11th Biennial IFC Conference on "Post-pandemic landscape for central bank statistics"

BIS Basel, 25-26 August 2022

A probabilistic method for reconstructing the FDI network in search of ultimate hosting economies[1]

Nadia Accoto, Valerio Astuti and Costanza Catalano,
Bank of Italy

[1] This presentation was prepared for the conference. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the event.

A probabilistic method for reconstructing the FDI network in search of ultimate hosting economies

Nadia Accoto, Valerio Astuti, Costanza Catalano
Department of Economics, Statistics and Research, Bank of Italy

Abstract

Bilateral Foreign Direct Investment (FDI) statistics struggle at identifying the ultimate hosting economies of a given country, due to the non-negligible presence of conduit jurisdictions. At the same time, determining the ultimate destinations of FDI investments is crucial for understanding the real paths followed by FDI investments among increasingly interdependent economies. In this paper, starting from the Coordinated Direct Investment Survey (CDIS) data collected and published by the International Monetary Fund (IMF), first we reconstruct the global FDI network through clustering techniques. Then we provide a method for computing an (approximate) distribution of the ultimate hosting economies of a country by using a probabilistic approach on this network, based on Markov chains. In particular, we analyse the Italian case.[1]

Keywords: foreign direct investments, FDI network, clustering, ultimate host economies, absorbing Markov chains

JEL classification: C51, C60, F23, G15

[1] We thank Marta Crispino, Valerio Della Corte, Andrea Del Monaco, Silvia Fabiani, Stefano Federico, Fadi Hassan, Giacomo Oddo, Giacomo Romanini, Alfonso Rosolia, Silvia Sabatini and Simonetta Zappa for useful discussions and suggestions. The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy and/or the Eurosystem.

Contents

A probabilistic method for reconstructing the FDI network in search of ultimate hosting economies 1

1. Introduction 3

2. The outward FDI network 4

   2.1 Missing data: mirroring the inward FDI 6

   2.2 Missing data: clustering techniques 6

3. In search of ultimate host economies 8

   3.1 The Markov chain model 9

   3.2 Estimation of the conduit parameters of the model 11

4. The Italian case: results 12

5. Conclusions and future work 15

   References 16

1. Introduction

Foreign Direct Investment (FDI) is a category of financial cross-border investment in which an investor (direct investor) of one economy makes an investment in an enterprise (direct investment enterprise) of another economy that allows having control or a significant degree of influence on the management of that enterprise.[2] FDI statistics, which provide information on investments by immediate counterparts between two different countries, are key indicators of countries' participation in the global economy. However, in a world that is more and more interconnected, such statistics are not sufficient to reconstruct the investment chains, due to the increasing presence of multinational enterprises and countries that act as tax havens[3] or investment hubs[4]. Indeed, investments can pass through the so-called Special Purpose Entities (SPEs), enterprises commonly created and registered in tax havens and investments hubs that allow tax optimization by channelling investments through economies, before arriving to the final investment recipient country. Therefore, a large portion of FDI transits in and out of some countries before reaching their final destination, producing no real economic value in the crossing country. This led international organizations and national compilers to consider the development of *extended* experimental statistics, such as inward FDI by Ultimate Investing Economies (UIE)[5] and outward FDI by Ultimate Host Economies (UHE)[6]. Compiling FDI statistics by UIE and UHE, other than by immediate counterpart country, in fact, would make the FDI statistics more complete and useful from a macroeconomic point of view, highlighting who ultimately controls the investments, the ultimate destination, and the financial connections between economies. The fourth edition of the OECD's Benchmark Definition of Foreign Direct Investment [1] recommends to compile inward investment positions according to the UIE; as of today, only few countries (Italy included[7]) provide such additional information as experimental statistics. FDI statistics by UHE belong to the research agenda of international organizations such as the IMF ([3], §1.43) and the OECD ([1], §672). A specific guidance and methodology concerning the UHE concept is being developed in the scope of the international statistical manuals' revision process ([4]), which is still ongoing. Recently, experimental methods have been developed by the FDI statistical community [5], [6], although no countries have yet published comprehensive statistics on that.

In this framework, we propose a model to estimate, for each country, the distribution of its outward FDI by UHE. The mathematical model is based on absorbing Markov chain, and has as input the reconstructed global outward FDI network.

[2] By definition, the direct investor owns 10 percent or more of the voting power in the direct investment enterprise [1].

[3] Countries whose economies are (almost) entirely dedicated to the provision of offshore services. In this paper we consider as tax havens the countries listed in the Balance of Payments vademecum [2].

[4] Jurisdictions that facilitate transit of investments due to favourable tax and investment conditions.

[5] Economies where the first investments are originated.

[6] Final recipient economies of the investment chains.

[7] Information on the residency of ultimate investor is collected directly from reporting firms in the annual FDI survey and then cross-validated with commercial databases (Orbis).

The data on outward FDI are taken from the Coordinated Direct Investment Survey (CDIS) database, where the IMF collects yearly, on a voluntary basis, information about FDI stocks from all the world's countries. Even though the process of compiling the survey can have its own characteristics in each respondent country, our model should not be heavily affected by these differences, given that we exploit only proportions of outward investments from each country and not the absolute declared values. Since the reporting countries in the CDIS database are only 89 over 246 world's countries, and some of the disseminated data are kept confidential by the reporting country itself, some imputation methods are necessary to reconstruct the full network. In particular, we make use of mirroring and clustering techniques. The Markov chain model (inspired by [7]) is then set up on this reconstructed network, providing, for each country, an estimate of the percentage of its outward FDI towards each other country as final recipient.

The paper is structured as follows: Section 2 contains the description of how the outward FDI network is fully reconstructed, addressing the problem of missing data, while Section 3 describes the mathematical model used to estimate the FDI distribution by UHE for every country. Finally, Section 4 reports the results of such model in the Italian case and Section 5 presents some conclusions and future work.

