l'étranger ou, plus généralement, pour intégrer cette dimension dans leurs analyses quantitatives.

RESUME COURT

Nous décrivons ici la méthodologie utilisée pour construire une base de donnée mondiale d'investissements direct à l'étranger en trois dimensions (pays investisseur, pays hôte et secteur), pour l'année 2004. À partir des données Eurostat, nous estimons des investissements théoriques pour l'ensemble des pays du monde. Nous contraignons ensuite nos estimations avec les données existantes de dimensions inférieures (1 ou 2) lors de l'équilibrage de la matrice, via une optimisation quadratique. Cette base de données est destinée à une utilisation pour des études quantitatives (modèles d'équilibre général).

Classification JEL : C 68, C 82, F 21

Mots-clefs : Modèles calculables d'équilibre général, Investissement direct à l'étranger, bases de données.

A FOREIGN DIRECT INVESTMENT DATABASE

FOR GLOBAL CGE MODELS

Christophe Gouel
Houssein Guimbard
David Laborde

1. INTRODUCTION

During the last twenty years, globalisation has mainly taken place through trade in goods and capital flows. Among the latter, foreign direct investment[4] (FDI) deserves a special emphasis.

Investment issues remain important under the Doha Development Agenda (DDA), through the mandate to negotiate on trade in services under the General Agreement on Trade in Services (GATS), which includes the right of establishment in services sectors (commercial presence or Mode 3). At the WTO (World Trade Organization) Ministerial Conference in Hong Kong, where a specific negotiating process was set up, it was agreed that the negotiations should account for the development level of WTO Members. Least Developed Countries are not expected to undertake new commitments. Furthermore, investment issues are crucial in the bilateral negotiations between the European Union (EU) and third countries, in particular with ASEAN countries, India, South Korea, Mercosur, Andean Pact, Central America, Ukraine and Russia.[5]

In this context, it is essential to better understand the consequences of foreign direct investment. Empirically and theoretically, recent years have seen an important development of studies on the relationship between investment and trade. These studies, however, have not yet produced any robust tool suitable to assess *ex-ante* consequences of political decisions regarding trade and investment. Using a monopolistic competition model of trade and multinational production, Lai and Zhu (2006) suggest that a worldwide trade liberalisation would increase U.S. exports by 3% and U.S. multinational production by 21.7%. However, this large expansion of overseas production is generally ignored in traditional computable general equilibrium (CGE) models used for analysing trade policies. The main reason for this is the lack of harmonised, balanced and detailed FDI data at the world level. Previous attempts to introduce FDI in CGE[6] relied on specific FDI database. For example, Petri (1997) constructs a 6-region, 3-sector database from APEC data, and Japan and US surveys. Walmsley (2002), and Dee et al. (2001), for the FTAP model, use also APEC data to build respectively an 11-region, 8-sector database and a 19-region, 3-sector one. Contrary to previous works that focus on the Asia-Pacific relationship, Lejour et al. (2007) work on the

[4] 10.8% of annual growth rate between 1980 and 2003.

[5] ASEAN stands for "Association of Southeast Asian Nations" ; MERCOSUR for "Mercado Común del Sur".

[6] For a survey, see Lejour et al. (2006).

liberalisation of services within Europe. Hence, they use EUROSTAT and OECD data to construct their 23-country, 10-sector FDI database (van Leeuwen et al., 2006). Giving theur problematic, all these works use only a part of the available FDI information to construct their databases.

In this work, we start to fill this gap by providing FDI data dedicated to CGE modelling. FDI data are already available from several international or regional institutions, like International Monetary Fund (IMF) or Eurostat, but they are not suitable for applied general equilibrium models. Indeed, they are not balanced; they have a lot of missing values and mirror values. Lastly they may miss one of the requested dimensions (investor, host, sector). To tackle these issues, we develop a methodology that estimates the missing values with econometrics and balances the database with a quadratic optimisation method.

This paper is organised as follows. Section 2 presents the principles of construction of the database (data collection, econometrics and balancing method).[7] Section 3 provides various sensitivity analyses. Section 4 concludes.

The original dataset and its updates are freely downloadable at:
http://www.cepii.fr/anglaisgraph/bdd/fdi.htm
Two .csv files (the separator is ";") are available: one for stocks and one for flows.
The variables are:
Column A (variable "r") refers to the investor.
Column B (variable "s") refers to the host country.
Column C (variable "i") refers to the sector.
Column D (variable "val") is the value of the corresponding FDI, in million of 2004 USD.
As example, in the flow dataset, the value "val" is thus the flow that goes from country "r" to country "s" and the "i" sector, following GTAP database logic.
Geographical classification and sectoral classification are presented in annex of this document.

2. CONSTRUCTION OF THE DATABASE

In this section, we present the construction of a FDI database suitable for trade and investment policy assessment and fitting with the GTAP framework, which could relieve modellers from constructing their own database for every new study. We fully document our method and propose a solution that allows any new piece of information to be integrated. So this database is not meant to be restricted to a specific geographical or sectoral coverage; we can easily improve it when new data are made available.

[7] Datasets are available at http://www.cepii.fr/fdi

2.1. Sources

2.1.1. Original sources

Building a database on FDI requires having a large coverage of countries and sectors. National sources provide comprehensive databases, but definitions, treatments and nomenclatures differ noticeably from one country to another. For this reason, we favoured multilateral sources collected and processed by international institutions. Their main asset is to provide harmonised data in terms of definition, nomenclature and treatment. However, information cannot be obtained in the three dimensions, namely investing country, host country and sector. To get a third dimension, we need to complement with regional sources that provide, in their restricted geographical coverage, more detailed information. The detailed geographical and sectoral coverage of the various sources is displayed in Table 4, Table 5 and Table 6.

International Monetary Fund (IMF) and UNCTAD data cover nearly every country in the world. However, they provide figures in a unilateral way: foreign direct investment by host country and investment abroad by origin country, no bilateral or sectoral breakdown is available. These databases constitute our benchmark for the aggregated values. Values between the two dataset are not necessarily consistent between them and we can choose only one. Even if UNCTAD filled a lot of missing values that are present in the IMF dataset, our first inclination was to rely only on IMF data.

Indeed, for some countries, UNCTAD data are constructed on principles that would lead to inconsistencies with the other databases we use. For example, UNCTAD corrects the Luxembourg FDI flows by removing all trans-shipped FDI (i.e., investments that do not stay in the host country, but that are channelled through a special purpose entity to another destination), which reduces inflows in Luxembourg by 95%. The fund data display also inconsistencies with other information, especially for stocks (for the flows, the two databases are almost equivalent). In the case of USA, IMF declares between 2000 and 2004 more than $2,000 billion of inward stock, whereas it is always fewer than 1,700 for UNCTAD. USITC website confirms the smallest number. French data presents the same discrepancies to the detriment of IMF. However, we decide to keep IMF values as the reference and to keep UNCTAD data when IMF does not provide any figure.

Through a joint questionnaire, **OECD and Eurostat** provide databases complementing the two previous ones. The geographical coverage remains partial but looking at all investments coming and leaving OECD and European countries allows capturing the broad picture of FDI in the world (it represents more than 75 % of FDI). Moreover, OECD is deeply involved with the IMF in defining methodology for FDI data collection (see for example the Survey of implementation of methodological standards for direct investment, SIMSDI [OECD/IMF, 1999]). OECD[8] provides figures on bilateral and sectoral flows and positions (but not in three

[8] OECD International Direct Investment Statistics extracted on August 06, 2007.

dimensions). It covers 30 countries as reporter and nearly all countries as partner. Eurostat provides FDI data by industry and by country of origin and destination. The Eurostat database[9] constitutes the core of our detailed information. It covers, at the detailed level, 15% of world flows.

