Correlation between upstreamness and downstreamness in random global value chains

Silvia Bartolucci
Fabio Caccioli
Francesco Caravelli
Pierpaolo Vivo

SRC Discussion Paper No 129
May 2024

Systemic Risk Centre

Discussion Paper Series

ISSN 2054-538X

Abstract
This paper is concerned with upstreamness and downstreamness of industries and countries in global value chains. Upstreamness and downstreamness measure respectively the average distance of an industrial sector from final consumption and from primary inputs, and they are computed from based on the most used global Input-Output tables databases, e.g., the World Input-Output Database (WIOD). Recently, Antràs and Chor reported a puzzling and counter-intuitive finding in data from the period 1995-2011, namely that (at country level) upstreamness appears to be positively correlated with downstreamness, with a correlation slope close to +1. This effect is stable over time and across countries, and it has been confirmed and validated by later analyses. We first analyze a simple model of random Input/Output tables, and we show that, under minimal and realistic structural assumptions, there is a natural positive correlation emerging between upstreamness and downstreamness of the same industrial sector/country, with correlation slope equal to +1. This effect is robust against changes in the randomness of the entries of the I/O table and different aggregation protocols. Secondly, we perform experiments by randomly reshuffling the entries of the empirical I/O table where these puzzling correlations are detected, in such a way that the global structural constraints are preserved. Again, we find that the upstreamness and downstreamness of the same industrial sector/country are positively correlated with slope close to +1, even though the random reshuffling has destroyed any underlying economic information about inter-sectorial connections and trends. Our results – rooted in the Complexity Science approach to economic problems – strongly suggest that the empirically observed puzzling correlation may rather be a necessary consequence of the few structural constraints (positive entries, and sub-stochasticity) that Input/Output tables and their surrogates must meet.

Keywords: Upstreamness, Downstreamness, Global Value Chains, Input/Output, Correlations.

This paper is published as part of the Systemic Risk Centre's Discussion Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1].

Silvia Bartolucci, University College London

Fabio Caccioli, University College London and Systemic Risk Centre, London School of Economics

Francesco Caravelli, Los Alamos National Laboratory

Pierpaolo Vivo, King's College London

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

ISSN 2054-538X

Correlation between upstreamness and downstreamness in random global value chains

Silvia Bartolucci[a,b], Fabio Caccioli[a,c,d], Francesco Caravelli[e], Pierpaolo Vivo[f]

[a]*Dept. of Computer Science, University College London, 66-72 Gower Street, London, WC1E 6EA, United Kingdom*
[b]*Centre for Financial Technology, Imperial College Business School, South Kensington, London, SW7 2AZ, United Kingdom*
[c]*Systemic Risk Centre, London School of Economics and Political Sciences, , London, WC2A 2AE, United Kingdom*
[d]*London Mathematical Laboratory, 8 Margravine Gardens, London, WC 8RH, United Kingdom*
[e]*Theoretical Division (T4), Condensed Matter & Complex Systems, Los Alamos National Laboratory, Los Alamos, 87545, New Mexico, U.S.A.*
[f]*Dept. of Mathematics, King's College London, Strand, London, WC2R 2LS, United Kingdom*

Abstract

This paper is concerned with upstreamness and downstreamness of industries and countries in global value chains. Upstreamness and downstreamness measure respectively the average distance of an industrial sector from final consumption and from primary inputs, and they are computed from based on the most used global Input-Output tables databases, e.g., the World Input-Output Database (WIOD). Recently, Antràs and Chor reported a puzzling and counter-intuitive finding in data from the period 1995-2011, namely that (at country level) upstreamness appears to be positively correlated with downstreamness, with a correlation slope close to +1. This effect is stable over time and across countries, and it has been confirmed and validated by later analyses. We first analyze a simple model of random Input/Output tables, and we show that, under minimal and realistic structural assumptions, there is a natural positive correlation emerging between upstreamness and downstreamness of the same industrial sector/country, with correlation slope equal to +1. This effect is robust against changes in the randomness of the entries of the I/O table and different aggregation protocols. Secondly, we perform experiments by randomly reshuffling the entries of the empirical I/O table where these puzzling correlations are detected, in such a way that the global structural constraints are preserved. Again, we find that the upstreamness and downstreamness of the same industrial sector/country are positively correlated with slope close to +1, even though the random reshuffling has destroyed any underlying economic information about inter-sectorial connections and trends. Our results – rooted in the Complexity Science approach to economic problems – strongly suggest that the empirically observed puzzling correlation may rather be a necessary consequence of the few structural constraints (positive entries, and sub-stochasticity) that Input/Output tables and their surrogates must meet.

Keywords: Upstreamness, Downstreamness, Global Value Chains, Input/Output, Correlations

1. Introduction

The structure of national and international trade flows has undergone a dramatic transformation in the past decades. Understanding how global value chains shape the exchange of goods and money at different scales (from industrial sectors to countries) has become of central importance. Researches on these issues usually rely on Input-Output analysis – the field pioneered by V. Leontief [1, 2]. This level of analysis is facilitated by the increasing availability and development of detailed Input/Output (I-O) tables for each country [3, 4].

To characterize the complexity of global value chains, metrics have been devised that take such empirical I-O tables as starting point. In particular, Antràs and Chor [5], Miller and Temurshoev [6] and Fally [7] introduced the notions of *upstreamness* and *downstreamness* to quantify the position of each economic sectors (and countries as a whole) with respect to final consumption, and raw materials, respectively (see Section 2 for details).

In a recent paper that has attracted much attention, Antràs and Chor [8] reported empirical observations of a puzzling correlation existing between the upstreamness and downstreamness of several countries[1] over many years (already noted in [6]). More precisely, they use data from the World Input-Output Database (WIOD) for the period 1995-2011, and observed that "countries that appear to be upstream according to their production-staging distance from final demand (U) are at the same time recorded to be downstream according to their production-staging distance from primary factors (D)", meaning that "countries that sell a disproportionate share of their output directly to final consumers (thus appearing to be downstream in GVCs according to U) tend to also feature high value-added over gross output ratios, reflecting a limited amount of intermediate inputs embodied in their production (thus appearing to be upstream in GVCs according to D)". A scatter plot of upstreamness vs. downstreamness at country level shows an evident linear relation with slope close to $+1$, an effect that persisted in all years of their sample – and that even intensified between 1995-2011 (see e.g. Figs. 4 and 5 in [8]). Similar effects are then shown also at the single country-industry level (see e.g. Fig. 10 in [8]).

Several explanatory factors have been put forward in [8] to make sense of these puzzling correlations, notably the possible persistence of large trade barriers across countries – which is however ruled out, as trade costs were found to have fallen off significantly over the period 1995-2011 – and the growing importance of the service sectors, which typically feature short production chain lengths and little use of intermediate inputs in production. We also mention that the standard definitions of upstreamness and downstreamness have been critically re-assessed, e.g. in [9], and alternative measures put forward there.

Besides looking at empirical data, it is sensible to corroborate the analysis with a complementary approach, namely the use of random models of interconnected economies, which have had a long and fruitful history in econometric studies [10–27]. The rationale is that whatever empirically observed effect survives randomization of the pairwise interaction between constituents cannot be due to any tailored and specific piece of information carried by the data, but must instead be generic and only due to global and structural constraints. In this spirit, we propose to look at the reported puzzling correlations between upstreamness and downstreamness at in Global Value Chains through the prism of (i) a random model of I-O tables, whose entries are drawn independently at random from a given distribution, preserving a few minimal structural constraints (essentially, non-negativity of the entries, and sub-stochasticity), for which the correlation between upstreamness and downstreamness can be tackled analytically, and (ii) a randomized process whereby the columns of an empirical I/O table where such correlations were detected are randomly reshuffled, therefore preserving the row sums of the original matrix. In both cases, we wish to see if randomization of the inter-sectorial dependencies destroys such correlations, as it would be natural to expect if these were due to finely tuned and subtle economic considerations. Contrary to our expectations, though, we find overwhelming evidence that it actually does not.