2. The outward FDI network

We consider the bilateral data on outward stocks of FDI by counterpart country on year 2019, taken from the CDIS database.

The data in the CDIS database are presented according to the Extended Directional Principle ([3], §6.42); a direct investment is shown as either an investment abroad (outward investment) or an investment in the reporting economy (inward investment). Direct investments abroad cover assets and liabilities between resident direct investors and their non-resident direct investment enterprises. Direct investment in the reporting economy includes all liabilities and assets between resident direct investment enterprises and their non-resident direct investors. Assets and liabilities between resident and non-resident fellow enterprises are classified as outward (inward) investment if the ultimate controlling parent is resident (non-resident). The data are broken down by financial instruments (equity and debt instruments).[8] The CDIS reporting economies are primarily requested to provide data on inward FDI, but they can complement them also providing data on outward FDI. Our aim is to reconstruct the (weighted, directed) network of bilateral outward FDI between the world's countries and to use it to gain information about the ultimate hosting economies of each country (in particular, of Italy). More specifically, we want to reconstruct the network having as nodes all the CDIS countries, and as (directed) links of weight w from country i to country j the total FDI stocks of w dollars from country i towards country j. This network will be the starting point of the subsequent Markov chain model that will provide approximated FDI statistics by UHE. The main difficulties in reconstructing the network using CDIS data can be summarized as follows:

1. Not all countries provide data on their outward FDI;

[8] Debt instruments refer to debt positions between affiliated enterprises, see also §6.26 [3].

2. due to the extended directional principle presentation, some reported imports are negative,[9] while we need positive weights for the Markov chain model;

3. reporting countries can decide to flag some of their data as *confidential*, thus masking the value of the investment.[10]

In particular, the reporting countries are only 89 over the 246 possible ones and the confidential data are 24.7% of the available data, spread over 37 reporting countries. The percentage of confidential data varies between countries: it goes from a maximum of 91.8% of confidential data for Hong Kong to a non-null minimum of 0.8% of confidential data for Guatemala, and no confidential data for 52 countries. See also Figure 1. We handled item 2. by considering the absolute value of such negative amounts.[11] In the next sections, we describe the solutions adopted in order to handle items 1. and 3.

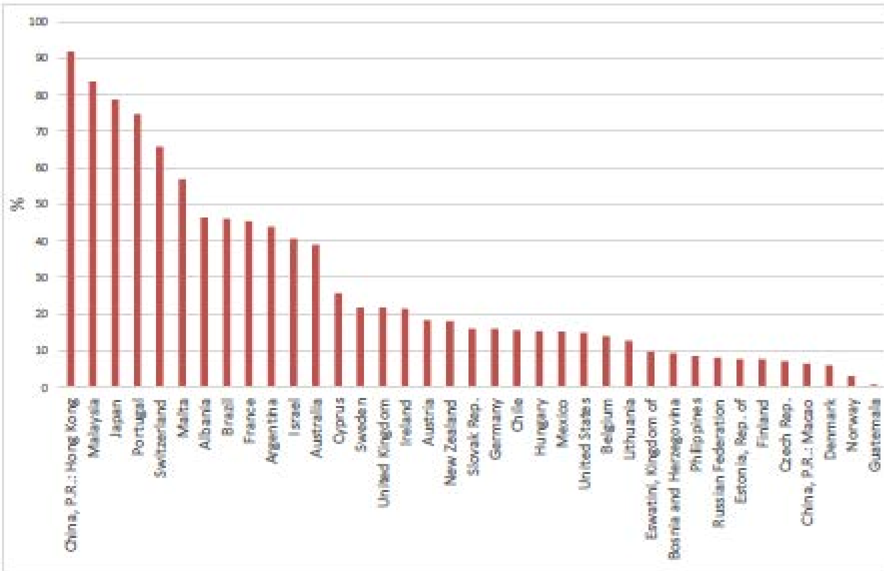


Figure 1: Percentage of confidential data in the outward FDI database by country.

9 FDI data presented according to the extended directional principle can be negative; this happens when the debt positions from the subsidiary to the parent company exceed the value of the investment of the parent in the subsidiary.

10 This expedient is often used to avoid that individual information are indirectly disclosed by deduction from reported data. In this case, we know that a link between two countries exists but we do not know its weight.

11 Setting these negative values to zero would have been in contrast with the information that such link exists.

2.1 Missing data: mirroring the inward FDI

The CDIS database provides both the bilateral inward and outward FDI statistics supplied by the reporting countries.[12] The inward and outward database, in theory, should be symmetric, as for every countries A and B it should hold:

Outward FDI from A to B = Inward FDI of B from A (1)

In real data, these two values might not coincide as the one on the left hand side of (1) is reported by country A, while the one on the right hand side is reported by country B. Property (1) can nonetheless be of help in (partially) solving items 1. and 3. listed above. Indeed, in the inward database the reporting countries result to be 119, thus providing many additional data with which we can complement the network. We complement the outward database with data taken form the inward database and not vice versa, despite the latter having more reporting countries than the former, because that is the direction of the investment we want to exploit in the model when looking for the ultimate recipient countries. In particular, we use the inward database to:

- impute the confidential data. If the outward FDI of country A to country B is confidential, we check the inward FDI reported by country B from A: if it exists and it is different from zero, we substitute the confidential data with this amount;
- add missing links. If the outward FDI of country A to country B is equal to zero or it is not reported, we check the inward FDI reported by country B from A: if it exists and it is different from zero, we add a link from A to B with this amount as weight.[13]

After all these operations, we again consider the absolute values of the negative amounts that might have been substituted.

This procedure let us impute 667 values over the 1918 starting confidential cells, and it let us add 7771 links (921 links that were reported as zeros in the outward FDI database plus 5915 extra links referring to non-reporting countries in the outward FDI database). The final network has 246 nodes (the total number of CDIS countries) and 14607 links, 3248 of which are still confidential (22.2% of the total data different from zero).[14] The network results to be strongly connected, i.e. there is a path of outward FDI from any country to any other country, and quite sparse (sparsity score=0.759, i.e. only 24.1% of links over all the possible ones are present). The next section describes how we impute the values of the remaining confidential data.