We also include data from **WiiW** [10] which covers twenty Central, East and Southeast Europe countries. Relying on published data from the National Banks of the FDI host countries, this database provides bilateral information based on national data for years 1990 to the most recent years, both for stocks and flows. Finally, we use data from the **Statistical Yearbook of China.**[11] The statistical yearbook of the National Bureau of Statistics of China provides annual statistics and historical data at the national level and, sometimes, at the local level of provinces. The 2006 yearbook contains bilateral and sectoral information on China for 2004 and 2005.

2.1.2. Some issues concerning FDI statistics

a) Discrepancies between reported flows

Discrepancies between global inflows and outflows are large. At the worldwide level, inflows are always larger than outflows.

Figure 1 shows difference between global inflows and outflows FDI for three different datasets: IMF, OECD and Eurostat. At the aggregated level, differences between datasets come from the different number of reporter countries: Eurostat concerns European countries as reporters, OECD only reports data for OECD members. IMF data concern almost all countries in the world. In 2001, in IMF database, global inflows exceed by 72.7 billion of dollars global outflows (see Figure 1).

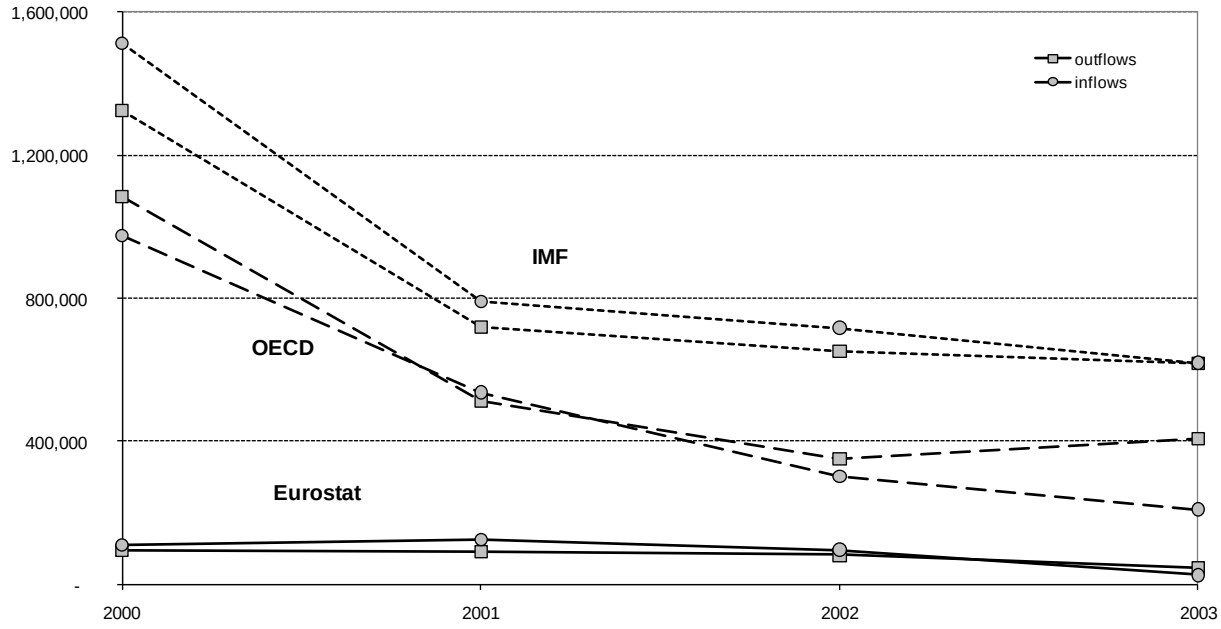
The joint work IMF/OECD has attributed the discrepancy between two mirror flows and so between global outflows and global inflows to a variety of reasons, from which the most notable are: (1) failure to compile data on reinvested earnings; (2) failure to follow international standards in relation to short-term financing between affiliated enterprises; (3) failure to record and properly classify the activities of "Special Purpose Entities"; (4) failure to record cross-border real estate transactions; and (5) failure to properly classify investment by affiliates in their parent companies. For a CGE model, the final database must be balanced, so we need to have a unique value for total FDI flow (as well as only one value for stocks). This will be addressed later in the methodological presentation.

[9] Extracted on July 07, 2007.

[10] The CD-ROM (2007) we used contains the WIIW's "Database on Foreign Direct Investment in Central, East and Southeast Europe".

[11] We used the 2006 CD-ROM of the Statistical Yearbook of China.

Figure 1 - Discrepancies between global inflows and outflows FDI, in millions of USD



Source: data come from IMF, OECD and Eurostat databases, Author's calculations.

b) Dealing with mirror values

FDI flows or positions reported by one country do not usually coincide with those reported by its partner. This problem concerns only bilateral databases, Eurostat and OECD. In fact, we often have only one value for a FDI in these data because, as they are from regional institutions, only European and OECD countries are reporting countries in the two databases. So, mirror values only concerns FDI between European or OECD countries. In the Eurostat database, 9.1% of flows observations are mirror flows, and 47.7% of them have the same values. For the other data, differences simply cannot be explained. 24% have opposite signs, and the differences between the two flows can amount to billions of dollars. For example, in 2001 for the sector of activity *Refined petroleum and other treatments*, USA reported 2 million euros of FDI from Netherlands, whereas, for the same flow, Netherlands reported a disinvestment of 12.4 billion euros. The median discrepancy is 21 million euros, but one-fourth of mirror flows show a discrepancy that exceed 106 million of euro.

The mirror values have, in part, the same origin than the discrepancy at the world level. "Many countries still deviate one way or another from the recommendations of the IMF and the OECD in their collection, definition and reporting of FDI data." (UNCTAD, 2005)

Concerning Eurostat data, we choose the inward values as reference because we are primarily interested in the sectoral inward data, which are more likely to be better reported by the host economy than by the investing country. Indeed, the host country will report the ultimate

sector of investment that can be different of the sector of origin. For example, in the case of a vertical FDI, the production chain is split between several countries; the host country produces a good different from the home country. Hence, the sector of origin is not the same than the sector of destination. The OECD dataset for country-country FDI also contains mirror values. To harmonise them, we use a simple average when both values are available.[12]

c) Dealing with negative values

Original sources present a lot of negative values,[13] for flows or stocks. According to UNCTAD website:[14] "FDI flows with a negative sign indicate that at least one of the three components of FDI (equity capital, reinvested earnings or intra-company loans) is negative and not offset by positive amounts of the remaining components. These are instances of reverse investment or disinvestment." Negative flows have real economic meaning, and, because of their numerical importance, we cannot get rid of them without losing all consistency. In the contrary, negative stock values are generally the consequences of accounting methods (they also be recorded when continuous losses in the direct investment enterprise lead to negative reserves) and we will treat them as zero.

2.2. Methodology

The FDI database construction involves two steps. Those are the same for stocks and flows but we treated them in two separated datasets because of the impossibility of reconciling them.

In the first step, we estimate all FDI relationships (flows and stocks) through a gravity equation, using Eurostat data (FDI as dependent variable). Adding to this estimated dataset our Eurostat data leads to a complete (but predicted) database for FDI.

In the second step, we finalise this database by modifying predicted values, trying to respect as far as possible values (Eurostat but also data of higher dimension) from reliable sources (Eurostat, OECD, IMF...). This section presents the procedure followed for the construction.

[12] A tentative version (presented at the 10th GTAP Conference, 2007) of the dataset, balanced with cross-entropy methods, treated the mirror values harmonization as a part of the matrix-balancing problem. The right value between the two mirror flows was the most consistent with the other information according to a metric.

[13] Negatives values account for 8% of Eurostat flows, and 1% for stocks. Even aggregate flows show negative values; IMF indicates a negative inflow of 3 billions of dollars in 2001 for Indonesia, one of the most affected countries by the East Asian financial crisis.