The paper is organized as follows. In Section 2 we provide the technical background, including the definition and interpretation of upstreamness and downstreamness, and how these measures are constructed from the I-O table. In Section 3 we first assume that the interaction matrix A between sectors is a random matrix, and then we construct the corresponding Upstreamness and Downstreamness matrices as well as the "covariance" and "slope" observables that we can monitor numerically and compute analytically in some cases. In Section 4, we provide our analytical results on a random model with exponential disorder, which shows that the scatter plot between upstreamness and downstreamness of the same sector has necessarily slope $+1$ for any matrix size N. These results are tested numerically in Section 5, along with numerical tests for other distributions of the entries of A, all confirming the same conclusions. In Section 6 we perform our random reshuffling experiment on empirical I-O matrices, which demonstrate that matrices satisfying the same structural constraints as the original one, but with any real economic information about inter-sectorial relations being wiped out, still display the same strong correlations between upstreamness and downstreamness as the original interaction matrix. Finally, in Section 6 we offer some critical discussion and concluding remarks.

[1]The upstreamness (or downstreamness) of a country is a weighted average upstreamness (downstreamness) of the economic sectors of the country (see Section 2 for details).

2. Definition of Upstreamness and Downstreamness

Antràs and Chor [5] considered a closed economy of N industries with no inventories – for instance, corresponding to a hypothetical single country that does not trade with others. For each industrial sector $i \in \{1, 2, \dots, N\}$ the value of gross output indicated with Y_i equals the sum of its use as a final good (F_i) and its use as an intermediate input to other industries (Z_i)

$$Y_i = F_i + Z_i = F_i + \sum_{j=1}^{N} a_{ij} \tag{1}$$

$$= F_i + \sum_{j=1}^{N} d_{ij} Y_j \ . \tag{2}$$

Here, d_{ij} is the dollar amount of sector i's output needed to produce one dollar's worth of sector j's output (see schematic structure of a I-O matrix for a single country in Fig. 1).

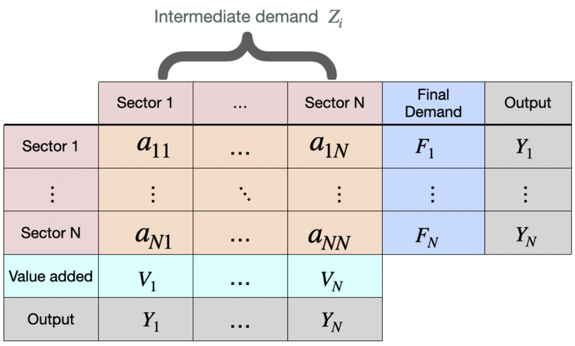

Figure 1: Scheme of the structure of a single-country input-output table [3, 4, 28].

Iterating the identity in (2), one obtains an infinite sequence of terms reflecting the use of sector i's output at different level in the value chain

$$Y_i = F_i + \sum_{j=1}^{N} d_{ij} F_j + \sum_{j=1}^{N} \sum_{k=1}^{N} d_{ik} d_{kj} F_j + \dots \ . \tag{3}$$

Using the matrix geometric series $\sum_{k \geq 0} D^k = [\mathbb{1}_N - D]^{-1}$, we can eventually rewrite (3) as

$$\boldsymbol{Y} = [\mathbb{1}_N - D]^{-1} \boldsymbol{F} \ , \tag{4}$$

where $\mathbb{1}_N$ is the $N \times N$ identity matrix, $D = (d_{ij})$ is the matrix of dollar values, and $\boldsymbol{F}$ the column vector of final demands.

Antràs and Chor [5] therefore proposed the following measure of upstreamness of the i-th industry, by multiplying each of the terms in (3) by their distance from final use, and dividing by Y_i

$$U_{1i} = 1 \cdot \frac{F_i}{Y_i} + 2 \cdot \frac{\sum_{j=1}^{N} d_{ij} F_j}{Y_i} + 3 \cdot \frac{\sum_{j,k=1}^{N} d_{ik} d_{kj} F_j}{Y_i} + \dots = \frac{([\mathbb{1}_N - D]^{-2} \boldsymbol{F})_i}{Y_i} \ , \tag{5}$$

where $(\cdot)_i$ indicates the i-th component of the vector.

Inserting (4) into (5), we can rewrite the upstreamness vector as

$$\boldsymbol{U}_1 = [\mathbb{1}_N - A_U]^{-1} \mathbf{1} \ , \tag{6}$$

where

$$A_U = Y^{-1} A = \begin{pmatrix} \frac{a_{11}}{Y_1} & \cdots & \frac{a_{1N}}{Y_1} \\ \vdots & \ddots & \vdots \\ \frac{a_{N1}}{Y_N} & \cdots & \frac{a_{NN}}{Y_N} \end{pmatrix} \tag{7}$$

and $Y = \mathrm{diag}(Y_1, \dots, Y_N)$. The matrix A_U has therefore non-negative elements, and is *row-substochastic* ($\sum_j (A_U)_{ij} \leq 1$ for all sectors j) because $(A_U)_{ij} = d_{ij} Y_j / Y_i$ is the share of sector i's total output that is purchased by industry j.

The upstreamness is defined in such a way that terms of the sum that are further upstream in the value chain have larger weight. By construction $U_{1i} \geq 1$ and is precisely equal to 1 if no output of industry i is used as input to other industries, that is the output of industry i is only used to satisfy the final demand.

Antràs et al. [29] later established an equivalence between their upstreamness measure and a measure – defined in a recursive fashion – of the "distance" of an industry from the final demand proposed independently by Fally [7]. Fally's upstreamnsess U_2 is defined as follows:

$$U_{2i} = 1 + \sum_{j=1}^{N} \frac{d_{ij} Y_j}{Y_i} U_{2j} \ . \tag{8}$$

The idea is that $\boldsymbol{U}_2$ aggregates information on the extent to which a sector in a given country produces goods that are sold directly to final consumers or that are sold to other sectors that themselves sell largely to final consumers. Sectors selling a large share of their output to relatively upstream industries should be therefore relatively upstream themselves. Using the fact that $d_{ij}Y_j = a_{ij}$ we have again that

$$\boldsymbol{U}_2 = [\mathbb{1}_N - A_U]^{-1}\mathbf{1} \ , \tag{9}$$

where A_U is defined in (7). In [8] an application of those measures for the analysis of empirical data on global value chains is presented.

On the input-side, there is an accounting identity that sector i's total input Y_i be equal to the value of its primary inputs (value added) V_i plus its intermediate input purchases from all other sectors:

$$Y_i = V_i + Z_i = V_i + \sum_{j=1}^{N} a_{ji} = V_i + \sum_{j=1}^{N} d_{ji}Y_j \ , \tag{10}$$

or in vector/matrix form

$$\boldsymbol{Y} = [\mathbb{1}_N - D^T]^{-1}\boldsymbol{V} \ . \tag{11}$$

Similarly to [5] (see (5)), Miller and Temurshoev [6] introduced the so-called *downstreamness* measuring the average distance between suppliers of primary inputs and sectors as input purchaser along the input demand supply chain as follows

$$D_{1i} = 1 \cdot \frac{V_i}{Y_i} + 2 \cdot \frac{\sum_{j=1}^{N} V_j d_{ji}}{Y_i} + \\ + 3 \cdot \frac{\sum_{j,k=1}^{N} V_j d_{jk} d_{ki}}{Y_i} + \cdots = \frac{([\mathbb{1}_N - D^T]^{-2}\boldsymbol{V})_i}{Y_i} \ . \tag{12}$$

As before, using (11), we obtain

$$\boldsymbol{D}_1 = [\mathbb{1}_N - A_D]^{-1}\mathbf{1} \tag{13}$$

with

$$\text{with } A_D = (AY^{-1})^T = \begin{pmatrix} \frac{a_{11}}{Y_1} & \cdots & \frac{a_{N1}}{Y_1} \\ \vdots & \ddots & \vdots \\ \frac{a_{1N}}{Y_N} & \cdots & \frac{a_{NN}}{Y_N} \end{pmatrix} \ . \tag{14}$$

Also the matrix A_D has therefore non-negative elements, and is *row-substochastic* ($\sum_j (A_D)_{ij} \leq 1$ for all sectors j). It is worth noting that by construction the matrices A_U and A_D share the diagonal elements a_{ii}/Y_i.