2.2 Missing data: clustering techniques

After the initial imputation described in the previous section, we still have a high number of confidential data, which could potentially distort the results of the analysis.

12 All the reporting countries of the outward database are also reporting countries in the inward database but not vice versa.

13 Note: in this case, we also keep the inward FDI data that are confidential, as they signal the presence of a link that was not reported in the outward FDI database.

14 The percentage of confidential data is increased because we have complemented the outward FDI database also with the confidential data from the inward FDI database.

We propose in this section a second step of imputation, based on the proximity of countries in the "*outward investment space*". More specifically, each country i defines a real-valued vector $\vec{v}_{\mathrm{i}}$ in a 246-dimensional space, where each j-th component $v_{\mathrm{i}}^{\mathrm{j}}$ of the vector denotes the amount invested by country i in country j. All of these components have non-negative values, such that the vectors span only a very small wedge of the embedding space. In addition, countries are obviously very different in terms of the total amount invested - which is linked to the magnitude of these vectors - but useful information on the investment pattern is contained in the *ratios* between the components. For this reason we define two different algorithms to study the similarity of the investment vectors: first, we consider the scalar product of the normalized investment vectors for two given countries i and j:

$$\sigma_{ij} = \hat{\vec{v}}_i \cdot \hat{\vec{v}}_j = \frac{\sum_{k=1}^{246} v_i^k \, v_j^k}{\|\vec{v_i}\| \, \|\vec{v_j}\|} \tag{2}$$

where for any vector $\vec{v}$, we define its norm $\|\vec{v}\| = [\sum_k v^k \cdot v^k]^{1/2}$. As an alternative, we consider the scalar product between the normalized vectors after having subtracted their means:

$$\rho_{ij} = \tilde{\vec{v}}_i \cdot \tilde{\vec{v}}_j = \sum_{k=1}^{246} \tilde{v}_i^k \, \tilde{v}_j^k \tag{3}$$

where $\tilde{v}_i^k = \frac{(\vec{v_i} - \bar{v}_i)}{\|\vec{v_i} - \bar{v_i}\|}$ and $\tilde{v}_j^k = \frac{(\vec{v}_j - \bar{v}_j)}{\|\vec{v}_j - \bar{v}_j\|}$ are the standardized vectors and $\bar{v}_i$ and $\bar{v}_j$ are the mean vectors.

While in principle leading to different results, the two proximity scores are in practice very similar for most of the countries; the reason is that many investment vectors have a large number of zero components, and few very large ones. This implies that subtracting the mean changes very little in terms of the components that most heavily influence the proximity scores. The underlying assumption of our imputation strategy is that countries that are close in this space have a similar investment pattern, such that, if country i has a confidential value for investments towards a given country k, the average proportion of investment towards country k from a set of countries which are close to i will function as a good estimator for the missing value.

Clearly, the existence of missing values will have an influence also on the evaluation of the proximity score, in that we ignore the portion of the investment pattern which is hidden in confidential data. If for a given country the number of confidential entries is too large, we will probably have a distorted representation of the behaviour of the country. In the evaluation of the proximity score, neglecting the confidential values altogether would be equivalent to assuming that no investment is present in these cases. This assumption seems however unrealistic, in that one would think there are more reasons to have confidentiality on something existing, than on an absent investment. For this reason, to represent confidential data we choose a small, positive number - for definiteness fixed at 1.[15] Whenever we encounter a confidential value for the investment from country i to country j, we select the $\mathrm{N}_{\mathrm{close}}$

[15] The exact value of this placeholder does not play any role in the clustering, and is exploited only for bookkeeping reasons.

countries with the highest values of the proximity scores with i having a non-confidential value for the investments towards j (such that the list of N_{close} countries can be different for any imputed confidential value). In addition we put a minimum threshold on the proximity score for the country to be considered as "close" to the one with the confidential value.[16] From this sample of neighbour countries, we compute the average of the proportion of the investments towards country j (i.e. the investment towards country j divided by the total outward investments). If the sample of countries selected is empty (for example, if every country with the proximity score above the chosen threshold has also a missing value for country j), we impute the confidential value with a null value. For each country, we know the total outward FDI stocks toward the rest of the world. From this and the average proportions, it is easy to derive the absolute value of the imputed outward links. In particular, we impose for the sum of the outward FDI (both previously known and imputed) to be equal to the known total outward FDI of each country.

To test the performance of the algorithm, we consider two out-of-sample tests: first, we select a pair of countries and hide the value of the investment from the first to the second. We impute the hidden values, and evaluate the difference between the imputed value and the real ones. As a second test, we hide 10 outward investments of a single country and impute them. In this way, we can evaluate the correlation between the series of true values and the series of imputed ones. With $N_{close} = 3$, the percentage of zero values after the imputation procedure is close to 5%. For comparison, 75% of non-confidential values in the data-set are null. We report some summary statistics of the error distribution of the imputation process in Table 1, excluding the cases in which the imputed value is equal to zero. The absolute error is simply the difference between the imputed proportion of investment toward a given country, and the true value. To have a comparative measure of these values, the average proportion, corresponding to the inverse of the number of countries in the network, is $246^{-1} \approx 0.0041$. For the correlation between the series of imputed and true values, we exclude the cases in which the series of true values is composed of null values only.

	Mean	25-th percentile	Median	75-th percentile
Proximity	0.50	0.28	0.53	0.70
Absolute error (10^{-3})	-0.39	-0.49	0.00	1.20
Correlation	0.53	0	0.69	0.99

Table 1: Summary statistics of the imputation process results

3. In search of ultimate host economies

Once we have reconstructed the network, we can proceed to build the model to estimate the distribution of FDI by ultimate host economies. We here provide a description on how such model works.