[14] http://www.unctad.org/Templates/Page.asp?intItemID=3153

2.2.1. Econometrics

The dataset built from the available databases has numerous missing values. Before balancing it to satisfy equilibrium constraints, we prefer to have rough estimates of these missing values. We rely on the gravity framework and on available data (Eurostat) to get these estimates. The gravity equation has long been used to analyse bilateral relationships such as migrations, trade or financial flows. Theoretical justifications for using this equation with FDI have been recently suggested in some papers. Bergstrand and Egger (2007) expand Markusen's knowledge-capital model by adding a third factor, capital, to skilled and unskilled labour. By introducing a third country in the model, they achieve to mimic the behaviour of a gravity-like[15] equation. Head and Ries (2007) see FDI as an outcome of the market for corporate control. FDI are explained by size and distance variable. The home country's size is the share of world's bidders. The host country's size is the asset value of the entire stock of targets. The distance is related to the cost of remote inspections, which increases with physical, but also cultural distance. Eventually, Kleinert and Toubal (2010) show that the very good results, obtained in empirical studies, from the gravity equation are due to the fact that it can be derived from various theoretical models of multinational firms.

Those studies, using the gravity equation, find that home and host country's market size have a positive effect on FDI flows, while the effect of distance between the two countries is negative. Our regression is thus theoretically funded for FDI flows. But we also need a prediction for FDI stocks. The logic at work behind FDI stocks behaviour is not necessary the same, but we decide to keep the same specification (gravity). We apply the same methodology and focus on the fit of the regression.

Data Issues

We estimate the relationship between FDI and other variables using two databases from Eurostat, one on flows and the other on stocks. They extend from 1986 to 2005 for flows and from 1994 to 2004 for stocks. The original sectoral coverage includes aggregated sectors. However, we keep only detailed information for our work (see Table 5**Table 6** for the list).

They present a lot of zero and negative values (Table 1).

As explained, negative flows are instances of disinvestment. We include them as prior values for our final database, but we keep only zero and positive values for the regressions.

As negative stocks do not have a real economic meaning, they are dropped from the final database for the estimations.

[15] Because of the highly non-linear relationships between the variables, there is no closed-form solution.

Table 1 - Signs of FDI values: Number of values (percent)

	Negative	**Zero**	**Positive**
Flows	5,520 (8.0%)	49,620 (72.2%)	13,586 (19.8%)
Stocks	731 (1.0%)	51,071 (67.7%)	23,646 (31.3%)

Source: Eurostat database

Following statistics (Table 2) are from the data used for the regressions: They come from Eurostat FDI database after exclusion of negative values, mirror outflows values and some points for which we lack explicative variables. Number of observations is increasing with the time, especially within recent years (see Table 2).

Table 2 - Number of observations per year

	1986	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	
Flows	236	226	176	190	164	166	181	149	576	1,785	1,537	
Stocks	0	0	0	0	0	0	0	0	633	2,819	2,444	
			1997	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**Total**
Flows			1,739	1,600	2,436	2,649	5426	8,877	9,763	12,024	9,906	59,806
Stocks			3,654	5,018	6,022	6,330	6,800	6,089	13,076	15,345		68,230

Source: Eurostat database (data used for the regressions)

FDI values are very dispersed: A large part is below $100 million and few points are above $10 billion. They also have a lot of zeroes, leading to the use of specific econometrics methods to estimate missing values at the sectoral level.

Estimation of missing values

We use the following gravity equation to explain the FDI stocks and flows:

$$\begin{aligned}\ln FDI_{rs}^{it} = {} & \alpha_1 + \alpha_2 \ln GDP_r^t + \alpha_3 \ln GDP_s^t + \alpha_4 \ln GDPCAP_r^t + \alpha_5 \ln GDPCAP_s^t \\ & + \alpha_6 \ln DIST_{rs} + \alpha_7 \ln COMLG_{rs} + \alpha_8 \ln COL_{rs} \\ & + \alpha_9 \ln DVing_r + \alpha_{10} \ln DVing_s + \beta^t year^t + \delta^i sector^i + \varepsilon_{rs}^{it}, \quad (1)\end{aligned}$$

where i, r and s stand respectively for sectors, origin countries and destination countries. $COMLG_{rs}$ stands for the use of a common official language between countries r and s; COL_{rs} is the dummy introducing a colonial link. By representing a geographical or cultural proximity, they can account for transaction costs not well represented by distance, $DIST_{rs}$.[16]

[16] Distance, contiguity, language and colonial link are taken from CEPII databases on bilateral distances, available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

We use also the GDP per capita[17] (*GDPCAP*) as a proxy of the level of development, a potential driver of inward and outward FDI flows. We introduce time dummies to catch a time evolution in FDI and sector dummies to represent heterogeneity of the various sectors. We use this equation, on Eurostat FDI data pooled over 1986-2005 (1994-2004 for stocks) after keeping inward flows or stock for non-identical mirror values.

We estimate the previous equation with a Poisson quasi-maximum-likelihood estimation (Santos Silva and Tenreyro, 2006). So FDI are taken in level and not in logarithm, which allows the inclusion of zero values (70% of all values). The negative values are dropped; they cannot be included in this framework: we may lose some information but we gain in robustness with more appropriate econometrics.

With a Poisson quasi-maximum-likelihood estimation, (1) becomes:

$$FDI_{rs}^{it} = \exp\big(\alpha_1 + \alpha_2 \ln GDP_r^t + \alpha_3 \ln GDP_s^t + \alpha_4 \ln GDPCAP_r^t + \alpha_5 \ln GDPCAP_s^t + \alpha_6 DIST_{rs} + \alpha_7 \ln COMLG_{rs} + \alpha_8 COL_{rs} + \alpha_9 \ln DVing_r + \alpha_{10} \ln DVing_s + \beta^t year^t + \delta^i sector^i\big) \times \varepsilon_{rs}^{it}. \quad (2)$$

Results are presented in Table 3.

[17] GDP, in current USD, and population are taken from CEPII database CHELEM.

Table 3 - PQMLE on FDI – Year and Sector Fixed Effects

Variables	Stocks	Flows
Intercept	-69.67a	-70.582a
	(1.058)	(1.972)
ln origin GDP	0.775a	0.717a
	(0.009)	(0.017)
ln destination GDP	0.858a	0.804a
	(0.009)	(0.018)
ln origin GDP per cap.	2.995a	2.781a
	(0.061)	(0.122)
ln destination GDP per cap.	2.417a	2.688a
	(0.061)	(0.115)
ln Distance	-0.756a	-0.671a
	(0.01)	(0.02)
Common language	0.683a	0.598a
	(0.029)	(0.057)
Colonial link	0.136a	0.057
	(0.033)	(0.065)
Developing (origin)	-0.345	-0.199
	(0.215)	(0.284)
Developing (destination)	1.558a	1.829a
	(0.086)	(0.133)
Nb. Obs.	68230	59806
R2 in levels	0.384	0.19
RMSE in levels	1551.269	397.206

Note: Standard errors in parentheses. "a" denotes significance at the 1% level

We are mainly interested in the overall quality of our regression. Almost all coefficients are significant at 1% and their signs correspond with theoretical predictions (those associated with GDP are positive and negative for the distance, for example). We retain those estimated by the technique PQMLE[18] because the measure of fit of this regression is the highest (R^2 of 0.384 for the stocks regression; 0.19 for the flows' one).

Using this regression, we can build, at the end of this first step, a database on FDI in three dimensions for almost all countries in the world. Values of those FDI correspond to predicted values from regression.

[18] See the sensitive analysis in annex.

A set of 19 countries[19] do not appear in the final database, while they are present in the original data at the aggregated level (IMF-UNCTAD). These are mainly tax heavens, overseas territories or territorial enclaves. Indeed, for them, we do not have the necessary data (e.g., GDP) to estimate FDI through our gravity model.

2.2.2. Matrix balancing

The first step of the construction provides us with a first version of our base in three dimensions. We move on to the second step which consists in balancing our estimates with existing data of inferior dimensions (1 or 2).