Finally, as in the upstreamness case, also for the downstreamness, Fally [7] introduced an analogous iterative definition of the form

$$D_{2i} = 1 + \sum_{j=1}^{N} d_{ji} D_{2j} \ , \tag{15}$$

which can be again mapped with simple manipulations into Eq. (13) using $Y_i d_{ji} = a_{ji}$.

The I-O Table in Fig. 1 can be modified in a conceptually simple way to account for inter-country trade by accommodating different inter-sectorial blocks (one for each country) – see scheme in Fig. 1 of [8]. The upstreamness (or downstreamness) of a country is then a suitably averaged (aggregate) version of the upstreamness (or downstreamness) of all industrial sectors of that country. In principle, there are two different ways to perform this aggregation. First, one could take the "giant" I-O table and collapse its entries at the country-by-country level by computing the total purchases of intermediate inputs by country j from country i – and then compute the upstreamness and the downstreamness on the collapsed (aggregate) table. Or, one could keep working with the giant table, compute the upstreamness and the downstreamness of industrial sectors within a country, and then perform a suitable average of those at country level. In [8], the two approaches were found to deliver extremely highly correlated country-level indices of GVC positioning.

3. The random model

Our randomized model is based on the closed-economy paradigm described in the previous section, and it assumes that the $N \times N$ matrices A_U and A_D (defined in Eqs. (7) and (14), respectively) are generated from a random interaction matrix A between sectors, i.e. without any structural information about the underlying dynamics of goods and prices apart from the constraint that their entries be non-negative, and that the matrices be row-substochastic. See subsection 3.2 for the precise definition of the random model.

3.1. Covariance and slope

Assuming therefore that the underlying model for the interaction matrix A is random, the covariance between the upstreamness $(\boldsymbol{U}_1)_i$ and downstreamness $(\boldsymbol{D}_1)_i$ (defined in Eqs. (6) and (13), respectively) of the same i-th sector is

$$\mathrm{Cov}((\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i) = \\ = \mathbb{E}\left[(\boldsymbol{U}_1)_i(\boldsymbol{D}_1)_i\right] - \mathbb{E}[(\boldsymbol{U}_1)_i]\mathbb{E}[(\boldsymbol{D}_1)_i] \ , \tag{16}$$

where the expectation $\mathbb{E}[\cdot]$ is taken w.r.t. the joint probability density function (pdf) of the entries of the matrix A (from which A_U and A_D are constructed). Since the upstreamness and downstreamness are defined in terms of a complicated matrix inversion, computing the covariance in Eq. (16) is a non-trivial task even for very simple joint pdfs of the entries of A.

However, we can take advantage of the results in Ref. [30, 31], which demonstrated that the "true" upstreamness and downstreamness (as defined in Eqs. (6) and (13), respectively) are individually correlated with simpler rank-1 estimators

$$\tilde{U}_i = 1 + \frac{r_i}{1 - (1/N)\sum_j r_j} \tag{17}$$

$$\tilde{D}_i = 1 + \frac{r'_i}{1 - (1/N)\sum_j r'_j} \ , \tag{18}$$

where $r_i = \sum_j (A_U)_{ij}$ are the row sums of A_U, and $r'_i = \sum_j (A_D)_{ij}$ are the row sums of A_D.

It is therefore sufficient to compute the covariance between the simpler estimators

$$\mathrm{Cov}(\tilde{U}_i, \tilde{D}_i) = \mathbb{E}[\tilde{U}_i \tilde{D}_i] - \mathbb{E}[\tilde{U}_i]\mathbb{E}[\tilde{D}_i] \tag{19}$$

to draw meaningful conclusions about the covariance between upstreamness and downstreamness as originally defined.

Noting that the quantities $(1/N)\sum_j r_j$ and $(1/N)\sum_j r'_j$ quickly converge to their non-fluctuating averages $\mathbb{E}[r]$ and $\mathbb{E}[r']$ by virtue of the Law of Large Numbers (LLN), we make the further simplifying move to replace these quantities with their non-fluctuating averages directly in the calculation of the covariance Eq. (19)[2]. Therefore, our covariance of interest reduces to the following object

$$\mathcal{C}_N = \frac{\mathbb{E}[rr'] - \mathbb{E}[r]\mathbb{E}[r']}{(1 - \mathbb{E}[r])(1 - \mathbb{E}[r'])} \ , \tag{20}$$

where we omitted the i-dependence (as every sector is statistically equivalent to any other in our random models). Therefore, r and r' can be viewed as the sum of, say, the first row of A_U and A_D, respectively.

We check with numerical simulations in Figs. 2 and 3 that indeed our conclusions are not affected by the fact that we considered simpler estimators *in lieu* of the original observables, as the former are perfectly correlated with the latter.

[2] More precisely, we make the approximation $\mathbb{E}\left[\frac{r_i}{1-(1/N)\sum_j r_j}\right] \approx \frac{\mathbb{E}[r]}{1-\mathbb{E}[r]}$, and similarly for r'.

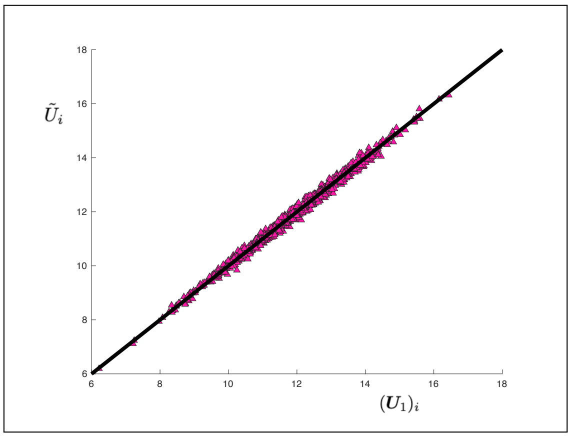

Figure 2: Scatter plot of the approximate upstreamness (Eq. (17)) vs. the "true" upstreamness $(\boldsymbol{U}_1)_i$ for our random model with exponential disorder. The parameters used are $\mu = 1$, $\mu_F = 0.1$, $N = 100$, $i = 7$. There are 1000 pairs of points in the figure, each obtained from a different instance of the random matrix A with exponential disorder.

The slope S of the scatter plot between $\tilde{D}_i$ and $\tilde{U}_i$ is easily determined from Eq. (19) by assuming first that there be a linear relation between the two, $\tilde{D}_i = S\tilde{U}_i$, and substituting in the expression for the covariance Eq. (19) we get

$$\mathrm{Cov}(\tilde{U}_i, \tilde{D}_i) = S\left\{\mathbb{E}[\tilde{U}_i^2] - \mathbb{E}[\tilde{U}_i]^2\right\} \ , \tag{21}$$

from which we deduce

$$S = \frac{\mathrm{Cov}(\tilde{U}_i, \tilde{D}_i)}{\mathrm{Var}[\tilde{U}_i]} \ , \tag{22}$$

where $\mathrm{Var}[\tilde{U}_i] = \mathbb{E}[\tilde{U}_i^2] - \mathbb{E}[\tilde{U}_i]^2$ is the variance of the approximate upstreamness.

Making again the further approximation that $(1/N)\sum_j r_j$ is replaced with its non-fluctuating average $\mathbb{E}[r]$ by virtue of the Law of Large Numbers, and after simple algebra from Eq. (18), we have that the slope S can be approximated by

$$S = \frac{\mathcal{C}_N\,(1 - \mathbb{E}[r])^2}{\mathbb{E}[r^2] - \mathbb{E}[r]^2} \ . \tag{23}$$

3.2. Model definition

In the random model we consider that the entries of A are drawn independently from an exponential probability density function (pdf) $p(a) = \mu \exp(-\mu a)$ with mean $1/\mu$. As such, the entries of A are all positive, and no economic or empirically motivated information whatsoever is injected

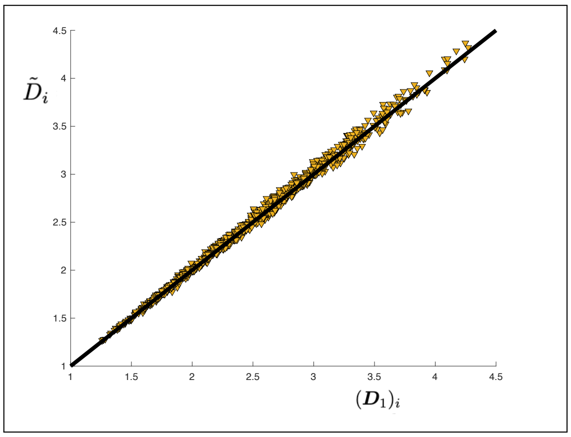

Figure 3: Scatter plot of the approximate downstreamness (Eq. (18)) vs. the "true" downstreamness $(\boldsymbol{D}_1)_i$ for our random model with exponential disorder. The parameters used are $\mu = 1$, $\mu_F = 0.01$, $N = 100$, $i = 7$. There are 1000 pairs of points in the figure, each obtained from a different instance of the random matrix A with exponential disorder.