16 The proximity score ρ_{ij} can assume negative values, so when using it to establish the proximity between countries we impose the constraint $\rho_{ij} > 0$. In Table 1 we report summary statistics of the proximity score between each country and its selected neighbours.

Conduit FDI arises when a multinational enterprise investing from home country A in host country B establishes an intermediate step through a third country C. This intermediate step is merely financial, as in country C no real "productive" investment takes place, and it is generally qualified as conduit FDI. In the model we allow for a conduit component in each country, representing the percentage of FDI received passing through the country. We now want to simulate the investment process on the FDI network following the investment from the investor to the final recipient. Starting from country A (the origin of the investment), we know from bilateral data the distribution of the investments of A towards its recipient countries (the out-neighbours of A in the network). If $w_{A,B}$ represents the magnitude of the outward FDI of country A towards B, and $s_A = \sum_B w_{A,B}$ is the total outward FDI coming from country A, we can say that an investment from A has a probability of $\pi_{A,B} = w_{A,B}/s_A$ to be invested in country B. Suppose now that the investment has been indeed made towards country B: it either stays there (country B behaves as a non-conduit jurisdiction) and the investments chain stops making B the UHE of A, or country B behaves as a conduit and makes the investment pass through towards another country. The investment will pass from country B to a country C in the network according the probability $\pi_{B,C}$ computed from country B's FDI bilateral data.[17] Again, either the investments stops there, making C the UHE, or country C behaves as a conduit and let the investment pass through towards another country. Eventually, the investment chain will stop in the final recipient economy, that is the UHE.

What we have just described is the behaviour of a random walk on (a modified version of) the FDI network. In the next sub-section we provide the rigorous mathematical framework of this model.

3.1 The Markov chain model

Consider the FDI network and suppose for each country/node A to be split into its conduit component A_c and its non-conduit component A_{nc}, thus producing a total of 246 · 2=492 nodes. To each country B we associate a probability $q_c(B)$ to act as a conduit: how we estimate this parameter for every country will be addressed in the next section. Here we underline that we are making the strong assumption that the percentage of passing-through investments of a country B *does not depend* on the country from which it receives the investment: this assumption comes from the scarcity of the available data, as will be explained in the next section.[18]

We now proceed to define a (discrete-time) Markov chain process[19] on such augmented network by specifying the transition probabilities between any pair of nodes. The nodes of the network, according to standard notation, will be also referred to as the *states* of the Markov chain. We remind that $\pi_{A,B}$ represents the percentage of outward FDI from country A to country B. For each pair of nodes (A_i, B_j), where

[17] It is not guaranteed that conduit investments will follow the same distribution $\pi_{B,C}$ as non-conduit ones; being unable to estimate the particular distribution of conduit investments, we make an uninformative assumption and conflate the two (in principle different) distributions.

[18] A generalized model where such probabilities depend also on the investing country will be object of future work.

[19] For formal definitions see [8].

$i, j \in \{c, nc\}$, we define the probability p_{A_i,B_j} that an investment passes from node A_i to node B_j in the following way:

- If $i = nc$, then $p_{A_{nc}B_j}$ = 1 if $B_j = A_{nc}$, otherwise $p_{A_{nc}A_{nc}} = 0$.

  This represents the fact that country A act as a non-conduit, so the investment remains in such country and does not pass through.

- If $i = c = j$, then $p_{A_cB_c} = \pi_{A,B} \cdot q_c(B)$.

  This represents the fact that A is investing on B and B will let the investment pass through (acts as a conduit).

- If $i = c$ and $j = nc$, then $p_{A_cB_{nc}} = \pi_{A,B} \cdot (1 - q_c(B))$.

  This represents the fact that A is investing on B and B does not let the investment pass through (acts as a non-conduit), i.e. it is the final recipient.

We call the non-conduit states A_{nc} an *absorbing* states, because once we enter that state we cannot leave. All the other states A_c are called *transient*. We can then arrange all the above probabilities in a matrix $P = [p_{A_i,B_j}]$ of dimension 492×492 ordering first by the absorbing states and then by the transient ones. The matrix P will then have the following structure:

$$P = \begin{array}{c c} & \begin{array}{cc} \text{abs} & \text{trans} \end{array} \\ \begin{array}{c} \text{abs} \\ \text{trans} \end{array} & \left[\begin{array}{cc} I & 0 \\ R & Q \end{array} \right] \end{array} \qquad (4)$$

Where I is the 246× 246 identity matrix, R represents the transition probabilities $p_{A_{c},B_{nc}}$ from transient states to absorbing states, and Q represents the transition probabilities p_{A_{c},B_c} between transient states. This defines a so-called *absorbing* Markov chain.[20] The initial state of the Markov chain is any stochastic vector[21] η of length 492. In particular, if we want the model to start from a given country A, the vector η will have all zero entries but $\eta_{A_c} = 1$. This initial condition on transient states is a mere modeling expedient to ensure that the investment moves towards another country in the first step (as otherwise it would not be a *foreign* direct investment).

In an absorbing Markov chain, a random walk starting from any of the states will eventually end up in one of the absorbing states (and remain there forever). The following theorem provides the long-run distribution of transition probabilities on an absorbing Markov chain (see also [9]).

Theorem.

The limiting distribution of an absorbing Markov Chain with transition matrix as in (4) is given by:

$$P^* = \lim_{n\to\infty} P^n = \begin{array}{c c} & \begin{array}{cc} abs & trans \end{array} \\ \begin{array}{c} abs \\ trans \end{array} & \left[\begin{array}{cc} I & 0 \\ (I-Q)^{-1}R & 0 \end{array} \right] \end{array} \qquad (5)$$

20 An absorbing state is a state that is impossible to leave (no outgoing links). An absorbing Markov chain is a Markov chain with at least one absorbing state that is reachable from every state.