In other words, all our information does not constitute a consistent database. Indeed, we obtain values inconsistent with country inflows and outflows from the IMF, or with bilateral information from OECD, WiiW or China's National sources. In order to obtain the desired database, we have to harmonise all these pieces of information into a single framework. This problem is quite common in CGE works, because estimating a social accounting matrix consists in finding a way to reconcile information from a variety of sources.[20]

Since our goal is to provide a database for CGE modellers, we want to propose a dataset that does not display too much volatility. Indeed, CGE models are usually used to assess mid-term or long-term policies. Hence, calibration is a crucial issue and results could be very sensitive to a particular configuration of the data not representative of structural issues, at the base year. So, we average the data in order to avoid such a problem, especially for flows, subject to high volatility.

Since a part of FDI flows is highly volatile, we rather rely for flows on a three-year average (2003-2005, when available), which smoothes the short-run volatility. Averaging will also help to limit the number of negative values, which can be an issue in CGE modelling.

A quadratic optimisation[21] procedure is then implemented to insure consistency of the database. As it would be very difficult to harmonise stocks and flows, two distinct databases, one for flows, the other for stocks, are built independently with the same methodology. It is important to keep in mind that both may present some inconsistencies on a chronological perspective. Indeed, no mechanism insures that stocks in year t will be based on stocks in $t-1$ and flows in $t-1$.

[19] These are : Anguilla, Cayman islands, Cook islands, East Timor, Falkland islands, French Polynesia, Gibraltar, Korea Dem. people's rep, Montserrat, New Caledonia, Niue, Northern Mariana islands, Palau, Palestinian territory, Swaziland, Tokelau, Turks and Caicos islands, Tuvalu, Virgin islands (British).

[20] For example, the construction of the GTAP database involves the aggregation of protection, trade, input-output or energy data.

[21] After numerous tests, we decided to deal the balancing problem by using a quadratic method which provides the best results both in terms of simplicity of implementation (few assumptions are needed) and rapidity of resolution. A version of this dataset has been built using a generalised cross-entropy method, minimising the divergence between prior and final FDI values.

Our objective is to minimise the discrepancies between the original and final values subject to consistency constraints and aggregate constraints. Our problem is thus stated as follows:

Minimise

$$\left(\sum_{irs}(\mathrm{FDI}_{irs}-\mathrm{FDI0}_{irs})^2/\mathrm{w}_{irs} + \sum_{r}(\mathrm{FDI}_{r}-\mathrm{FDI0}_{r})^2/\mathrm{w}_{r} + \sum_{s}(\mathrm{FDI}_{s}-\mathrm{FDI0}_{s})^2/\mathrm{w}_{s} + \sum_{rs}(\mathrm{FDI}_{rs}-\mathrm{FDI0}_{rs})^2/\mathrm{w}_{rs} + \sum_{is}(\mathrm{FDI}_{is}-\mathrm{FDI0}_{is})^2/\mathrm{w}_{is}\ \text{(in)} + \sum_{is}(\mathrm{FDI}_{is}-\mathrm{FDI0}_{is})^2/\mathrm{w}_{is}\ \text{(out)}\right)$$

Subject to

$$\sum_{is}\mathrm{FDI}_{irs} = \mathrm{FDI}_{r} \qquad \text{for all r (Total outward)}$$

$$\sum_{ir}\mathrm{FDI}_{irs} = \mathrm{FDI}_{s} \qquad \text{for all s (Total inward)}$$

$$\sum_{i}\mathrm{FDI}_{irs} = \mathrm{FDI}_{rs} \qquad \text{Bilateral}$$

$$\sum_{r}\mathrm{FDI}_{irs} + \sum_{r}\mathrm{FDI}_{isr} = \mathrm{FDI}_{is} \qquad \text{Sectoral Unilateral}$$

$$\sum_{r}\mathrm{FDI}_{r} = \sum_{s}\mathrm{FDI}_{s} \qquad \text{Total World}$$

FDI0 is the original data, that is the combination of the raw data and the estimated values. FDI stands for the final data resulting from the quadratic optimisation process that satisfies the different constraints. *w* are item-specific reliability weights. They represent the confidence we have in the data. Choosing *w* is not an easy task and requires knowledge about the potential errors in the raw data. We try to be as objective as possible, allowing more flexibility to estimated values than to raw data.

As we have already shown, FDI data are full of inaccuracies. To account for this, we use as benchmark the minimum variance of estimated data. For collected data, we assume that they cannot vary as much as estimated one. To do that, we applied a multiplier coefficient to the coefficient of variation of estimated data. We use 0.01 for Eurostat data (since we consider Eurostat information as quite reliable, we allow only small variations around these values. However, we do assume that they are not exempt of inaccuracies) and 0.0001 for data in one dimension (because aggregated data must contain less errors than detailed data). For bilateral data, we use an intermediary value of 0.001. This gives us our weights w.

Although negative FDI flows represent a significant phenomenon, they are not usually consider in microeconomic theory of FDI, and we do not have any theoretically sound way to include them into a CGE model. Before incorporating FDI data into a model, one may need to get rid of all negative values. For this purpose, we also propose a balanced matrix without negative values.[22]

In practice, this formulation accounts for information coming from all raw data. Our final database has 26 sectors (cf. Table 5) and more than 180 countries (see the geographical breakdown in Table 6).

2.2.3. Differences with original values

The principles used for the construction of our dataset will necessarily lead to changes in the original values to satisfy the various constraints. A large share of the original data is stable: lot of them does not vary by more than 1%. Indeed, our weighting system (the same for stocks and flows) aims, in a sense, at minimizing large discrepancies between the raw data and the final ones.

Stocks data

At the aggregate level, we decided to stay close to the original data. Indeed, these are reported by countries and it is necessary not to stray too far from the overall image provided by the original data. 132 (202 in total) country's inward stocks vary by less than 1%. 30 experienced a variation by less than 10%. Only two see their inward stocks rising by more than 90% (Liberia and Myanmar). Countries experiencing a fall by 100% are those that declare FDI to IMF or CNUCED but for which we do not have predicted values (lack of explanatory variables).

[22] We initialize negative values to zero and apply the same balancing methods to these data. CEPII website also provides this other version of the database.

Bilateral data (OECD, WiiW and China) contain 2,390 observations. Few points vary by more than 100% (4). Almost 15% of the observations have a range of variation between 10% and 90%. Around 62% do not move by more than 1%.

Nearly all values from Eurostat are very similar to their original values. 1,712 observations (on 2,381) vary by less that 1% (72% of the Eurostat data). 81 observations experience variations exceeding 50%. Few vary by more than 400%.

Flows data

As for stocks, obtained data on flows do not display too much variations.

At the aggregated level, excepting the 19 countries for which we do not have predicted data, only 12 countries see their FDI inflows changing by more than 1% (from which USA: +5%, UK: +5%). Only Congo (-29%) and Liberia (-12%)'s FDI inflows fall by more than 5.5%.

Bilateral data contains 3,436 observations. Our methodology leads to a variation higher than 10% for 402 values (267 increase whereas 135 decrease by 100%, corresponding to countries for which global predicted FDI are null). 41 pairs vary by more than 10%.

Eurostat data exhibit 2,651 observations. The majority is closed to its original value (variation inferior to 1%). 126 triplets increase more than 1% but less than 10%. In the same interval, 115 triplets see their values decreasing. Finally, only 16 values vary by more than 100%.

3. CONCLUSION

In this work, we have proposed a framework to construct a harmonised FDI database suitable for CGE modeling. It fills a gap, because, until now, most trade policy analysis missed effects due to multinational firms' activity due to the lack of consistent data. This paper describes a first release of the database. The econometrics and the developed optimisation procedure allow accounting for any other sources of information (The design principles make it useless for econometrics). When data are collected for a specific country or regions, for a PTA study for example, the whole dataset could be updated, improving its quality.

However, some issues may still be of concerns. As we have used FDI information provided by countries to international institutions, they do not always represent the real flows that interest us in CGE modelling. Some flows are trans-shipped to another country and do not represent real flows. A first case happens when a country treats differently domestic and foreign investors. If foreign investors are favoured, then domestic investors have the incentive to export their funds before importing them back (round-tripping). This particularly concerns China: Chinese firms use their affiliates in Hong Kong to benefit from special treatment when investing in China.[23] The second case concerns the countries that are only used for tunnelling

[23] According UNCTAD (2003), round-tripping accounts for 25% of FDI in China.