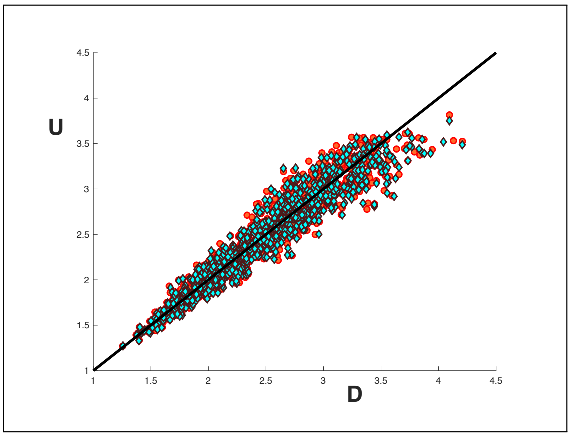

Figure 4: Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 200$ random model with exponential disorder with parameters $\mu = 1$, $\mu_F = 0.005$. Each of the 1000 blue [red] points is obtained from one instance of the random "source" matrix A, and represents the pair of values $((\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i)$ $[(\tilde{U}_i, \tilde{D}_i)$, respectively]. The thick black line has slope $+1$.

in the construction of A. The final demand values F_i $(i = 1, \ldots, N)$ are further modeled as i.i.d. exponential random variables with mean $1/\mu_F$.

From the matrix A, we construct the matrices A_U and A_D (see Eq. (7) and (14), together with the definition of Y_i in Eq. (1)) as

$$(A_U)_{ij} = \frac{a_{ij}}{\sum_j a_{ij} + F_i} \tag{24}$$

$$(A_D)_{ij} = \frac{a_{ji}}{\sum_j a_{ij} + F_i} \ , \tag{25}$$

where we used that $\sum_j a_{ij} + F_i = \sum_j a_{ji} + V_i$ for all i, as follows from the accounting identity. Therefore, provided that μ_F is sufficiently small[3], both A_U and A_D as defined above have non-negative elements, and are row sub-stochastic (as they should).

For each instance of the random matrix A and of the vector of final demands $\boldsymbol{F}$, we construct the matrices A_U and A_D as above, and from those we compute the pairs $(\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i$ and $\tilde{U}_i, \tilde{D}_i$ for any sector i that we choose. These are all random variables, whose pairwise covariance is of interest in this paper.

4. Results

Our results are summarized in the theorem and corollary below. We show that even in our completely random model (with no economic or empirically motivated information whatsoever injected in constructing the I/O table), the upstreamness and downstreamness of an industrial sector of a single country are necessarily positively correlated, and that for any N the slope of the scatter plot between the two is always equal to $+1$.

In Figs. 5 and 6 we further numerically check that our results do not crucially depend on the specific choice of the way the random matrices A_U and A_D are generated, so the positive correlation between upstreamness and downstreamness of economic sectors – and their correlation slope being $+1$ – seem to be very robust results and rather insensitive to the fine details of the inter-sectorial I/O matrix.

Theorem 1. *Let $N \times N$ matrices be defined as*

$$(A_U)_{ij} = \frac{a_{ij}}{\sum_j a_{ij} + F_i} \tag{26}$$

$$(A_D)_{ij} = \frac{a_{ji}}{\sum_j a_{ij} + F_i} \ , \tag{27}$$

where a_{ij} are i.i.d. variables drawn from an exponential pdf $p(a) = \mu \exp(-\mu a)$, and the F_i's are i.i.d. variables drawn from an exponential pdf $p_F(F) = \mu_F \exp(-\mu_F F)$ with $\mu_F \ll \mu$ to ensure that A_U and A_D are row sub-stochastic. Let $r = \sum_j (A_U)_{1j}$ and $r' = \sum_j (A_D)_{1j}$. Then the simplified covariance between upstreamness and down-

[3] This condition is necessary to ensure that F_i's will be (typically) sufficiently large that $\sum_j (A_D)_{ij}$ in (25) will be smaller than 1.

streamness (see Eq. (20))

$$\mathcal{C}_N(\mu,\mu_F) = \frac{\mathbb{E}[rr'] - \mathbb{E}[r]\mathbb{E}[r']}{(1-\mathbb{E}[r])(1-\mathbb{E}[r'])} \qquad (28)$$

is given by the exact formula $\mathcal{C}_N(\mu,\mu_F) = -\mathcal{N}_N(\phi)/\mathcal{D}_N(\phi)$*, where*

$$\begin{aligned}\mathcal{N}_N(\phi) = (\phi-1)\big[&N(\phi-1)\mathrm{B}(1,N+1)^2\,{}_2F_1(1,N+1;N+2;\phi)^2\\ &+(N-1)N(\phi-1)\mathrm{B}(1,N+1)\,{}_2F_1(1,N+1;N+2;\phi)(\mathrm{B}(1,N)\,{}_2F_1(1,N;N+1;\phi)+1)\\ &+(N+1)(\phi-1)\mathrm{B}(1,N+2)\,{}_2F_1(1,N+2;N+3;\phi)+N\big]\\ \mathcal{D}_N(\phi) = ((N-1)&(\phi-1)\mathrm{B}(1,N)\,{}_2F_1(1,N;N+1;\phi)+(\phi-1)\mathrm{B}(1,N+1)\,{}_2F_1(1,N+1;N+2;\phi)\\ &+1)(N(\phi-1)\mathrm{B}(1,N+1)\,{}_2F_1(1,N+1;N+2;\phi)+1)\ ,\end{aligned}$$

where $\phi = 1-\mu_F/\mu$*. Here,* $\mathrm{B}(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ *is the Beta function, and* ${}_2F_1$ *is the Gaussian hypergeometric function.*

Corollary 1. *In the hypotheses of Theorem 1, the slope* $S(\mu,\mu_F)$ *of the scatter plot between the rank-1 estimators of downstreamness and upstreamness (see Eq. (23)) is equal to* $+1$ *for any* N*, irrespective of the values of* μ,μ_F*.*

The proofs are deferred to the Appendix.

5. Numerical simulations

We have performed numerical simulations on our random model, generating m instances of the $N\times N$ matrix A with i.i.d. exponential entries with mean $1/\mu$. We also generate random vectors of final demands $\boldsymbol{F}$ of size N, with i.i.d. entries with mean $1/\mu_F$ (with $\mu_F \ll \mu$).

For each generated instance of the matrix A, we formed the matrices A_U and A_D (as defined in Eqs. (7) and (14)), which have by construction non-negative elements, and are row sub-stochastic[4].

From the matrices A_U and A_D so generated, we constructed the vectors of upstreamness $\boldsymbol{U}_1$ and downstreamness $\boldsymbol{D}_1$ values according to the inversion formulae Eq. (6) and (13), respectively. We then pick a certain sector index i (for example, $i=7$), and for that index we compute the estimators $\tilde{U}_i$ and $\tilde{D}_i$ according to Eqs. (17) and (18) respectively.

We first show in Figs. 2 and 3 that the "true" upstreamness (downstreamness) of sector i – computed from the full inversion formulae – is perfectly correlated (with correlation slope = +1) with its approximate estimator. It is therefore perfectly justified to use the approximate estimators (instead of the full definition) to study correlations, as those are much simpler to handle analytically.