21 Entry-wise nonnegative vector whose entries sum up to 1.

The interpretation of the limiting matrix P^* is the following: for any country that acts as source of the investment (the rows of the matrix), the process will end up after a sufficiently large number of investment steps, with probability 1, in one of the UHE, modeled as absorbing states (the columns of the matrix). In particular, if we want to retrieve the UHE distribution of a country A, we just need to take the row indexed by A_c of the fundamental matrix $(I - Q)^{-1}R$: this is a distribution by UHE over the (non-conduit version of the) world's countries. In other words, the share of FDI from country A that ends up in country B as final recipient is $P^*_{A_c,B_{nc}} = ((I - Q)^{-1}R)_{A_c,B_{nc}}$, i.e. the entry of P^* on row A_c and column B_{nc}; this is equivalent to perform the vector-matrix multiplication $\eta^\top P^*$, with $\eta_{A_c} = 1$ and all its other entries equal to zero.

3.2 Estimation of the conduit parameters of the model

We here describe, for each country A, how we estimate the parameter $q_c(A)$ i.e. the probability that an investment arriving in A would pass through it heading to another country.

As we have already mentioned in the introduction, most conduit FDI in the world take place through a limited set of jurisdictions that act as global FDI hubs. Such hubs can be divided into two groups:

a) the *tax havens*, that are small jurisdictions whose economy is entirely, or almost entirely, dedicated to the provision of offshore financial services;

b) other *investment hubs*, that are countries that have a substantial real economic activity but also act as conduit jurisdictions due to their favourable tax and investment regime.

We consider as tax havens 38 countries, [22] listed by the European Commission in the BoP Vademecum [2], plus 4 other countries[23] identified by Casella in [7]. Note that we consider Hong Kong and Singapore belonging to group (b) instead of group (a), despite appearing in the BoP Vademecum, due to their relevant size in term of population, comparable to other investments hubs such as Luxembourg. Since tax havens act fully as conduit jurisdictions, we associate to them a probability $q_c = 1$.

Regarding the investments hubs belonging to group (b), there are some countries that report[24] their yearly outward FDI investments made through Special Purpose Entities (SPEs). These data can then be used to compute their conduit probability as the ratio between the inward FDI through SPEs and total inward investment:

$$q_c(A) = \frac{\text{Inward FDI of } A \text{ in resident SPEs}}{\text{Total Inward FDI of } A}. \tag{6}$$

[22] Andorra, Anguilla, Antigua and Barbuda, Aruba, Bahamas, Bahrain, Barbados, Belize, Bermuda, Cayman Islands, Cook Islands, Curaçao, Dominica, Gibraltar, Grenada, Guernsey, Isle of Man, Jersey, Lebanon, Liberia, Liechtenstein, Marshall Islands, Mauritius, Montserrat, Nauru, Niue, Panama, Philippines, Saint Lucia, Saint Vincent and the Grenadines, Samoa, Seychelles, Sint Maarten, St Kitts and Nevis, Turks and Caicos Islands, Vanuatu, British Virgin Islands, U.S. Virgin Islands.

[23] Malta, Monaco, Netherlands Antilles, San Marino.

[24] Dataset: FDI positions - Main aggregates BMD4, outward directional principle, OECD statistics, https://stats.oecd.org/.

In particular, as conduit probability we take the average of the above ratio on the years 2017-2019.[25] The reason why we use in the computation of the conduit probability (6) for country A its inward FDI is that we are trying to estimate the percentage of investments made to country A that will pass through it towards other countries, relative to the total investments received. The computed conduit probabilities of such self-reporting countries are reported in Table 2. To them we added four countries, namely United Kingdom, Ireland, Hong Kong and Singapore, where we consider as conduit probability the values estimated by Casella [7], through a regression method based on GDP[26] (end of Table 2). The countries listed in Table 2 are the countries that we consider belonging to group (b).

To all the countries that belong neither to group (a) nor to group (b) we assign a conduit probability of 0, thus implying that they always act as final recipient.

Country	q_c (%)
Luxembourg	94.94
Netherlands	67.16
Hungary	57.28
Switzerland	20.39
Denmark	18.90
Austria	18.81
Sweden	7.30
Spain	6.41
Belgium	6.10
Portugal	5.83
Norway	4.95
Iceland	4.17
Lithuania	3.78
Finland	3.19
Estonia	2.77
Chile	0.91
South Korea	0.41
Poland	0.31
Hong Kong	78.94
Ireland	60.31
Singapore	25.08
United Kingdom	20.16

Table 2: Conduit probability of self-reporting SPEs and estimated hubs.

4. The Italian case: results

The weighted FDI network reconstructed in Section 2 and the conduit probabilities computed in Section 3.2 are all the ingredients we need to run our Markov chain

25 If some annual data are missing, we consider the average on the years with available data.

26 Since Casella in [7] uses the outward FDI to compute such probabilities, for each country A we corrected the values by the coefficient (Total Outward FDI of A/Total Inward FDI of A). Singapore does not report its outward FDIs, so we used the sum of the imputed values.

model (Section 3.1). In particular, we focus on the Italian case; Figure 2 and 3 compare the Italian bilateral data on outward FDI from the CDIS database (presented as percentage over the total outward FDI) with the results we obtain from the model in terms of UHE (percentage of outward FDI towards final recipient countries).

We can notice that countries such as The Netherlands and Luxembourg, which are ranking respectively first and fifth according to Italian bilateral data, often let these investments pass through to other destinations: indeed, in the Italian FDI distribution by UHE, their ranks drop respectively to the sixth and 22th position. This result was somehow expected, as both countries are characterized by favorable tax regimes and by the presence of a large number of SPEs in their territory. These countries are also the ones that show the biggest (absolute) percentage difference between bilateral data and the model output by UHE (see Figure 4).