FDI to another destination. MNFs trans-ship their FDI because they may want to locate profits in a country with low taxes. So it concerns mainly tax havens and offshore financial centres.[24]

The database has been already used in different works involving a new modelling of investment decisions performed by multinational firms and on effects of bilateral investment treaties. This approach have been used by Laborde and Lakatos (2009) to measure the political economy effects of foreign investments in trade policy incentives; by Bouet, Laborde and Lakatos (2010) to assess the role of FDI in shifting the factor endowments of countries in the baseline of a CGE; and by Bouet, Estrades and Laborde (2010) for a policy oriented analysis on regional integration between Asia and Latin America. Chappuis and Decreux (2011) modify the demand side of the MIRAGE model (Decreux and Valin, 2007), differentiating goods according to their origin of capital, and simulate the consequences of a reduction of a tax on capital.

[24] For Luxembourg, UNCTAD (2006) estimates that 95% of FDI inflows during 2002-2005 were trans-shipped[24].

REFERENCES

Bergstrand J. H., Egger P. (2007), A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises, *Journal of International Economics*, in press.

Bouet A., Laborde D., Lakatos C. (2010), Economic baselines for CGE modelling of trade liberalization in the world economy: a sensitivity analysis, prepared for the workshop on dynamic baseline for CGE in Shangai.

Bouet A., Estrades C., Laborde D., (2010), Trade and Investment in Latin America and Asia: Lessons from the Past and Potential Perspectives from Further Integration, in the book "*Opportunities for Mutual Learning in Asia and Latin America: Macroeconomic Growth and Reductions in Poverty and Hunger*".

Chappuis T., Decreux Y., Applied general equilibrium modelling of FDI: illustration with the Mirage model, *CEPII Working Paper*, Forthcoming 2012

Decreux Y.,Valin H. (2007), MIRAGE, Updated Version of the Model for Trade Policy Analysis: Focus on Agriculture and Dynamics, *CEPII Working Paper*, N°2007-15.

Dee P., Hanslow K. (2001), Multilateral Liberalization of Services Trade, in Robert M. Stern, ed., *Services in the international economy*, University of Michigan Press, pp. 117--139.

Feenstra R.C. (2004), *Advanced international trade: Theory and evidence*, Princeton University Press.

Gaulier G.; Zignago S.; Sondjo D.; Sissoko A.A., Paillacar R. (2007), BACI: A World Database of International Trade Analysis at the Product-level 1995-2004 Version, *CEPII Working Paper*.

Gehlhar M. (1996), Reconciling Bilateral Trade Data for Use in GTAP, *Technical report, GTAP*, 10.

Golan A., Judge G., Miller D. (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Wiley.

Golan A., Judge G., Robinson S. (1994), Recovering Information from Incomplete or Partial Multisectoral Economic Data, *Review of Economics and Statistics* **76**(3), 541--549.

Head K., Ries J. C. (2007), FDI as an Outcome of the Market for Corporate Control: Theory and Evidence, *Journal of International Economics*, in press.

Kleinert J., Toubal F., (2010), Gravity for FDI, Review of International Economics, Wiley Blackwell, vol. 18(1), pages 1-13, 02.

Lai H., Zhu S.C. (2006), U.S. Exports and Multinational Production, *The Review of Economics and Statistics* **88**(3), 531-548.

Laborde D., Lakatos C. (2009), Does Foreign Investment Shape Trade Policies? A CGE Assessment of the Doha Stalemate, GTAP Conference 2009.

Lejour A., Rojas-Romagosa H. (2006), Foreign Direct Investment in Applied General Equilibrium Models, *CPB Memoranda*, 169.

Lejour A., Rojas-Romagosa H., Verweij G. (2007), Opening Services Markets within Europe: Modelling Foreign Establishments in a CGE Framework, *CPB Discussion Papers*, 80.

Markusen J. R., (2002), Multinational firms and the theory of international trade, *MIT Press*.

OECD/IMF (1999), Report on the Survey of Implementation of Methodological Standards for Direct Investment, (DAFFE/IME(99)14).

Petri P.A. (1997), Foreign Direct Investment in a Computable General Equilibrium Framework, Brandeis-Keio Conference on "Making APEC Work: Economic Challenges and Policy Alternatives", Keio University, Tokyo.

R Development Core Team (2005), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

Robinson S., Cattaneo A., El-Said M. (2001), Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods, *Economic Systems Research* **13**(1), 47--64.

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

Smith A. (1987), Strategic Investment, Multinational Corporations and Trade Policy, *European Economic Review*, **31**(1/2), 89--96.

UNCTAD (1996), *World Investment Report 1996. Investment, Trade and International Policy Arrangements*, United Nations Publications.

UNCTAD (2003), *World Investment Report 2003. FDI Policies for Development: National and International Perspectives*, United Nations Publications.

UNCTAD (2005), *World Investment Report 2005. Transnational Corporations and the Internationalization of R&D*, United Nations Publications.

UNCTAD (2006), *World Investment Report 2006. FDI from Developing and Transition Economies: Implications for Development*, United Nations Publications.

van Leeuwen N., Lejour, A. (2006), Bilateral FDI Stocks by sector, *CPB Memoranda*, 164.

Wächter A., Biegler L. T., On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, *Mathematical Programming* 106(1), pp. 25-57, 2006

Walmsley T.L. (2002), Incorporating international ownership of endowments into a global applied general equilibrium model, *Economic Modelling*, 19(5): 679--707.

APPENDIX A: COVERAGE OF THE GLOBAL SOURCES

Table 4 - Coverage of sources used in the database construction

Level of aggregation	Source	Type	Information		2004	Total value for 2004 ($ billion)
Global	IMF	Flows	Inward		150	738
			Outward		108	931
		Stocks	Inward		97	12042
			Outward		84	12413
	UNCTAD	Flows	Inward		205	686
			Outward		181	802
		Stocks	Inward		216	9449
			Outward		152	10188
Bilateral	OECD	Flows	Inward	Origin	219	480
				Destination	29	
				Number of pairs	3476	
			Outward	Origin	29	617
				Destination	219	
				Number of pairs	3565	
		Stocks	Inward	Origin	219	5374
				Destination	23	
				Number of pairs	2602	
			Outward	Origin	23	6537
				Destination	219	
				Number of pairs	2647	
Sectoral inward	OECD	Flows	Number of countries		26	287
			Average number of sectors per country		24	
		Stocks	Number of countries		20	4633
			Average number of sectors per country		26	
Sectoral bilateral	Eurostat[25]	Flows	Number of origin countries		69	63
			Number of destination countries		63	

[25] After treatment of mirror values.

APPENDIX B: SECTORAL AND GEOGRAPHICAL BREAKDOWN

We have included in the database sectors that do not overlap in order to be able to balance the database without counting two times the same flow/stock. We start from the sectoral breakdown used in the Eurostat / OECD questionnaire. In the table below, the sectors introduced in the database are those that are matched with a Eurostat code. The sectors that can be made consistent with GTAP sectors have been aggregated (for example, sectors ELE, MVH and TRD). Table 5 presents the sectoral classification and Table 6 the geographical breakdown.

