



## A FOREIGN DIRECT INVESTMENT DATABASE

### FOR GLOBAL CGE MODELS

Christophe Gouel<sup>1</sup>  
Houssein Guimbard<sup>2</sup>  
David Laborde<sup>3</sup>

#### 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.

---

<sup>1</sup> World Bank, Washington and CEPII, Paris.

<sup>2</sup> CEPII, Paris, Correspondance : houssein.guimbard (at) cepii.fr.

<sup>3</sup> 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 conducting 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 Guimbar  
David Laborde

#### RESUME NON TECHNIQUE

Plusieurs organisations internationales collectent auprès de leurs membres les statistiques d'investissement direct à l'étranger (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.