Contrarily, countries such as the United States, Germany and France show the opposite behaviour: the volume of Italian investments that they receive as ultimate recipient is larger than what is reported in bilateral data. This means that some investments originated in Italy have been channeled through investments hubs and/or tax havens before ending up to such countries. Figure 3 shows that the main Italian partners in terms of final recipient of investment chains are the United States, Germany, Spain and France (first four positions), while ranking respectively 3rd, 4th, 2nd and 6th when considering bilateral data.

It is also interesting to observe that Italy itself appears as an Italian UHE: the results indeed show that a small percentage of Italian FDI (around 0.5%) returns to Italy, highlighting the presence of round-tripping phenomena.[27] Finally, the results obtained at the country level, in particular for Luxembourg, The Netherlands and USA, are reflected in the aggregation across main areas (see Figure 5).

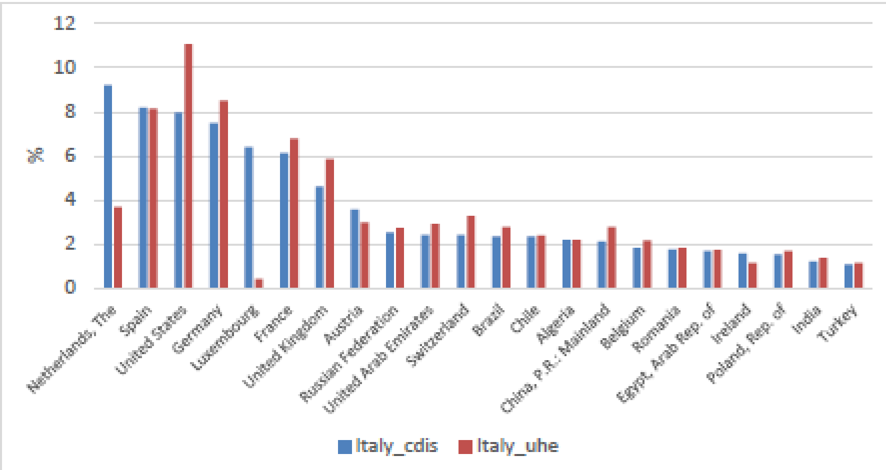


Figure 2: Comparison between the Italian bilateral FDI from CDIS and the estimated distribution of Italian FDI by ultimate host economies, sorted in decreasing order by bilateral data. Only values over 1% are displayed.

[27] Round-tripping refers to capital that leaves the economy and then goes back to it, see also §6.46 [3]

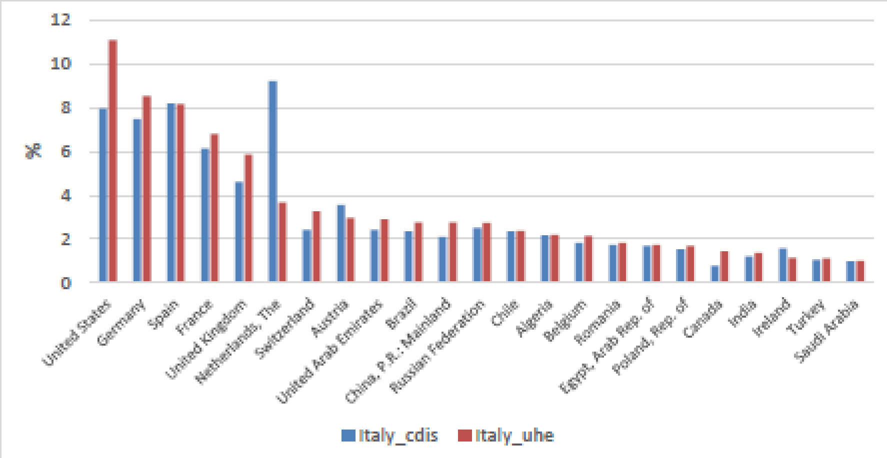


Figure 3: Comparison between the Italian bilateral FDI from CDIS and the estimated distribution of Italian FDI by ultimate host economies, sorted in decreasing order by ultimate host economies. Only values over 1% are displayed.

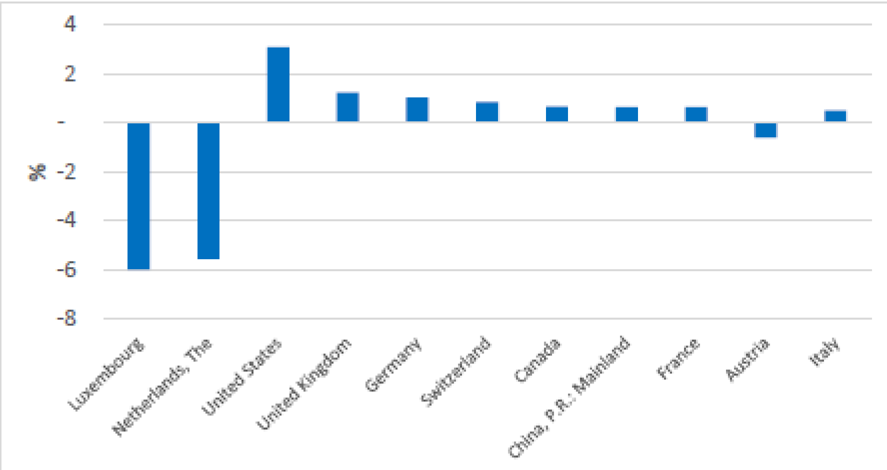


Figure 4: Difference between the estimated distribution by ultimate host economies and the Italian bilateral FDI. Only differences over 0.5 in absolute value are displayed.