Table 5 - List of sectors included in the FDI database

Economic activity	Eurostat code	FDI Database	ISIC Rev.3	GTAP 7
Agriculture and fishing	0595	0595	1-2, 5	1-14
Mining and quarrying	1495	1495	10-14	15-18
Manufacturing				
Food products	1605	1605	15-16	19-26
Total textiles & Wood				
Textiles and wearing apparel	1805	1805	17-18	27-28
Wood, publishing and printing	2205	2205	20-22	30-31
Total petroleum, Chemicals, Rubber & Plastic products				
Refined petroleum & other treatments	2300	2300	23	32
Chemical products	2400	2400	24	27, 33
Rubber and plastic products	2500	2500	25	33
Total Metal & Mechanical				
Metal products	2805	2805	27-28	35-36
Mechanical products	2900	PoOME	29	41
Total Office machinery & Radio				
Office machinery and computers	3000	ELE	30	40
Radio, TV, communication equipments	3200	ELE	32	40
Medical, precision and optical instruments, watches and clocks	3300	PoOME	33	41
Total Motor vehicles & Other transport				
Motor vehicles	3400	MVH	34	38
Other transport equipments	3500	OTN	35	39
Manufacturing n.i.e.	3990	3990	19, 26, 31, 36, 37	29, 34, 41-42
Electricity, gas and water	4195	4195	40-41	43-45
Construction	4500	CNS	45	46
Services				
Trade and repairs	5295	TRD	50-52	47
Hotels and restaurants	5500	TRD	55	47
Transport and communication				
Land transport	6000	OTP	60	48
Water transport	6100	WTP	61	49
Air transport	6200	ATP	62	50
Supporting and auxiliary transport activities; activies of travel agencies	6300	OTP	63	48
Post and telecommunications	6400	CMN	64	51
Financial intermediation				
Financial intermediation,except insurance and pension funding	6500	OFI	65	52
Insurance and pension funding, except compulsory social security	6600	ISR	66	53
Activities auxiliary to financial intermediation	6700	OFI	67	52
Real estate & business services				
Real estate	7000	OBS	70	54
Renting of machinery and equipment without operator and of personal and household goods	7100	OBS	71	54
Computer activities	7200	OBS	72	54
Research and development	7300	OBS	73	54
Other business activities	7400	OBS	74	54
Other services	9995	9995	L, M, N, O, P, Q / 75, 80, 85, 90-93, 95, 99	55-56
Priv. purchases & sales of real estate	9998	DWE		57

Note: n.i.e. stands for not included elsewhere

Table 6 - List of the countries in the FDI database

List of countries (226)		IS	OS	IF	OF	List of countries (226)		IS	OS	IF	OF	List of countries (226)		IS	OS	IF	OF	List of countries (226)		IS	OS	IF	OF
ISO	*Name*					*ISO*	*Name*					*ISO*	*Name*					*ISO*	*Name*				
004	Afghanistan	1	0	1	0	231	Ethiopia	1	0	1	0	454	Malawi	1	1	1	0	678	Sao tome and principe	1	0	1	0
008	Albania	1	1	1	1	232	Eritrea	1	0	1	0	458	Malaysia	1	1	1	1	682	Saudi arabia	1	1	1	1
012	Algeria	1	1	1	1	233	Estonia	1	1	1	1	462	Maldives	1	0	1	1	686	Senegal	1	1	1	1
016	American samoa	0	0	0	0	234	Faroe islands	0	0	0	0	466	Mali	1	1	1	1	690	Seychelles	1	1	1	1
020	Andorra	0	0	0	0	238	Falkland islands	0	0	0	0	470	Malta	1	1	1	1	694	Sierra leone	1	0	1	1
024	Angola	1	1	1	1	242	Fiji	1	1	1	1	474	Martinique	0	0	0	0	699	India	1	1	1	1
028	Antigua and barbuda	1	0	1	0	246	Finland	1	1	1	1	478	Mauritania	1	1	1	1	702	Singapore	1	1	1	1
031	Azerbaijan	1	1	1	1	251	France	1	1	1	1	480	Mauritius	1	1	1	1	703	Slovakia	1	1	1	1
032	Argentina	1	1	1	1	254	French guiana	0	0	0	0	484	Mexico	1	1	1	1	704	Viet nam	1	0	1	1
036	Australia	1	1	1	1	258	French polynesia	0	0	0	0	490	Taiwan	1	1	1	1	705	Slovenia	1	1	1	1
040	Austria	1	1	1	1	262	Djibouti	1	0	1	0	492	Monaco	0	0	0	0	706	Somalia	1	0	1	0
044	Bahamas	1	1	1	0	266	Gabon	1	1	1	1	496	Mongolia	1	0	1	1	710	South africa	1	1	1	1
048	Bahrain	1	1	1	1	268	Georgia	1	0	1	1	498	Moldova, rep.of	1	1	1	1	716	Zimbabwe	1	1	1	1
050	Bangladesh	1	1	1	1	270	Gambia	1	1	1	1	500	Montserrat	0	0	0	0	724	Spain	1	1	1	1
051	Armenia	1	1	1	1	275	Palestinian territory	0	0	0	0	504	Morocco	1	1	1	1	736	Sudan	1	0	1	0
052	Barbados	1	1	1	1	276	Germany	1	1	1	1	508	Mozambique	1	1	1	1	740	Suriname	0	0	1	0
056	Belgium	1	1	1	1	288	Ghana	1	1	1	1	512	Oman	1	1	1	1	748	Swaziland	0	0	0	0
060	Bermuda	1	1	1	1	292	Gibraltar	0	0	0	0	516	Namibia	1	1	1	1	752	Sweden	1	1	1	1
064	Bhutan	1	0	1	0	296	Kiribati	1	0	1	0	520	Nauru	0	0	0	0	757	Switzerland	1	1	1	1
068	Bolivia	1	1	1	1	300	Greece	1	1	1	1	524	Nepal	1	0	1	0	760	Syrian arab republic	1	0	1	0
070	Bosnia and herzegovina	1	1	1	1	304	Greenland	0	0	0	0	528	Netherlands	1	1	1	1	762	Tajikistan	1	0	1	0
072	Botswana	1	1	1	1	308	Grenada	1	0	1	0	530	Netherland antilles	0	1	1	1	764	Thailand	1	1	1	1
076	Brazil	1	1	1	1	312	Guadeloupe	0	0	0	0	533	Aruba	1	1	1	1	768	Togo	1	0	1	1
084	Belize	1	1	1	1	316	Guam	0	0	0	0	540	New caledonia	0	0	0	0	772	Tokelau	0	0	0	0
090	Solomon islands	1	0	1	0	320	Guatemala	1	1	1	1	548	Vanuatu	1	1	1	1	776	Tonga	1	0	1	0
092	Virgin islands (british)	0	0	0	0	324	Guinea	1	1	1	0	554	New zealand	1	1	1	1	780	Trinidad and tobago	1	1	1	1
096	Brunei darussalam	1	1	1	1	328	Guyana	1	1	1	0	558	Nicaragua	1	1	1	1	784	United arab emirates	1	1	1	1
100	Bulgaria	1	0	1	1	332	Haiti	1	1	1	1	562	Niger	1	1	1	1	788	Tunisia	1	1	1	1
104	Myanmar	1	0	1	0	340	Honduras	1	0	1	1	566	Nigeria	1	1	1	1	792	Turkey	1	1	1	1
108	Burundi	1	1	1	1	344	Hong kong	1	1	1	1	570	Niue	0	0	0	0	795	Turkmenistan	1	0	1	0
112	Belarus	1	1	1	1	348	Hungary	1	1	1	1	574	Norfolk Island	0	0	0	0	796	Turks and caicos islands	0	0	0	0
116	Cambodia	1	1	1	1	352	Iceland	1	1	1	1	579	Norway	1	1	1	1	798	Tuvalu	0	0	0	0
120	Cameroon	1	1	1	1	360	Indonesia	1	0	1	1	580	Northern mariana islands	0	0	0	0	800	Uganda	1	1	1	0
124	Canada	1	1	1	1	364	Iran	1	1	1	1	583	Micronesia, fed. st.	0	0	0	0	804	Ukraine	1	1	1	1
132	Cape verde	1	1	1	1	368	Iraq	1	0	1	0	584	Marshall islands	0	0	0	0	807	Macedonia	1	1	1	1
136	Cayman islands	0	0	0	0	372	Ireland	1	1	1	1	585	Palau	0	0	0	0	818	Egypt	1	1	1	1
140	Central african republic	1	1	1	1	376	Israel	1	1	1	1	586	Pakistan	1	1	1	1	826	United kingdom	1	1	1	1
144	Sri lanka	1	1	1	1	381	Italy	1	1	1	1	591	Panama	1	1	1	1	834	Tanzania	1	0	1	0
148	Chad	1	1	1	1	384	Côte d'ivoire	1	1	1	1	598	Papua new guinea	1	1	1	1	842	United states of america	1	1	1	1
152	Chile	1	1	1	1	388	Jamaica	1	1	1	1	600	Paraguay	1	1	1	1	850	Virgin islands (u.s.)	0	0	0	0
156	China	1	1	1	1	392	Japan	1	1	1	1	604	Peru	1	1	1	1	854	Burkina faso	1	1	1	1
170	Colombia	1	1	1	1	398	Kazakstan	1	0	1	1	608	Philippines	1	1	1	1	858	Uruguay	1	1	1	1
174	Comoros	1	1	1	0	400	Jordan	1	0	1	1	616	Poland	1	1	1	1	860	Uzbekistan	1	0	1	0
175	Mayotte	0	0	0	0	404	Kenya	1	1	1	1	620	Portugal	1	1	1	1	862	Venezuela	1	1	1	1
178	Congo	1	0	1	1	408	Korea, dem. people's rep	0	0	0	0	624	Guinea-bissau	1	1	1	1	876	Wallis and futuna island	0	0	0	0
180	Congo (democratic rep.)	1	0	1	0	410	Korea	1	1	1	1	626	East timor	0	0	0	0	882	Samoa	1	0	1	1
184	Cook islands	0	0	0	0	414	Kuwait	1	1	1	1	630	Puerto rico	0	0	0	0	887	Yemen	1	1	1	0
188	Costa rica	1	1	1	1	417	Kyrgyzstan	1	1	1	1	634	Qatar	1	1	1	1	891	Serbia and Montenegro	1	0	1	0
191	Croatia	1	1	1	1	418	Lao people's democratic	1	1	1	1	638	Reunion	0	0	0	0	894	Zambia	1	0	1	0
192	Cuba	1	0	1	0	422	Lebanon	1	1	1	1	642	Romania	1	1	1	1	**Total**		**183**	**142**	**185**	**146**
196	Cyprus	1	1	1	1	426	Lesotho	1	1	1	1	643	Russian federation	1	1	1	1						
203	Czech republic	1	1	1	1	428	Latvia	1	1	1	1	646	Rwanda	1	1	1	1						
204	Benin	1	1	1	1	430	Liberia	1	1	1	1	654	Saint Helena	0	0	0	0						
208	Denmark	1	1	1	1	434	Libyan arab jamahiriya	1	1	1	1	659	Saint kitts and nevis	1	0	1	0						
212	Dominica	1	0	1	0	438	Liechtenstein	0	0	0	0	660	Anguilla	0	0	0	0						
214	Dominican republic	1	1	1	1	440	Lithuania	1	1	1	1	662	Saint lucia	1	0	1	0						
218	Ecuador	1	1	1	0	442	Luxembourg	1	1	1	1	666	St. pierre and miquelon	0	0	0	0						
222	El salvador	1	1	1	1	446	Macau	1	1	1	1	670	Saint vincent and the gr	1	0	1	0						
226	Equatorial guinea	1	1	1	1	450	Madagascar	1	1	1	0	674	San marino	0	0	0	0						