Next, in Table 1, we report values of the "true" covariance between upstreamness and downsteamness of sector $i=7$, obtained from averaging over $m=10000$ numerically generated instances of our random model, against the values of $\mathcal{C}_N(\mu,\mu_F)$ analytically computed, and we observe an excellent agreement between the two.

In Fig. 4 we further provide scatter plots of upstreamness vs. downstreamness of sector i (both "true" and approximate) - where each generated instance contributes a single point to the scatter plot. Again, we observe an excellent collapse onto the diagonal line with slope +1, further confirming that a strong positive correlation between upstreamness and downstreamness of the same sector is a generic feature of "structure-less" matrices – provided they have non-negative entries and are sub-stochastic.

Finally, in Figs. 5 and 6 we provide the same scatter plots as in Fig. 4, but this time for the "original" matrix A (and similarly for the vector $\boldsymbol{F}$) having i.i.d. non-negative entries drawn from a log-normal ($p(a) = (a\sqrt{2\pi})^{-1}\exp(-(\ln(a)-\mu')^2/2)$) and uniform (with mean $1/\mu$ and $1/\mu_F$) pdf, respectively. Although we do not provide analytical results for these cases, these plots further confirm that the positive correlation with slope +1 between upstreamness and downstreamness keeps holding irrespective of the precise details of the way the "source" matrix A is generated – provided that A_U and A_D are non-negative and sub-stochastic.

[4] For A_U substochasticity is guaranteed by construction. For A_D this is true with overwhelming likelihood provided that $\mu_F \ll \mu$.

Table 1: Covariance between Upstreamness and Downstreamness in the random model for $i = 7$ and taken over $m = 10000$ samples

μ	μ_F	N	$\mathrm{Cov}((\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i)$	$\mathcal{C}_N(\mu, \mu_F)$
1	0.001	200	0.10450	0.10385
2	0.005	400	0.30415	0.29494
3	0.001	300	0.06136	0.06158
1.2	0.001	500	0.17346	0.17260
1.5	0.003	350	0.24224	0.23955

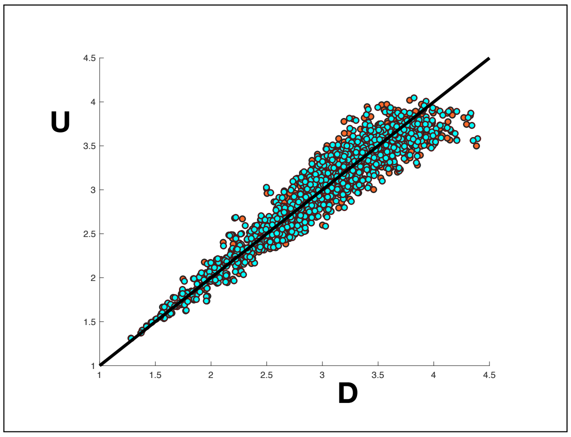

Figure 5: Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 400$ random model with log-normal disorder with parameters $\mu' = 1$, $\mu'_F = 6.67$. Each of the 1000 light blue [orange] points is obtained from one instance of the random "source" matrix A, and represents the pair of values $((\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i)$ $[(\tilde{U}_i, \tilde{D}_i)$, respectively]. The thick black line has slope +1.

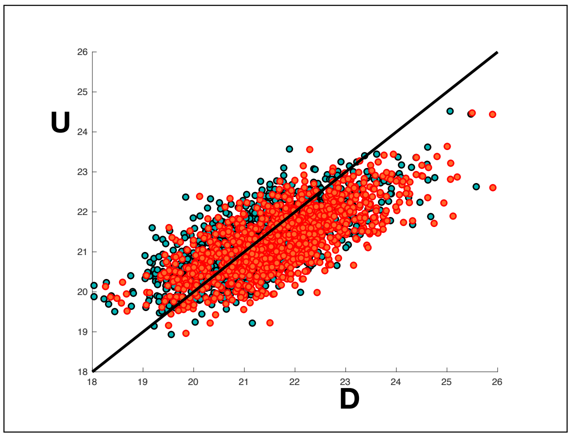

Figure 6: Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 400$ random model with uniform disorder with parameters $\mu = 1$, $\mu_F = 0.05$. Each of the 1000 light blue [orange] points is obtained from one instance of the random "source" matrix A, and represents the pair of values $((\boldsymbol{U}_1)_i, (\boldsymbol{D}_1)_i)$ $[(\tilde{U}_i, \tilde{D}_i)$, respectively]. The thick black line has slope +1.

6. Random reshuffling of the I-O Table

To perform the second series of experiments, we have taken the empirical I-O matrices including 39 countries for the years 1995-2011 (from WIOD 2013 release). For each country and each year, we have computed the upstreamness and downstreamness of that country using Eq. (9) and Eq. (13) respectively, averaging over sectors. In fig. 7, we show the values of upstreamness for all countries in all years, spanning a period identical to that considered in the paper [8]. In Fig. 8, we provide the scatter plot of upstreamness vs. downstreamness of each country for three selected years (1996-2003-2011). As expected, we confirm the general trend observed in [8], namely that the two measures appear to be strongly correlated with a slope of the scatter plot very close to +1.

In Fig. 9, though, we took the same I-O matrices for the entire period 1995-2011 and we randomly reshuffled the columns of such matrices according to a random permutation of the set $\{1, \ldots, N\}$. The resulting matrices satisfy the same aggregate constraints (namely, the row sums) of the orginal, actual matrices, but the interactions between sectors have been randomly scrambled, resulting in an entirely fictitious Global Value Chain, where all economic forces at play in the real world have been neutralized. Still, and quite surprisingly, we find that the same linear correlation with slope close to +1 between upstreamness and downstreamness survives. We have checked that this result is not an artifact of the specific random permutation of columns chosen, but keeps holding irrespective of what new "strength" of interaction is attributed to pairs of sectors/countries via a random reshuffling of the old, actual one.

The fact that U-D correlations are so strong and stable that they survive a complete overhauling of the actual economic interactions at play in the real world provides a further strong confirmation that most – if not all – of such correlations cannot be due to sophisticated and finely-tuned economic factors leading to a *specific* set of inter-sectorial interactions, otherwise any random reshuffling would have completely annihilated them. These experiments therefore lend further support to the claim that U-D correlations are mostly due to structural constraints that the matrices A_U and A_D must meet simply because of the way they are constructed from the interaction matrix A.

Of course, it remains an open problem to then devise a better measure of downstreamness that is truly independent of the upstreamness.

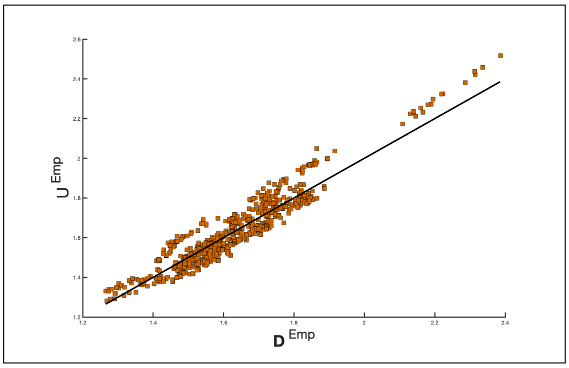

Figure 7: Empirical upstreamness vs. empirical downstreamness (averaged over 35 sectors) for 39 countries for the years 1995-2001.

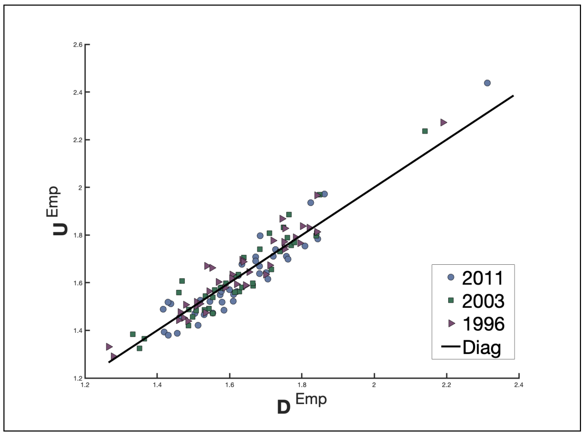

Figure 8: Empirical upstreamness vs. empirical downstreamness (averaged over 35 sectors) for 39 countries in selected years (1996-2003-2011).