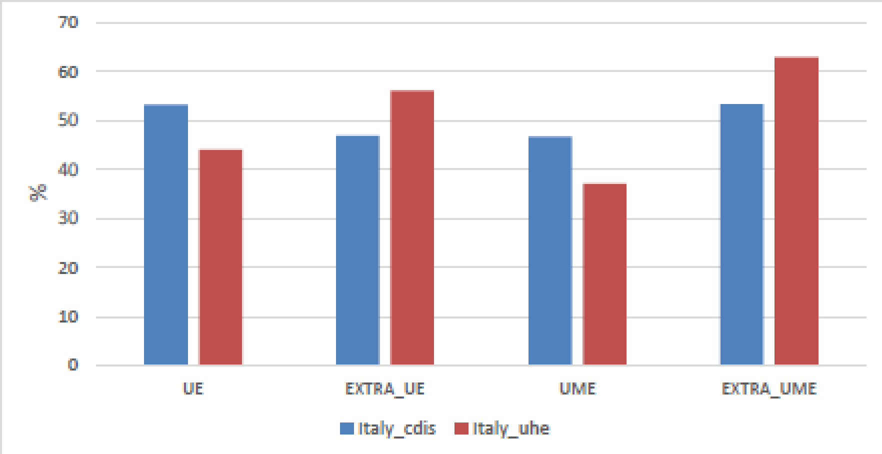


Figure 5: Comparison between the Italian bilateral FDI from CDIS and the estimated distribution of Italian FDI by ultimate host economies for main areas.

5. Conclusions and future work

In this paper we first presented a methodology on how to reconstruct the full outward FDI network starting from the (incomplete and with confidentiality issues) CDIS database. Second, we proposed a model to estimate the FDI distribution by ultimate host economies for any given country. The results of the model in the Italian case show that some of the main Italian partners in terms of bilateral FDI receive much smaller volumes of investments as final recipients.

Future work would involve testing the robustness of the model for the different assumptions that were made, as well as possibly validating it by comparing the results with other experimental FDI statistics by UHE. Moreover, we plan to consider only the equity part of the FDI data to build (and reconstruct) the network, which should let us avoid negative weights and then maybe obtain a more realistic network. It would be of interest also to refine the model by considering conduit probabilities that depend also on the country making the investment, and not only on the receiving country. Also the search of newer techniques to compute the conduit probabilities would make it possible to extend the list of investments hubs and provide more reliable estimates. Finally, we plan to perform a deeper analysis on the FDI network in terms of connectivity, resistance to shock propagation and centrality measures [10].

References

[1] OECD (2008), Benchmark Definition of Foreign Direct Investment, *fourth edition*.

[2] EUROSTAT. Balance of Payment vademecum. European Commission, Update December 2021 – February 2022.

[3] IMF. Balance of Payments and International Investment Position Manual. Version Sixth Edition (BPM6). Washington D.C., 2009. https: //www.imf.org/external/pubs/ft/bop/2007/pdf/bpm6.pdf.

[4] IMF - Direct Investment Task Team (DITT). Draft guidance note D6 -Ultimate Investing Economy (UIE)/Ultimate Host Economy (UHE) and Pass-through Funds. 2022. https://www.imf.org/en/Data/Statistics/BPM/DITT.

[5] Deutsche Bundesbank. 2022. www.bundesbank.de/resource/blob/811622/48c0cfe22fb4050ca950b0650ac1c754/mL/erlaeuterungen-data.pdf.

[6] Deutsche Bundesbank. 2022. www.bundesbank.de/en/press/press-releases/foreign-direct-investment-stocks-at-the-end-of-2020-890218.

[7] B. Casella. "Looking through conduit FDI in search of ultimate investors–a probabilistic approach". In: Transnational Corporations 26.1 (2019), pp. 109–146.

[8] J. R. Norris. Markov Chains. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[9] C. M. Grinstead and J. L. Snell. Introduction to probability. American Mathematical Society, 1997.

[10] F. Lima et al. "Foreign direct investment – using network analysis to understand the position of Portugal in a global FDI network". In: IFC Bulletins chapters 52 (2020).

A probabilistic method for reconstructing the FDI network in search of ultimate host economies

Nadia Accoto, Valerio Astuti, Costanza Catalano

26/08/2022 11th IFC Biennial Conference, Basel

Department of Economics, Statistics and Research - Bank of Italy

Foreign Direct Investments (FDI) and CDIS database

FDI: When an investor of one economy makes an investment in an enterprise of another economy, that allows to have control or a significant degree of influence. *[Source: IMF]*

The Coordinated Direct Investment Survey (CDIS)

- Promoted by IMF, annual and voluntary
- Inward FDI (investments received) and Outward FDI (investments made) by counterpart economy

Investments hubs and tax havens: facilitate transit of investments due to favourable tax regimes and off-shore services.

economies ⇒ **Goal of MultiNational Enterprises:** Tax optimization by channeling investments through several

Ultimate investing economy and Ultimate host economy

Difficult to interpret FDI statistics by immediate partner economy as it does not show the ultimate sources and destinations of FDI.

Ultimate investing economy (UIE): where the investment originated

Ultimate host economy (UHE): final recipient of the investment

- who ultimately controls the investments/ultimate destination
- reveals the financial connections between economies
- info on businesses using offshore centers

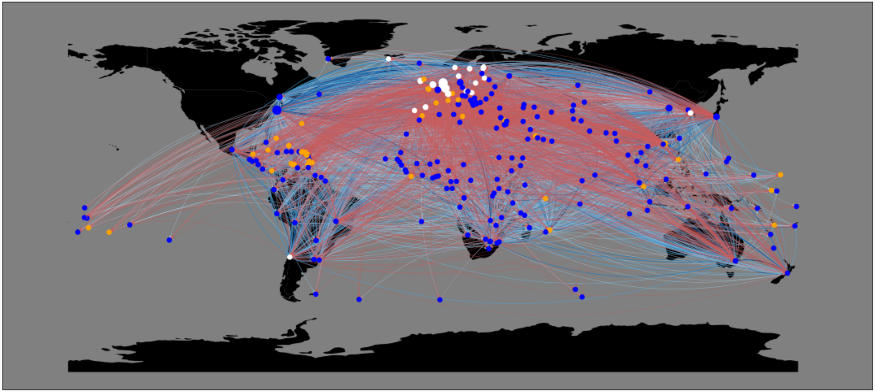
Difficult to collect/provide data on UHE (no data available).