Note : I=Inward, O=Outward,S=Stocks, F=Flows

APPENDIX C: SENSITIVITY ANALYSIS OF ECONOMETRICS ESTIMATION

The number of zero imposes to use specific methods of estimation. Simple ordinary least squares (OLS) with log specification can only be applied on positive values, which obliges us to drop the majority of the sample and may lead to sampling-bias. To avoid it, we can include zero by adding a small amount to each FDI values. This strategy is equivalent to consider zero as small values, which is probably not appropriate. Instead, it is possible to account for the possible sampling bias by using a Heckman method.

Various specifications are tested, in complement to the Poisson Quasi-Maximum Likelihood Estimation which provides the best fit: OLS on all values by adding 1 to each value and on all positive values and 2-step Heckman estimation. All estimations are made on pooled data on the period 1994-2004.

Results for stocks are provided in. The last two columns contain estimation of PQMLE on cross-section data for years 2001 and 2004.[26]

[26] All statistical analysis and graphics were made with R (R Development Core Team, 2005).

Table 7 - FDI stocks regressions – Year and Sector Fixed Effects

Variables	PQMLE	OLS(FDI+1)	OLS(FDI>0)	Heckman	PQMLE - 2001	PQMLE - 2004
ln origin GDP	0.775a	0.464a	0.595a	1.162a	0.825a	0.789a
	(0.009)	(0.004)	(0.01)	(0.019)	(0.023)	(0.018)
ln destination GDP	0.858a	0.333a	0.548a	0.945a	0.939a	0.873a
	(0.009)	(0.004)	(0.009)	(0.015)	(0.025)	(0.019)
ln origin GDP per cap.	2.995a	0.639a	1.881a	3.629a	3.553a	2.993a
	(0.061)	(0.014)	(0.056)	(0.075)	(0.153)	(0.118)
ln destination GDP per cap.	2.417a	0.345a	0.933a	0.912a	2.705a	2.05a
	(0.061)	(0.014)	(0.037)	(0.036)	(0.176)	(0.111)
ln Distance	-0.756a	-0.409a	-0.59a	-1.17a	-0.847a	-0.896a
	(0.01)	(0.006)	(0.012)	(0.021)	(0.027)	(0.022)
Common language	0.683a	0.763a	0.574a	0.75a	0.831a	0.565a
	(0.029)	(0.027)	(0.045)	(0.044)	(0.086)	(0.058)
Colonial link	0.136a	0.42a	0.554a	0.997a	0.011	0.168a
	(0.033)	(0.029)	(0.045)	(0.046)	(0.1)	(0.068)
Developing (origin)	-0.345	0.341a	0.437a	0.995a	-0.078	-0.284
	(0.215)	(0.026)	(0.087)	(0.087)	(1.622)	(0.306)
Developing (destination)	1.558a	0.5a	0.73a	0.831a	-0.178	1.352a
	(0.086)	(0.026)	(0.061)	(0.06)	(1.838)	(0.146)
Nb. Obs.	68230	68230	20747	68230	6800	15345
R^2 in levels	0.384	0.128	0.191	0.11	0.417	0.466
R^2 in logs	0.529	0.449	0.462	0.491	0.542	0.506
RMSE in levels	1551.269	1845.023	3184.577	3338.952	1734.587	1670.104
RMSE in logs	1.928	1.619	1.721	1.673	1.814	1.8

Note: Standard errors in parentheses. a, b and c denote respectively significance at the 1%, 5% and 10% levels

OLS results are very sensitive to the inclusion of zero values, their estimations changing a lot depending on this choice. Alternatives to PQMLE lead to quite poor measures of fit (R^2 and RMSE in levels). Moreover, PQMLE seems rather robust to the period of estimation, since estimates on different sample lead to similar results.

We apply the same methods on flows and results are presented in Table 8.

However, FDI flows are much more volatile than stocks, and their short run fluctuation may not always be linked to our explanatory variables. It is, indeed, difficult to predict big instances of merger and acquisitions at this level of aggregation. It may be one explication to poorest measures of fit for flows than for stocks. Estimations on 2001 and 2004 show also important differences with results on pooled data, which points the inter-annual volatility

problem out to us. Differences exhibited by PQMLE regressions by year comfort us in pooling the data.