7. Discussion

In summary, we have considered two classes of *random* Input/Output matrices A to test whether the "puzzling" correlation detected between upstreamness and downstreamness at the sector and country level [8] would survive even if the underlying economic forces and the inter-sectorial dependencies had nothing to do with the real world.

First, we constructed a random model for the matrix A that mimics a closed economy composed of N economic sectors. We showed analytically that

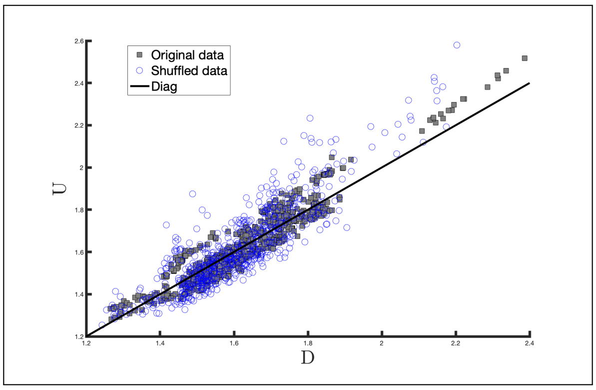

Figure 9: Upstreamness vs. downstreamness (averaged over 35 sectors) calculated on the empirical and reshuffled matrices for 39 countries for the years 1995-2001. Blue circles represent values calculated on I-O matrices where columns have been randomly reshuffled. Grey squares are upstreamness/donwstreamness pairs calculated on the original data.

the resulting upstreamness and downstreamness of a given sector are generically positively correlated, with a slope of the scatter plot between the two equal to $+1$, if the entries of the matrix A are independently drawn from an exponential pdf. We also showed numerically that our results do not depend very strongly on the pdf of matrix entries. At least at the level of single countries, our work provides a comforting "proof of principle" that a strong positive correlation between upstreamness and downstreamness of individual sectors as originally defined (see Eqs. (6) and (13)) is bound to materialize even on a structure-less and zero-information I-O matrix: one would have to try very hard to concoct an I-O matrix so extraordinary and finely tuned, that such correlations were *not* observed.

Secondly, we started from a real, empirical I-O matrix A taken from the WIOD Dataset (2013 Release), which displayed the same kind of correlations between observables as originally detected in [8]. We performed several experiments where we simply randomly reshuffled the columns of the interaction matrix prior to computing the matrices A_U and A_D from which upstreamness and downsteamness of the same sector can be determined. The resulting shuffled interaction matrix A' is a new, perfectly legitimate interaction matrix, which shares the row sums and all other structural constraints with the original matrix A, but whose economic fabric and inter-sectorial dependencies are entirely made up: the Global Value Chain that A' embodies does not respond to any realistic economic force nor is connected to any realistic economic scenario. Yet, we find that the correlations survive unscathed.

Although derived in the context of a closed-economy I-O table (see Fig. 1), our results are nevertheless also relevant to the more general setting of international trade considered in [8] and the puzzling correlation highlighted thereof, for a few reasons: (i) the "giant" I-O matrix that includes inter-country trade blocks satisfies the same constraints (non-negative entries, and row-stochasticity) as the single-country one (and thus as well as the random model we presented). (ii) Preliminary numerical experiments played on random A matrices with a block structure similar to the giant I-O matrix did not present evidence of any combination of heterogeneity in the distribution of entries, sparsity patterns of blocks, or aggregation of outcomes (either before or after the country upstreamness/downstreamness were computed) that was ever able to break the robust positive correlation between upstreamness and downstreamness observed again in that more general setting – this time, at the country level – and entirely akin to the simpler setting described in this paper. (iii) The model considered here – after a trivial re-interpretation of the matrix entries – is at the very least expected to mimic rather accurately what would happen in the more general inter-country setting in the two extreme cases of *zero* and *infinite* trade barriers between different countries. In the former case, the "giant" I-O matrix will have inter-country and intra-country blocks that do not differ much (statistically), therefore – after country-wise aggregation – the resulting A matrix will look very similar to the one we considered in this paper. In the latter case, the "giant" I-O matrix will be block-diagonal – with inter-country blocks full of zeros due to the absence of trade – with each non-zero (intra-country) block being an independent replica of the closed economy model proposed here. While a deeper investigation of the intermediate trade barriers setting in the random "giant" model is surely needed, the aforementioned observations sharply point towards the observed correlation between upstreamness and downstreamness also at country level being simply due to structural and unavoidable algebraic constraints that I-O tables and their surrogates must satisfy.

Our first series of results rest on the following assumptions and simplifications:

1. The correlation between "true" upstreamness and downstreamness of a sector can be faithfully probed by using the rank-1 approximants defined in [30, 31]. This assumption was tested on empirical I/O data in [30, 31], and on the random model here in Figs. 2 and 3, by showing that the "true" upstreamness (or downstreamness) is indeed perfectly correlated with its rank-1 approximant. Such rank-1 approximation could only become less reliable if the true Input/Output matrix were exceedingly sparse, i.e. with a very large number of zero entries (see discussion in [31–33]), a situation that does not often materialize in practice. By considering national I-O tables available from the 2013 release of the WIOD [4], we indeed obtain quite high average densities of nonzero elements – between 0.92 and 0.93 across 40 countries for the years 1995-2011. We have further checked that a moderate sparsification of our random model does not qualitatively change our conclusions, however in future experiments it will be appropriate to test the consequences of sparsity more thoroughly.

2. We have assumed that the entries of the matrix A were independent and identically distributed (i.i.d.). Some preliminary results (not shown) where this assumption has been relaxed indicate that heterogeneity in the pdfs of the entries of A may not play a major role and is generally insufficient to change the conclusions of our analysis.

3. We used some simplifications (for instance, appealing to the LLN) to make some progress in the analytical calculations. All approximations are controlled and have been carefully tested.

Apart from performing a more thorough study on the effect of sparsity and heterogeneity in random models of I-O tables, in future studies it will be interesting to try to compute analytically the *full* covariance Eq. (16) for our random model and for various different pdfs of the entries of the I/O matrix A, i.e. without employing any rank-1 proxy and/or LLN approximations. This task will require handling the average of (products of) inverse matrices (coming from the definitions of upstreamness and downstreamness, see Eq. (6) and (13)), which is possible in some cases using techniques from statistical physics [30].

Appendix A. Derivation of Theorem 1 and Corollary 1

We need to compute $\mathbb{E}[r]$, $\mathbb{E}[r']$ and $\mathbb{E}[rr']$ separately. We have

$$\mathbb{E}[r] = \int_0^\infty \prod_{i=1}^N da_i p(a_i) \int_0^\infty dF\ p_F(F) \frac{\sum_k a_k}{\sum_k a_k + F}\ , \tag{A.1}$$

where for simplicity we denoted $a_k \equiv a_{1k}$ and $F \equiv F_1$. Using the identity

$$\frac{1}{\xi} = \int_0^\infty ds\ e^{-\xi s} \qquad \xi > 0 \tag{A.2}$$

we have

$$\begin{aligned}\mathbb{E}[r] &= \mu^N \mu_F \int_0^\infty \prod_{i=1}^N da_i \int_0^\infty dF ds\ \sum_k a_k \times \\ &\times e^{-\mu \sum_k a_k - \mu_F F - s(\sum_k a_k + F)} \\ &= \mu^N \mu_F N \int_0^\infty ds \left[\int_0^\infty da e^{-\mu a - sa}\right]^{N-1} \times \\ &\times \int_0^\infty dy\ y e^{-\mu y - sy} \int_0^\infty dF e^{-\mu_F F - sF} \\ &= \boxed{\mu^N \mu_F N J(N+1)}\end{aligned} \tag{A.3}$$

where

$$\begin{aligned}J(k) &= \int_0^\infty ds \frac{1}{(\mu+s)^k} \frac{1}{\mu_F + s} = \\ &= \frac{1}{\mu^k} \int_0^\infty dt (1+t)^{-1} \left(1 + \frac{\mu_F}{\mu} t\right)^{-k} = \\ &= \frac{1}{\mu^k} \mathrm{B}(1,k)\ {}_2F_1(k,1;k+1;1-\mu_F/\mu)\end{aligned} \tag{A.4}$$

from [34], formula 3.197.5 (pag. 335) with $\lambda = 1$, $\nu = -1$, $\alpha = \mu_F/\mu$ and $\tilde{\mu} = -k$. Here, $\mathrm{B}(\cdot,\cdot)$ is the Beta function, and ${}_2F_1$ is a hypergeometric function.