The Outward FDI Network

246 countries represented by their capitals. Reporting countries: 89

Orange: Tax havens, **White**: investment hubs, **Blue**: all the others

Red links: Confidential data. The import of the investment is censored ($\sim$ 24.7% of the reported data)



Reconstructing the network

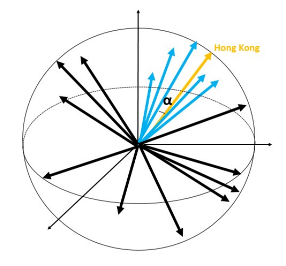
Imputation of confidentials and adding missing links:

- **Mirroring the Inward FDI database**:
  Outward of A by counterpart B = Inward of B by counterpart A

Imputation of remaining confidentials:

- **Clustering techniques**: "countries that are *similar* have the same outward ratios"

1. We cluster the countries according to how similar their outward FDI ratio are;
2. If the FDI of country A towards B is confidential, we impute it by averaging the FDI towards B of the first N_{close} countries of A that report such data.



In search of UHE: the Markov chain model

We simulate the investment process on the FDI network following the investment from the investor to the final recipient.[1]

We allow for a **conduit component** $q_c(j)$ in each country j = % of FDI received passing through it:

$$q_c(j) = \frac{SPE_{out}(j)}{FDI_{out}(j)}$$

[*OECD statistics, average over 3 years.*]

Tax haven: $q_c(j) = 1$.
All the investments pass through, no real investment is made in such countries.

Otherwise $q_c(j) = 0$: no investment passes through.

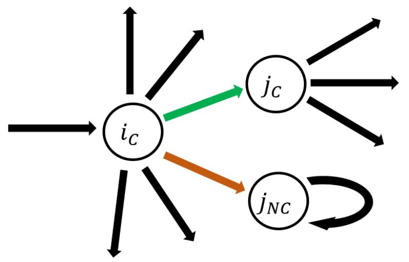
Country	$\mathbf{q_c}$ **(perc)**
Luxembourg	94.42%
Hungary	80.92%
Netherlands	59.83%
Austria	17.33%
Lithuania	12.21%
Denmark	11.66%
Switzerland	10.68%
Portugal	9.08%
Belgium	7.72%
Spain	7.48%
Estonia	7.33%
Iceland	6.73%
Norway	3.45%
Sweden	3.37%
Poland	2.42%
Chile	2.11%
Finland	0.14%
South Korea	0.01%
Hong Kong	78.00%
Ireland	65.00%
Singapore	57.00%
United Kingdom	21.00%

[1] see also B. Casella, *Looking through conduit FDI in search of ultimate investors*, UNCTAD

In search of UHE: the Markov chain model

We simulate the investment process on the FDI network following the investment from the investor to the final recipient.

1. Starts from country i: the investment has a probability of (%FDI from i to j) to be invested in country j;
2. it either stays in j (j is the UHE of i) with prob $(1 - q_c(j))$ or it passes through to another country with prob $q_c(j)$ (j behave as a conduit);
3. Reiterate steps 1. and 2. from j.



(%FDI of country i towards country j)*(prob country j acts as a conduit)

(%FDI of country i towards country j)*(prob country j does not act as a conduit)

In search of UHE: the Markov chain model

Eventually the investment chain (random walk) will end up in one non-conduit node, which will be the final recipient economy (UHE).

$\Rightarrow$ Absorbing Markov chain.
We can retrieve the final FDI distribution by UHE for each country.

Theorem

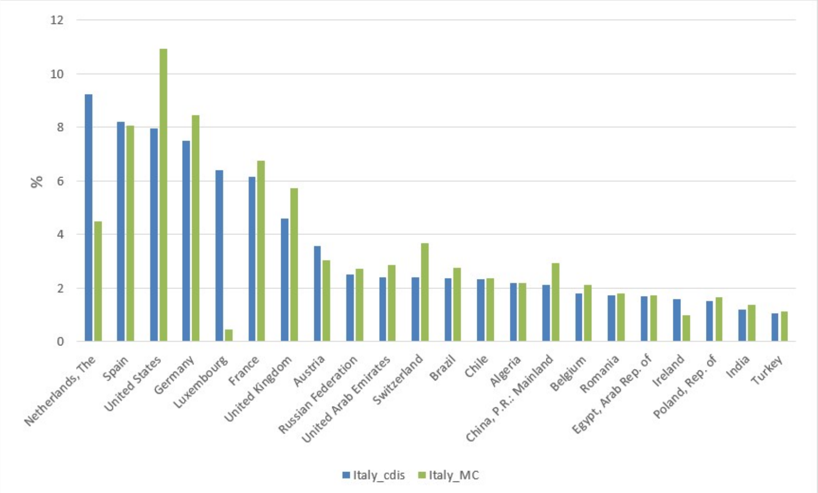
Starting country

$$P = \begin{array}{cc} & \begin{array}{cc} \text{abs} & \text{trans} \end{array} \\ \begin{array}{c} \text{abs} \\ \text{trans} \end{array} & \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix} \end{array} \xrightarrow[\text{as } n \to \infty]{} P^* = \begin{array}{cc} & \begin{array}{cc} \text{abs} & \text{trans} \end{array} \\ \begin{array}{c} \text{abs} \\ \text{trans} \end{array} & \begin{bmatrix} I & 0 \\ \boxed{(I-Q)^{-1}R} & 0 \end{bmatrix} \end{array}$$

UHE distribution

Results: Estimated distribution of Italian FDI by UHE

Comparison between Italian FDI bilateral data (blue) and the estimated distribution of Italian FDI by UHE (green), sorted in decreasing order by bilateral data. Only values over 1% are displayed.



Thank you for your attention!

Questions?