Table 8 - FDI flows regressions - Year and Sector Fixed Effects

Variables	PQMLE	OLS(FDI+1)	OLS(FDI>0)	Heckman	PQMLE - 2001	PQMLE - 2004
ln origin GDP	0.717a	0.262a	0.407a	1.065a	0.751a	0.687a
	(0.017)	(0.003)	(0.012)	(0.028)	(0.043)	(0.026)
ln destination GDP	0.804a	0.197a	0.393a	0.872a	0.708a	0.606a
	(0.018)	(0.003)	(0.011)	(0.021)	(0.043)	(0.026)
ln origin GDP per cap.	2.781a	0.45a	1.306a	3.535a	3.254a	1.562a
	(0.122)	(0.012)	(0.07)	(0.11)	(0.317)	(0.162)
ln destination GDP per cap.	2.688a	0.228a	0.637a	0.882a	3.748a	1.871a
	(0.115)	(0.012)	(0.042)	(0.042)	(0.348)	(0.114)
ln Distance	-0.671a	-0.26a	-0.436a	-1.139a	-0.7a	-0.848a
	(0.02)	(0.005)	(0.015)	(0.031)	(0.057)	(0.031)
Common language	0.598a	0.758a	0.309a	0.842a	-0.632b	0.656a
	(0.057)	(0.025)	(0.051)	(0.054)	(0.3)	(0.086)
Colonial link	0.057	0.059b	0.568a	0.924a	1.202a	0.76a
	(0.065)	(0.024)	(0.055)	(0.055)	(0.314)	(0.093)
Developing (origin)	-0.199	0.064a	0.218b	0.847a	-2.002	-0.427
	(0.284)	(0.02)	(0.104)	(0.104)	(5.075)	(0.297)
Developing (destination)	1.829a	0.194a	0.314a	0.72a	1.585	1.792a
	(0.133)	(0.02)	(0.068)	(0.068)	(2.104)	(0.143)
Nb. Obs.	59806	59806	12808	59806	5426	12024
R^2 in levels	0.19	0.071	0.102	0.067	0.396	0.393
R^2 in logs	0.439	0.351	0.349	0.381	0.457	0.418
RMSE in levels	397.206	425.372	893.206	910.14	446.418	196.109
RMSE in logs	1.493	1.189	1.626	1.58	1.393	1.264

Note: Standard errors in parentheses. a, b and c denote respectively significance at the 1%, 5% and 10% levels

LIST OF WORKING PAPERS RELEASED BY CEPII

An Exhaustive list is available on the website: \\www.cepii.fr.

No	*Tittle*	*Authors*
2012-07	On currency misalignments within the euro area	V. Coudert, C. Couharde & V. Mignon
2012-06	How frequently firms export? Evidence from France	G. Békés, L. Fontagné B. Muraközy & V. Vicard
2012-05	Fiscal Sustainability in the Presence of Systemic Banks: the Case of EU Countries	A. Bénassy-Quéré & G. Roussellet
2012-04	Low-Wage Countries' Competition, Reallocation across Firms and the Quality Content of Exports	J. Martin & I. Méjean
2012-03	The Great Shift: Macroeconomic Projections for the World Economy at the 2050 Horizon	J. Fouré, A. Bénassy-Quéré & L. Fontagné
2012-02	The Discriminatory E_ect of Domestic Regulations on International Services Trade: Evidence from Firm-Level Data	M. Crozet, E. Milet & D. Mirza
2012-01	Optimal food price stabilization in a small open developing country	C. Gouël & S. Jean
2011-33	Export Dynamics and Sales at Home	N. Berman, A. Berthou & J. Héricourt
2011-32	Entry on Difficult Export Markets by Chinese Domestic Firms: The Role of Foreign Export Spillovers	F. Mayneris & S. Poncet
2011-31	French Firms at the Conquest of Asian Markets: The Role of Export Spillovers	F. Mayneris & S. Poncet

No	*Tittle*	*Authors*
2011-30	Environmental Policy and Trade Performance: Evidence from China	L. Hering & S. Poncet
2011-29	Immigration, Unemployment and GDP in the Host Country: Bootstrap Panel Granger Causality Analysis on OECD Countries	E. Boubtane D. Coulibaly & C. Rault
2011-28	Index Trading and Agricultural Commodity Prices: A Panel Granger Causality Analysis	G. Capelle-Blancard & D. Coulibaly
2011-27	The Impossible Trinity Revised: An Application to China	B. Carton
2011-26	Isolating the Network Effect of Immigrants on Trade	M. Aleksynska & G. Peri
2011-25	Notes on CEPII's Distances Measures: The GeoDist Database	T. Mayer & S. Zignago
2011-24	Estimations of Tariff Equivalents for the Services Sectors	L. Fontagné, A. Guillin & C. Mitaritonna
2011-23	Economic Impact of Potential Outcome of the DDA	Y. Decreux & L. Fontagné
2011-22	More Bankers, more Growth? Evidence from OECD Countries	G. Capelle-Blancard & C. Labonne
2011-21	EMU, EU, Market Integration and Consumption Smoothing	A. Christev & J. Mélitz
2011-20	Real Time Data and Fiscal Policy Analysis	J. Cimadomo
2011-19	On the inclusion of the Chinese renminbi in the SDR basket	A. Bénassy-Quéré & D. Capelle
2011-18	Unilateral trade reform, Market Access and Foreign Competition: the Patterns of Multi-Product Exporters	M. Bas & P. Bombarda
2011-17	The " Forward Premium Puzzle" and the Sovereign Default Risk	V. Coudert & V. Mignon

No	*Tittle*	*Authors*
2011-16	Occupation-Education Mismatch of Immigrant Workers in Europe: Context and Policies	M. Aleksynska & A. Tritah
2011-15	Does Importing More Inputs Raise Exports? Firm Level Evidence from France	M. Bas & V. Strauss-Kahn
2011-14	Joint Estimates of Automatic and Discretionary Fiscal Policy: the OECD 1981-2003	J. Darby & J. Mélitz
2011-13	Immigration, vieillissement démographique et financement de la protection sociale : une évaluation par l'équilibre général calculable appliqué à la France	X. Chojnicki & L. Ragot
2011-12	The Performance of Socially Responsible Funds: Does the Screening Process Matter?	G. Capelle-Blancard & S. Monjon
2011-11	Market Size, Competition, and the Product Mix of Exporters	T. Mayer, M. Melitz & G. Ottaviano
2011-10	The Trade Unit Values Database	A. Berthou & C. Emlinger
2011-09	Carbon Price Drivers: Phase I versus Phase II Equilibrium	A. Creti, P.-A. Jouvet & V. Mignon
2011-08	Rebalancing Growth in China: An International Perspective	A. Bénassy-Quéré, B. Carton & L. Gauvin
2011-07	Economic Integration in the EuroMed: Current Status and Review of Studies	J. Jarreau
2011-06	The Decision to Import Capital Goods in India: Firms' Financial Factors Matter	A. Berthou & M. Bas
2011-05	FDI from the South: the Role of Institutional Distance and Natural Resources	M. Aleksynska & O. Havrylchyk
2011-04b	What International Monetary System for a fast-changing World Economy?	A. Bénassy-Quéré & J. Pisani-Ferry

No	*Tittle*	*Authors*
2011-04a	Quel système monétaire international pour une économie mondiale en mutation rapide ?	A. Bénassy-Quéré & J. Pisani-Ferry
2011-03	China's Foreign Trade in the Perspective of a more Balanced Economic Growth	G. Gaulier, F. Lemoine & D. Ünal
2011-02	The Interactions between the Credit Default Swap and the Bond Markets in Financial Turmoil	V. Coudert & M. Gex
2011-01	Comparative Advantage and Within-Industry Firms Performance	M. Crozet & F. Trionfetti

Organisme public d'étude et de recherche en économie internationale, le CEPII est placé auprès du Centre d'Analyse Stratégique. Son programme de travail est fixé par un conseil composé de responsables de l'administration et de personnalités issues des entreprises, des organisations syndicales et de l'Université.

Les *documents de travail* du CEPII mettent à disposition du public professionnel des travaux effectués au CEPII, dans leur phase d'élaboration et de discussion avant publication définitive. Les *documents de travail* sont publiés sous la responsabilité de la direction du CEPII et n'engagent ni le conseil du Centre, ni le Centre d'Analyse Stratégique. Les opinions qui y sont exprimées sont celles des auteurs.

Les *documents de travail* du CEPII sont disponibles sur le site : http//www.cepii.fr.