Similarly

$$\begin{aligned}\mathbb{E}[r'] &= \mu^{2N-1} \mu_F \int_0^\infty \prod_{i=1}^N da_i p(a_i) \prod_{j=2}^N db_j p(b_j) \times \\ &\times \int_0^\infty dF\ p_F(F) \frac{a_1 + \sum_{k\geq 2} b_k}{\sum_k a_k + F} \\ &= \boxed{\mu^{2N-1} \mu_F \left[\frac{1}{\mu^{N-1}} J(N+1) + \frac{N-1}{\mu^N} J(N)\right]}\ ,\end{aligned} \tag{A.5}$$

where for simplicity we denoted $a_{1j} \equiv a_j$ (for $j = 1, \ldots, N$), and $a_{k1} \equiv b_k$ (for $k = 2, \ldots, N$). To prove this, we write

$$\mathbb{E}[r'] = \mu^{2N-1} \mu_F [I_1 + I_2]\ , \tag{A.6}$$

where

$$\begin{aligned}I_1 &= \int_0^\infty ds \left[\int_0^\infty dx e^{-\mu x - sx}\right]^{N-1} \int_0^\infty dy\ y e^{-\mu y - sy} \\ &\times \left[\int_0^\infty dz\ e^{-\mu z}\right]^{N-1} \int_0^\infty dF e^{-\mu_F F - sF} \\ &= \frac{1}{\mu^{N-1}} \int_0^\infty ds \frac{1}{(\mu+s)^{N+1}(\mu_F+s)} = \boxed{\frac{J(N+1)}{\mu^{N-1}}}\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}I_2 &= (N-1) \int_0^\infty ds \left[\int_0^\infty dx\ e^{-\mu x - sx}\right]^N \times \\ &\times \int_0^\infty dF e^{-\mu_F F - sF} \left[\int_0^\infty dy e^{-\mu y}\right]^{N-2} \times \\ &\int_0^\infty dz\ z e^{-\mu z} = \frac{N-1}{\mu^N} \int_0^\infty ds \frac{1}{(\mu+s)^N (\mu_F+s)} \\ &= \boxed{\frac{N-1}{\mu^N} J(N)}\ .\end{aligned}$$

Finally

$$\begin{aligned}&\mathbb{E}[rr'] = \mu^{2N-1} \mu_F \int_0^\infty \prod_{i=1}^N da_i p(a_i) \prod_{j=2}^N db_j p(b_j) dF \\ &\frac{\sum_\ell a_\ell}{\sum_k a_k + F} \frac{a_1 + \sum_{k\geq 2} b_k}{\sum_k a_k + F} = \\ &= \mu^{2N-1} \mu_F \left[\frac{2}{\mu^{N-1}} L(N+2) + \frac{N-1}{\mu^N} L(N+1)\right. \\ &\left. + \frac{N-1}{\mu^{N-1}} L(N+2) + \frac{(N-1)^2}{\mu^N} L(N+1)\right]\ ,\end{aligned} \tag{A.8}$$

where

$$\begin{aligned}L(k) &= \int_0^\infty ds dt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^k} = \\ &= \frac{\mu^{1-k}}{k-1} - \mu_F J(k)\ .\end{aligned} \tag{A.9}$$

Eq. (A.8) follows from writing $\sum_\ell a_\ell = a_1 + \sum_{\ell \neq 1} a_\ell$, and applying the "lifting-up" identity (A.2) twice, which yields

$$\mathbb{E}[rr'] = \mu^{2N-1} \mu_F [K_1 + K_2 + K_3 + K_4]\ , \tag{A.10}$$

where

$$K_1 = \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N dsdt$$
$$e^{-\mu \sum_k x_k - \mu \sum_{k\geq 2} y_k - \mu_F F_1} x_i^2 \times$$
$$\times\, e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} =$$
$$= \frac{1}{\mu^{N-1}} \int dsdt \frac{1}{\mu_F + s + t} \left[\int dx\; e^{-\mu x - (s+t)x} \right]^{N-1}$$
$$\times \int dx\; x^2 e^{-\mu x - (s+t)x}$$
$$= \frac{2}{\mu^{N-1}} \int_0^\infty dsdt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+2}} =$$
$$= \boxed{\frac{2}{\mu^{N-1}} L(N+2)} \,. \tag{A.11}$$

$$K_2 = \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N dsdt$$
$$\times\, e^{-\mu \sum_k x_k - \mu \sum_{k\geq 2} y_k - \mu_F F_1} x_i \sum_{k\geq 2} y_k \times$$
$$\times\, e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} =$$
$$= (N-1) \int dsdt \frac{1}{\mu_F + s + t} \left[\int dx\; e^{-\mu x - (s+t)x} \right]^{N-1}$$
$$\times \int dx\; x e^{-\mu x - (s+t)x}$$
$$\times \left[\int dy e^{-\mu y} \right]^{N-2} \int dy\; y e^{-\mu y}$$
$$= \frac{N-1}{\mu^N} \int_0^\infty dsdt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+1}} =$$
$$= \boxed{\frac{N-1}{\mu^N} L(N+1)} \,. \tag{A.12}$$

$$K_3 = \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N dsdt$$
$$\times\, e^{-\mu \sum_k x_k - \mu \sum_{k\geq 2} y_k - \mu_F F_1} x_i \sum_{\ell \neq i} x_\ell \times$$
$$\times\, e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} =$$
$$= (N-1) \int dsdt \frac{1}{\mu_F + s + t} \left[\int dx\; x e^{-\mu x - (s+t)x} \right]^2$$
$$\times \left[\int dx\; e^{-\mu x - (s+t)x} \right]^{N-2} \times \left[\int dy e^{-\mu y} \right]^{N-1}$$
$$= \frac{N-1}{\mu^{N-1}} \int_0^\infty dsdt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+2}} =$$
$$= \boxed{\frac{N-1}{\mu^{N-1}} L(N+2)} \,. \tag{A.13}$$

$$K_4 = \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N dsdt$$
$$\times\, e^{-\mu \sum_k x_k - \mu \sum_{k\geq 2} y_k - \mu_F F_1} \sum_{k\geq 2} y_k \sum_{\ell \neq i} x_\ell \times$$
$$\times\, e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} =$$
$$= (N-1)^2 \int dsdt \frac{1}{\mu_F + s + t}$$
$$\times \left[\int dx\; x e^{-\mu x - (s+t)x} \right] \left[\int dx\; e^{-\mu x - (s+t)x} \right]^{N-1}$$
$$\times \left[\int dy e^{-\mu y} \right]^{N-2} \int dy\; y e^{-\mu y}$$
$$= \frac{(N-1)^2}{\mu^N} \int_0^\infty dsdt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+1}} =$$
$$= \boxed{\frac{(N-1)^2}{\mu^N} L(N+1)} \,. \tag{A.14}$$

Collecting all terms and simplifying, we arrive at the formula announced in Theorem 1. Plotting the covariance formula as a function of N for different values of μ, μ_F reveals that the covariance is always positive and increasing (see Fig. A.10).

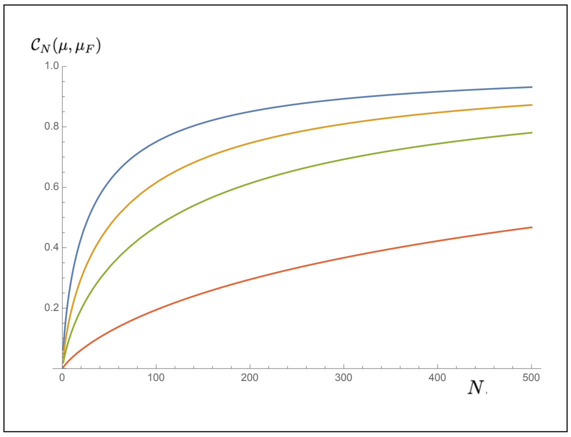

Figure A.10: Covariance $\mathcal{C}_N(\mu, \mu_F)$ between the approximate upstreamness and downstreamness for the random model (exact formula in Theorem 1). The parameters (μ, μ_F) are $(1, 0.1)$ (blue), $(2, 0.1)$ (orange), $(2, 0.05)$ (green), $(2, 0.01)$ (red).

To prove the Corollary 1, we need to further compute $\mathbb{E}[r^2]$ and then simplify the resulting expression for the slope (23), yielding a slope $= +1$ for any N.

$$\mathbb{E}[r^2] = \int_0^\infty \prod_{i=1}^N da_i p(a_i) \int_0^\infty dF\ p_F(F) \left[\frac{\sum_k a_k}{\sum_k a_k + F}\right]^2$$

$$= \mu^N \mu_F \int d\boldsymbol{a} dF ds\ s e^{-\mu \sum_k a_k - \mu_F F} \times$$

$$\left(\sum_k a_k\right)^2 e^{-s(\sum_k a_k + F)}$$

$$= \mu^N \mu_F \left[N \int d\boldsymbol{a} dF ds\ s e^{-\mu \sum_k a_k - \mu_F F} a_1^2 e^{-s(\sum_k a_k + F)} + \right.$$

$$\left. + (N^2 - N) \int d\boldsymbol{a} dF ds\ s e^{-\mu \sum_k a_k - \mu_F F} a_1 a_2 e^{-s(\sum_k a_k + F)}\right]$$

$$= \mu^N \mu_F N \int_0^\infty ds\ s \left[\int dx e^{-\mu x - sx}\right]^{N-1} \int dy\ y^2 e^{-\mu y - sy} \times$$

$$\times \int dF e^{-\mu_F F - sF} + \mu^N \mu_F (N^2 - N) \int_0^\infty ds\ s \times$$

$$\times \left[\int dx e^{-\mu x - sx}\right]^{N-2} \left[\int dy\ y e^{-\mu y - sy}\right]^2 \int dF e^{-\mu_F F - sF}$$

$$= 2\mu^N \mu_F N \int_0^\infty ds \frac{s}{(\mu + s)^{N+2}} \frac{1}{\mu_F + s} +$$

$$+ \mu^N \mu_F (N^2 - N) \int_0^\infty ds \frac{s}{(\mu + s)^{N+2}} \frac{1}{\mu_F + s}$$

$$= \boxed{\mu^N \mu_F (N^2 + N) L(N + 2)}\ \blacksquare$$

References

[1] W. Leontief (1986). Input-output economics. Oxford University Press.

[2] W. Leontief (1936). Quantitative Input-Output Relations in the Economic System of the United States. Rev. Econ. Stat. **18**, pp. 105–125.

[3] United Nations Department for Economic and Social Affairs Statistics Division (1999). Handbook of Input-Output Table Compilation and Analysis.

[4] M. P. Timmer, E. Dietzenbacher, B. Los, R. Stehrer, and G. J. de Vries (2015). An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production. Review of International Economics **23**, 575–605.

[5] P. Antràs and D. Chor (2012). Organizing the global value chain. Econometrica **81**(6), 2127-2204.

[6] R. E. Miller and U. Temurshoev (2015). Output upstreamness and input downstreamness of industries/countries in world production. International regional science review **40**(5), 443-475.

[7] T. Fally (2012). Production Staging: Measurement and Facts (unpublished).

[8] P. Antràs and D. Chor (2018). On the measurement of upstreamness and downstreamness in global value chains. Working Paper 24185 `http://www.nber.org/papers/w24185`

[9] Z. Wang, S. J. Wei, X. Yu, and K. Zhu (2017). Characterizing global value chains: production length and upstreamness. Working paper No. w23261. National Bureau of Economic Research.

[10] B. Peterson and M. Olinick (1982). Leontief models, Markov chains, Substochastic matrices, and positive solutions of matrix equations. Mathematical modelling **3**, 221-239.

[11] J. McNerney, C. Savoie, F. Caravelli, and J. D. Farmer (2021). How production networks amplify economic growth. PNAS **119** (1) e2106031118.

[12] J. McNerney (2012). Applications of Statistical Physics to Technology Price Evolution. Ph.D. thesis, Boston University.

[13] P. S. M. Kop Jansen (1994). Analysis of multipliers in stochastic input-output models. Regional Science and Urban Economics **24**, 55-74.

[14] P. Kop Jansen and T. Ten Raa (1990). The choice of model in the construction of Input-Output coefficients matrices. International Economic Review **31**, No. 1, pp. 213- 227.

[15] W. D. Evans (1954). The Effect of Structural Matrix Errors on Interindustry Relations Estimates. Econometrica **22** (n. 4), pp. 461-480.

[16] R. E. Quandt (1958). Probabilistic errors in the Leontief system. Naval Research Logistics Quarterly **5**, pp. 155-170.

[17] A. Simonovits (1975). A Note on the Underestimation and Overestimation of the Leontief Inverse. Econometrica **43**, pp. 493-498.

[18] G. R. West (1986). A stochastic analysis of an input-output model. Econometrica: Journal of the Econometric Society **54**, 363-374.

[19] H. Kogelschatz (2007). On the Solution of Stochastic input-output-Models. Discussion Paper Series n. 447, University of Heidelberg.

[20] M. Kozicka (2019). Novel approach to stochastic Input-Output modeling. RAIRO-Oper. Res. **53**, 1155–1169 `https://doi.org/10.1051/ro/2018046`.

[21] J. L. Katz and R. L. Burford (1985). Shortcut formulas for output, income and employment multipliers. The Annals of Regional Science **19**(2), 61-76.

[22] R. L. Burford (1977). Regional input-output multipliers without a full IO table. The Annals of Regional Science **11**(3), 21-38.

[23] R. L. Drake (1976). A short-cut to estimates of regional input-output multipliers: methodology and evaluation. International Regional Science Review **1**(2), 1-17.

[24] P. J. Phibbs and A. J. Holsman (1981). An evaluation of the Burford Katz short cut technique for deriving input-output multipliers. The Annals of Regional Science **15**(3), 11-19.

[25] R. C. Jensen and G. J. D. Hewings (1985). Shortcut 'Input-Output' Multipliers: The Resurrection Problem (a Reply). Environment and Planning A **17**(11), 1551-1552.

[26] R. C. Jensen and G. J. D. Hewings (1985). Shortcut 'input-output' multipliers: A requiem. Environment and Planning A **17**(6), 747-759.

[27] R. L. Burford and J. L. Katz (1985). Shortcut 'input-output' multipliers, alive and well: Response to Jensen and Hewings. Environment and Planning A **17**(11), 1541-1549.

[28] Suganuma, K. (2016) Upstreamness in the global value chain: Manufacturing and services. In Meeting of the

Japanese Economic Association at Nagoya University on June (Vol. 18, p. 19) .

[29] P. Antràs, D. Chor, T. Fally, and R. Hillberry (2012). Measuring the Upstreamness of Production and Trade Flows. American Economic Review: Papers & Proceedings **102**(3): 412–416.

[30] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo (2024). Upstreamness and downstreamness in input-output analysis from local and aggregate information. Preprint arXiv:2009.06350v4.

[31] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo (2023). Ranking influential nodes in networks from aggregate local information. Phys. Rev. Research **5**, 033123.

[32] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo (2021). "Spectrally gapped" random walks on networks: a Mean First Passage Time formula. SciPost Phys. **11**, 088.

[33] M. J. Crumpton, Y. V. Fyodorov, P. Vivo (2022). Statistics of the largest eigenvalues and singular values of low-rank random matrices with non-negative entries. Preprint arXiv:2208.09430.

[34] I. S. Gradshteyn and I. M. Ryzhik (2000). Table of Integrals, Series and Products. Academic Press; 6th edition.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Economic and Social Research Council

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk