



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 Margrave Gardens, London, WC 8RH, United Kingdom

^e Theoretical Division (T4), Condensed Matter & Complex Systems, Los Alamos National Laboratory, Los Alamos, 87545, NM, USA

^f Dept. of Mathematics, King's College London, Strand, London, WC2R 2LS, United Kingdom

ARTICLE INFO

Dataset link: <https://www.rug.nl/ggdc/valuechain/wiod/wiod-2013-release?lang=en>

Keywords:

Upstreamness

Downstreamness

Global value chains

Input/output

Correlations

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 factors of production, and they are computed from 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 (i) extra care is needed when interpreting these measures as simple representations of each sector's positioning along the value chain, as the “curse of the input–output identities” and labor effects effectively force the value chain to acquire additional links from primary factors of production, and (ii) 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, in turn making other proposed measures of sector inter-linkages more suitable and intuitive.

* Corresponding author.

E-mail address: pierpaolo.vivo@kcl.ac.uk (P. Vivo).

<https://doi.org/10.1016/j.jebo.2025.106945>

Received 15 February 2024; Received in revised form 5 February 2025; Accepted 12 February 2025

Available online 5 March 2025

0167-2681/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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 (Leontief, 1986, 1936). This level of analysis is facilitated by the increasing availability and development of detailed Input/Output (I–O) tables for each country (United Nations Department for Economic and Social Affairs Statistics Division, 1999; Timmer et al., 2015).

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 (2012), Miller and Temurshoev (2015) and Fally (2012) introduced the notions of *upstreamness* and *downstreamness* to quantify the position of each economic sectors (and countries as a whole) with respect to final demand, and primary factors of production, respectively (see Section 2 for details).

In a recent paper that has attracted much attention, Antràs and Chor (2018) reported empirical observations of a puzzling correlation existing between the upstreamness and downstreamness of several countries¹ over many years (already noted in Miller and Temurshoev (2015)). More precisely, they used 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 Antràs and Chor (2018)). Similar effects are then shown also at the single country-industry level (see e.g. Fig. 10 in Antràs and Chor (2018)).

Several explanatory factors have been put forward in Antràs and Chor (2018) 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 Wang et al. (2017), and alternative measures put forward there (see also (Colon et al., 2020) for an alternative proposal based on a dynamical model of supply chains rooted in evolutionary game theory). Our analysis lends further support to the warning issued in Wang et al. (2017) concerning the subtleties involved in the traditional measures of upstreamness and downstreamness when interpreted as simple proxies for a sector’s (country’s) mere positioning along the GVC.

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 (Peterson and Olinick, 1982; McNeerney et al., 2021; McNeerney, 2012; Kop Jansen, 1994; Kop Jansen and Ten Raa, 1990; Evans, 1954; Quandt, 1958; Simonovits, 1975; West, 1986; Kogelschatz, 2007; Kozicka, 2019; Katz and Burford, 1985; Burford, 1977; Drake, 1976; Phibbs and Holsman, 1981; Jensen and Hewings, 1985,a; Burford and Katz, 1985b). 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 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 row sub-stochasticity² of a closely related matrix, see Section 2 below), 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.

Investigating further the economic factors that lead to correlations as hard to destroy as difficult to make intuitive sense of, we find that a crucial role is played by the input–output identities expressing the fundamental equilibrium rule that “outputs + final demand” should perfectly balance “inputs + value added” at the sectorial level. We will refer in the following to such accounting identity as *input–output identity*. Such identities not only generate a *de facto* denser graph of interactions between sectors, which in turn tends to bring U and D closer to each other, but they also blur the natural interpretation one would expect to assign to these measures (at least in simple GVC topologies), namely as proxies for the positioning of each sector along the production chain — so that sectors closer to the final demand should have low *U* but simultaneously high *D*, and vice versa for sectors closer to primary factors of production. We will see that this natural intuition may actually fail even in simple examples, a further indication that such measures should collectively be interpreted with much caution.

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 This means that the sum of elements in each row is smaller than or equal to 1.

	Intermediate demand Z_i				
	Sector 1	...	Sector N	Final demand	Output
Sector 1	a_{11}	...	a_{1N}	F_1	Y_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
Sector N	a_{N1}	...	a_{NN}	F_N	Y_N
Value added	V_1	...	V_N		
Output	Y_1	...	Y_N		

Fig. 1. Scheme of the structure of a single-country input–output table (Timmer et al., 2015; United Nations Department for Economic and Social Affairs Statistics Division, 1999; Saganuma, 2016).

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 will analyze in detail the simplest case of a linear production chain to show that the input–output identities (in the form of labor injected in the production process) (i) forces the downstreamness values to deviate from the naïve interpretation of positioning index along the chain starting from primary factors of production, and (ii) together with the non-negativity constraint on I–O matrix elements, it pushes U and D to be more aligned with each other, effectively preventing these measures from pointing in opposite directions (as it would be quite natural to expect) unless the production chain is truly trivialized – a set of consequences that we collectively dub “the curse of the input–output identity”. In Section 4 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 5, 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 6, along with numerical tests for other distributions of the entries of A , all confirming the same conclusions. In Section 7 we perform our random reshuffling experiment on empirical I–O matrices, which demonstrates 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 7 we offer some critical discussion and concluding remarks.

2. Definition of upstreamness and downstreamness

Antràs and Chor (2012) 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)

$$Y_i = F_i + Z_i = F_i + \sum_{j=1}^N a_{ij} Y_j \quad (1)$$

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

Here, a_{ij} is the total value in monetary units (e.g. US dollars) of i 's output used to produce j 's output, while d_{ij} is the amount of monetary units of sector i 's output needed to produce one monetary unit's worth of sector j 's output (see schematic structure of a I–O matrix for a single country in Fig. 1).

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

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

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

$$Y = [\mathbb{1}_N - D]^{-1} F, \quad (4)$$

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

Antràs and Chor (2012) 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

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

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

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

$$U_1 = [\mathbb{1}_N - A_U]^{-1} \mathbf{1}, \quad (6)$$

where

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

and $Y = \text{diag}(Y_1, \dots, Y_N)$. 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. (2012) 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 (2012). Fally's upstreamness U_2 is defined as follows:

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

The idea is that 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

$$U_2 = [\mathbb{1}_N - A_U]^{-1} \mathbf{1}, \quad (9)$$

where A_U is defined in (7). In Antràs and Chor (2018) an application of those measures for the analysis of empirical data on global value chains is presented.

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

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

or in vector/matrix form

$$Y = [\mathbb{1}_N - D^T]^{-1} V. \quad (11)$$

Similarly to Antràs and Chor (2012) (see (5)), Miller and Temurshoev (2015) introduced the so-called *downstreamness* measuring the average distance between suppliers of primary factors of production and sectors as input purchasers, along the input demand chain as follows

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

As before, using (11), we obtain

$$D_1 = [\mathbb{1}_N - A_D]^{-1} \mathbf{1} \quad (13)$$

with

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

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 (2012) introduced an analogous iterative definition of the form

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

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 Antràs and Chor (2018). 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



Fig. 2. Schematic representation of a production chain of five sectors, where each sector receives goods from a single other sector (or primary factors of production, in the case of 5), and transfer it to another one (or to the consumer to meet the Final Demand, in the case of 1).

“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 Antràs and Chor (2018), the two approaches were found to deliver extremely highly correlated country-level indices of GVC positioning.

3. The curse of the input–output identity

In this Section, we illustrate a few subtleties with the interpretation of upstreamness and downstreamness defined earlier as simple indicators of a sector’s positioning along the production process by considering the simplest example of a production (linear) chain, where sectors are arranged as in Fig. 2.

Each sector receives goods as input from a single other sector, while sector 1 sells directly to final consumers, and sector 5 handles primary factors of production directly.

In such a simple geometry, we would expect to be able to read off the upstreamness ($U_1 = (1, 2, 3, 4, 5)^T$) and downstreamness ($D_1 = (5, 4, 3, 2, 1)^T$) values of each sector by simply identifying their positioning along the production chain with respect to final consumers and primary factors of production, respectively. However, we will see shortly that this natural expectation is violated even in this simple geometry due to the input–output identity (i.e. the need to take into account labor injected in the production chain at every step).

The I–O matrix A that corresponds to the chain in Fig. 2 reads

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 & 0 \\ 0 & 0 & a_{43} & 0 & 0 \\ 0 & 0 & 0 & a_{54} & 0 \end{pmatrix}, \quad (16)$$

with corresponding final demand vector $F = (F_1, 0, 0, 0, 0)^T$, with $F_1 \neq 0$, to reflect the fact that the sector 1 does not transfer output to any other sectors, but only to the final consumer, while all other sectors do not interact directly with the final consumer. As a consequence, the matrices A_U and A_D defined in Eqs. (7) and (14) read

$$A_U = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (17)$$

and

$$A_D = \begin{pmatrix} 0 & a_{21}/F_1 & 0 & 0 & 0 \\ 0 & 0 & a_{32}/a_{21} & 0 & 0 \\ 0 & 0 & 0 & a_{43}/a_{32} & 0 \\ 0 & 0 & 0 & 0 & a_{54}/a_{43} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (18)$$

Interestingly, by imposing the chain topology *from the viewpoint of interaction with final consumer* (i.e. fixing the vector of final demands while leaving the value added vector V unconstrained), we trivialize the matrix A_U (killing any parametric dependence on the actual “weights” of the goods transferred along the chain).

We can now compute the upstreamness and downstreamness vectors U_1 and D_1 from the definitions in Eqs. (6) and (13) obtaining

$$U_1 = (1, 2, 3, 4, 5)^T \quad (19)$$

$$D_1 = \begin{pmatrix} 1 + \frac{a_{21}}{F_1} + \frac{a_{32}}{F_1} + \frac{a_{43}}{F_1} + \frac{a_{54}}{F_1} \\ 1 + \frac{a_{32}}{a_{21}} + \frac{a_{43}}{a_{21}} + \frac{a_{54}}{a_{21}} \\ 1 + \frac{a_{43}}{a_{32}} + \frac{a_{54}}{a_{32}} \\ 1 + \frac{a_{54}}{a_{43}} \\ 1 \end{pmatrix}. \quad (20)$$

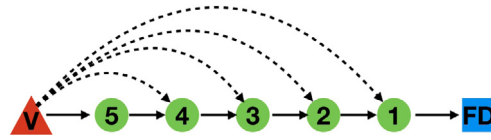


Fig. 3. Schematic representation of the same production chain of five sectors as in Fig. 2, this time including the value added at each node, for instance in the form of labor utilized to process goods.

Perhaps unexpectedly, while the upstreamness vector precisely returns the positioning of each vector along the chain with respect to final consumers, the downstreamness vector has a more complicated and less intelligible structure, which does not match our naïve expectation $D_1 = (5, 4, 3, 2, 1)^T$. Incidentally, this point was mentioned in passing also in the paper (Wang et al., 2017), where the authors stated: “Curiously, sector rankings by “upstreamness” and “downstreamness” measures do not coincide with each other. This implies certain inconsistency in the way that these measures are defined”.

It is interesting to further stress the following:

1. We have so far ignored the input–output identities (1) and (10) that further imply

$$V_i + \sum_{j=1}^N a_{ji} = F_i + \sum_{j=1}^N a_{ij} \quad (21)$$

for every sector i . Specializing to our linear chain, we obtain the following set of identities

$$\begin{aligned} V_1 + a_{21} &= F_1 \\ V_2 + a_{32} &= a_{21} \\ V_3 + a_{43} &= a_{32} \\ V_4 + a_{54} &= a_{43} \\ V_5 &= a_{54}, \end{aligned} \quad (22)$$

which can be visually represented as additional directed links (see Fig. 3).

2. The “natural” downstreamness vector $D_1 = (5, 4, 3, 2, 1)^T$ (representing simply the positioning of sectors along the chain with respect to primary factors of production) could be obtained from (20) only if $a_{21} = a_{32} = a_{43} = a_{54} = F_1$, which would in turn imply $V_1 = V_2 = V_3 = V_4 = 0$ and $V_5 \neq 0$ from the input–output identities above. This is a highly degenerate case where sector 5 handles primary factors of production and then transfer them seamlessly along the chain, with all other sectors playing just the role of “paper pushers” until the unprocessed goods reach the final consumer. In all other cases where sectors actually operate injecting labor on the incoming goods, the downstreamness level of each sector may get distorted compared to their basic positioning, to the point that even the *relative* ordering between sectors (e.g. sector 1 should be more downstream than sector 3) may be subverted. Take for instance $a_{21} = 1.5$, $a_{32} = \alpha a_{21}$, $a_{43} = \alpha^2 a_{21}$, $a_{54} = \alpha^3 a_{21}$, with $\alpha = 1/3$ and $F_1 = 100$. This choice leads to the downstreamness of sector 1 to be *lower* than that of sectors 2, 3, 4, which seems paradoxical at first sight. However, this happens because the input–output identities (22) force V_1 to be ~ 100 times larger than V_2 , and ~ 1000 times larger than V_3 and V_4 : a large external demand, coupled with a relatively low volume of incoming goods to process, forces a strong demand for labor at sector 1, which in turn considerably lowers its expected downstreamness value given its position along the chain (farthest sector from primary factors of production).

In summary, this simple example already shows that U and D do not simply reflect the positioning of a sector (even along a linear production chain), since the input–output identities force an effectively denser network geometry, with extra links corresponding to value added injected at each step (compare Figs. 2 and 3).

Moreover, we can study the U - D Pearson correlation coefficient $\rho \in [-1, 1]$. At least in this simple geometry, we should expect a (maximally) *negative* correlation (close to -1), irrespective of the actual production parameters — as sectors closer to final demand appear to be automatically farther from primary factors of production. Surprisingly, we find that the economic parameters can be chosen to produce *any* value of ρ (including *positive* values!): a sector can therefore be simultaneously quite upstream with respect to final demand, and quite downstream with respect to primary factors of production, unless there is no value added along the chain (the “paper pushers” scenario described above), in which the two measures are perfectly anticorrelated (see below).

To make the calculation more transparent, we put ourselves in a simplified scenario and take a_{21} as reference point, setting $a_{32} = a_{54} = \alpha a_{21}$, $a_{43} = \beta a_{21}$, and $a_{21}/F_1 = t$. The input–output identities (22) force the elementary constraints $0 \leq \alpha, \beta, t \leq 1$. The Pearson correlation coefficient ρ has a lengthy but fully explicit expression as a function of α, β, t , given by $\rho = \Omega_1/\Omega_2$, with

$$\Omega_1 = \alpha - 2\alpha\beta(2t + 1) - \beta(\beta + 2(\beta + 1)t) \quad (23)$$

and

$$\Omega_2 = 2\beta \left[\alpha^2 \left(\frac{2 - 2\beta}{\beta^2} + 8 \right) + \frac{2\beta^2}{\alpha^2} + \alpha \left(8\beta - \frac{1}{\beta} - 3 \right) \right]$$

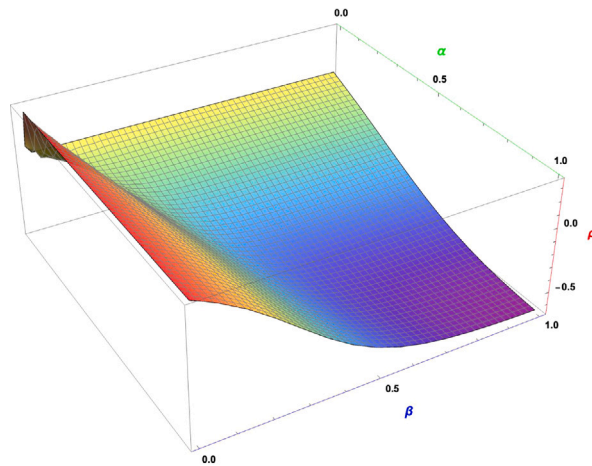


Fig. 4. U-D Pearson correlation coefficient ρ for the simplified scenario described in the main text, with $t = 1/3$, as a function of α and β .

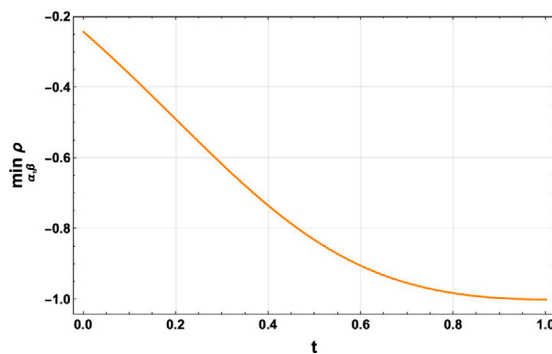


Fig. 5. Minimal value of the U-D Pearson correlation coefficient as a function of t . For t strictly smaller than 1, ρ is lower bounded away from -1 .

$$-\frac{(\beta-4)\beta}{\alpha} + \beta(2\beta-3) + 2t^2(2\alpha+\beta+1)^2 - \frac{t(2\alpha+\beta+1)(\alpha^2(2\beta+1) + \alpha\beta(\beta+1) + \beta^2)}{\alpha\beta} + 1 \Big]^{1/2},$$

which can be plotted and analyzed in detail (see Fig. 4 for $t = 1/3$).

First, we observe that the limit $t, \alpha, \beta \rightarrow 1$ corresponds to sending all added values V_i to zero, reproducing the “paper pusher” scenario described above: in this case, the Pearson correlation coefficient ρ correctly tends to -1 , signaling that in this case U and D are maximally anti-correlated (the closest sector to final demand must be the farthest sector from primary factors of production).

However, a more striking observation is that, for t strictly smaller than 1, the minimum value of ρ (while still negative), is bounded away from -1 (see Fig. 5), which means that having just V_1 strictly larger than zero is already enough to push U and D closer to each other. Moreover, the maximum value of $\rho = 1/2\sqrt{2} \approx 0.3535\ldots$ – obtained for vanishing β , irrespective of α, t – is *positive* (and all intermediate values are possible). So, even in a perfectly linear chain, where the positioning of a sector according to final demand should be perfectly specular to its positioning according to primary factors of production, it is eminently possible to have U and D weakly anti-correlated, uncorrelated, or even positively correlated instead, for instance provided one of the links is sufficiently weak that the value added at the corresponding node must “step up” to improve the flow along the chain.

In summary, the simple “chain” geometry already shows that the natural interpretation of U and D as measures of each sector’s positioning along the production process must be handled with care, and that a positive correlation between the two (or at least a much weaker anti-correlation than one might naïvely expect) may naturally arise because of the effectively denser structure of the production chain once value added (e.g. in the form of labor) is included in the picture due to the input–output identities (22). In this context, it is worth highlighting that this “weakened anti-correlation” was precisely at the heart of a different proposal to measure the relative positioning of a sector within GVC, put forward in Wang et al. (2017). Citing (Wang et al., 2017) again: “We propose a new production position measure as the relative distance of a particular production stage (country-sector) to the both ends³ of a value chain. Using our definitions, the sector ranking by upstreamness and downstreamness would be exactly *inversely related*.”

³ Italics added for emphasis.

This removes one inconsistency with the existing measures in the literature. ”

In the next sections, we will further investigate these correlations in the framework of randomized GVC models.

4. The random model

Our randomized model is based on the closed-economy paradigm described in Section 2, 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 Section 4.2 for the precise definition of the random model.

4.1. Covariance and slope

Assuming therefore that the underlying model for the interaction matrix A is random, the covariance between the upstreamness $(U_1)_i$ and downstreamness $(D_1)_i$ (defined in Eqs. (6) and (13), respectively) of the same i th sector is

$$\text{Cov}((U_1)_i, (D_1)_i) = \mathbb{E}[(U_1)_i(D_1)_i] - \mathbb{E}[(U_1)_i]\mathbb{E}[(D_1)_i], \quad (24)$$

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. (24) 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 Bartolucci et al. (2025, 2023), which demonstrated that the “true” upstreamness and downstreamness (as defined in Eqs. (6) and (13), respectively) for sufficiently dense matrices are individually correlated with simpler rank-1 estimators

$$\tilde{U}_i = 1 + \frac{r_i}{1 - (1/N) \sum_j r_j} \quad (25)$$

$$\tilde{D}_i = 1 + \frac{r'_i}{1 - (1/N) \sum_j r'_j}, \quad (26)$$

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

$$\text{Cov}(\tilde{U}_i, \tilde{D}_i) = \mathbb{E}[\tilde{U}_i \tilde{D}_i] - \mathbb{E}[\tilde{U}_i]\mathbb{E}[\tilde{D}_i] \quad (27)$$

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. (27).⁴ Therefore, our covariance of interest reduces to the following object

$$C_N = \frac{\mathbb{E}[rr'] - \mathbb{E}[r]\mathbb{E}[r']}{(1 - \mathbb{E}[r])(1 - \mathbb{E}[r'])}, \quad (28)$$

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. 6 and 7 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.

The slope S of the scatter plot between $\tilde{D}_i$ and $\tilde{U}_i$ is easily determined from Eq. (27) 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. (27) we get

$$\text{Cov}(\tilde{U}_i, \tilde{D}_i) = S \{ \mathbb{E}[\tilde{U}_i^2] - \mathbb{E}[\tilde{U}_i]^2 \}, \quad (29)$$

from which we deduce

$$S = \frac{\text{Cov}(\tilde{U}_i, \tilde{D}_i)}{\text{Var}[\tilde{U}_i]}, \quad (30)$$

where $\text{Var}[\tilde{U}_i] = \mathbb{E}[\tilde{U}_i^2] - \mathbb{E}[\tilde{U}_i]^2$ 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 LLN, and after simple algebra from Eq. (27), we have that the slope S can be approximated by

$$S = \frac{C_N (1 - \mathbb{E}[r])^2}{\mathbb{E}[r^2] - \mathbb{E}[r]^2}. \quad (31)$$

⁴ 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' .

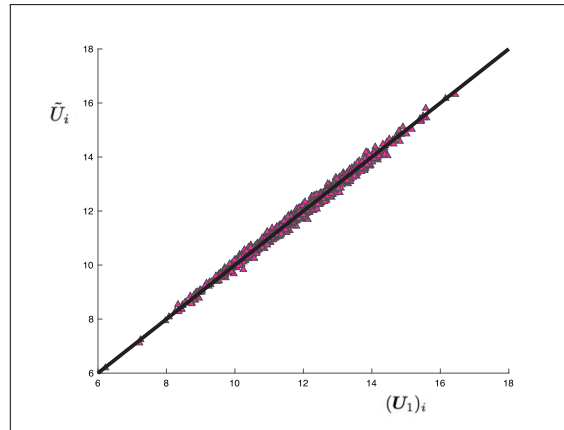


Fig. 6. Scatter plot of the approximate upstreamness (Eq. (25)) vs. the “true” upstreamness $(U_1)_i$ for our random model with exponential disorder. The parameters used are $\mu = 1$, $\mu_F = 0.1$, $N = 100$, $i = 7$. There are 1000 pairs of points in the figure, each obtained from a different instance of the random matrix A with exponential disorder.

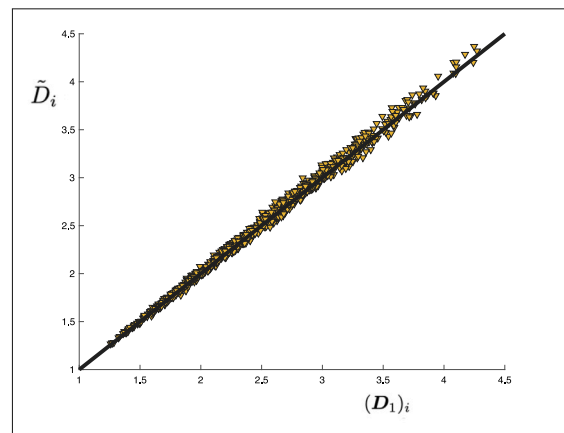


Fig. 7. Scatter plot of the approximate downstreamness (Eq. (26)) vs. the “true” downstreamness $(D_1)_i$ for our random model with exponential disorder. The parameters used are $\mu = 1$, $\mu_F = 0.01$, $N = 100$, $i = 7$. There are 1000 pairs of points in the figure, each obtained from a different instance of the random matrix A with exponential disorder.

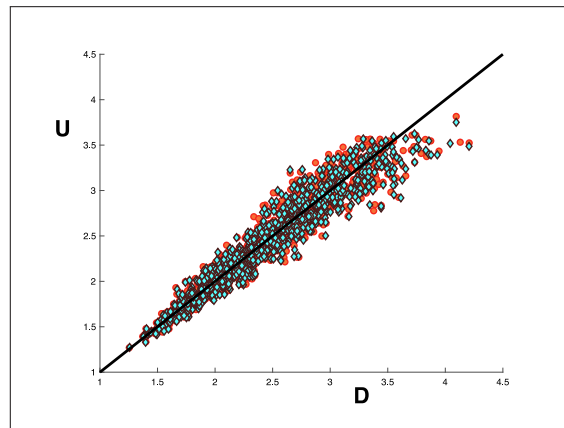


Fig. 8. Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 200$ random model with exponential disorder with parameters $\mu = 1$, $\mu_F = 0.005$. Each of the 1000 blue [red] points is obtained from one instance of the random “source” matrix A , and represents the pair of values $((U_1)_i, (D_1)_i)$ [$(\tilde{U}_i, \tilde{D}_i)$], respectively]. The thick black line has slope +1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Model definition

In the random model we consider that the entries of A are drawn independently from an exponential 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 in the construction of A . The final demand values F_i ($i = 1, \dots, N$) are further modeled as i.i.d. exponential random variables with mean $1/\mu_F$. The choice of an exponential pdf for the entries of the random model instead of a fatter distribution, closer to the empirical one (see Section 7 below) makes the analytical calculations simpler and the interpretation of results more transparent. However, we tested numerically that the results remain robust even for other distributions (for which we do not provide an analytical treatment), therefore we can safely conclude that the choice of an exponential pdf is not in any way unduly restrictive or affecting the trustworthiness of our conclusions.

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

$$(A_U)_{ij} = \frac{a_{ij}}{\sum_j a_{ij} + F_i} \quad (32)$$

$$(A_D)_{ij} = \frac{a_{ji}}{\sum_j a_{ij} + F_i}, \quad (33)$$

where we used that $\sum_j a_{ij} + F_i = \sum_j a_{ji} + V_i$ for all i , as follows from the input–output identities. Therefore, both A_U and A_D as defined above have non-negative elements, and are row sub-stochastic (as they should), provided that μ_F is sufficiently small. This condition $\mu_F \ll \mu$ is necessary to ensure that F_i 's will be (typically) large enough to make $\sum_j (A_D)_{ij}$ in (33) smaller than 1.

For each instance of the random matrix A and of the vector of final demands F , we construct the matrices A_U and A_D as above, and from those we compute the pairs $(U_1)_i, (D_1)_i$ and $\bar{U}_i, \bar{D}_i$ for any sector i that we choose. These are all random variables, whose pairwise covariance is of interest in this paper.

5. 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. 9 and 10 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, unless the matrix is exceedingly sparse and/or it represents a very *ad hoc* structure (e.g. the linear chain of Section 3), which of course occur with zero likelihood in the random model.

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

$$(A_U)_{ij} = \frac{a_{ij}}{\sum_j a_{ij} + F_i} \quad (34)$$

$$(A_D)_{ij} = \frac{a_{ji}}{\sum_j a_{ij} + F_i}, \quad (35)$$

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 downstreamness (see Eq. (28))

$$C_N(\mu, \mu_F) = \frac{\mathbb{E}[rr'] - \mathbb{E}[r]\mathbb{E}[r']}{(1 - \mathbb{E}[r])(1 - \mathbb{E}[r'])} \quad (36)$$

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

$$\begin{aligned} \mathcal{N}_N(\phi) = & (\phi - 1) \left[N(\phi - 1)B(1, N + 1) {}_2F_1(1, N + 1; N + 2; \phi)^2 + (N - 1)N(\phi - 1) \right. \\ & \left. \times B(1, N + 1) {}_2F_1(1, N + 1; N + 2; \phi)B(1, N) {}_2F_1(1, N; N + 1; \phi) + (N + 1)(\phi - 1)B(1, N + 2) {}_2F_1(1, N + 2; N + 3; \phi) + N \right] \end{aligned}$$

$$\begin{aligned} \mathcal{D}_N(\phi) = & ((N - 1)(\phi - 1)B(1, N) {}_2F_1(1, N; N + 1; \phi) + (\phi - 1)B(1, N + 1) {}_2F_1(1, N + 1; N + 2; \phi) + 1) \\ & \times (N(\phi - 1)B(1, N + 1) {}_2F_1(1, N + 1; N + 2; \phi) + 1), \end{aligned}$$

where $\phi = 1 - \mu_F/\mu$. Here, $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ 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. (31)) is equal to +1 for any N , irrespective of the values of μ, μ_F .

The proofs are deferred to the Appendix.

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

μ	μ_F	N	$\text{Cov}((\mathbf{U}_1)_i, (\mathbf{D}_1)_i)$	$C_N(\mu, \mu_F)$
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

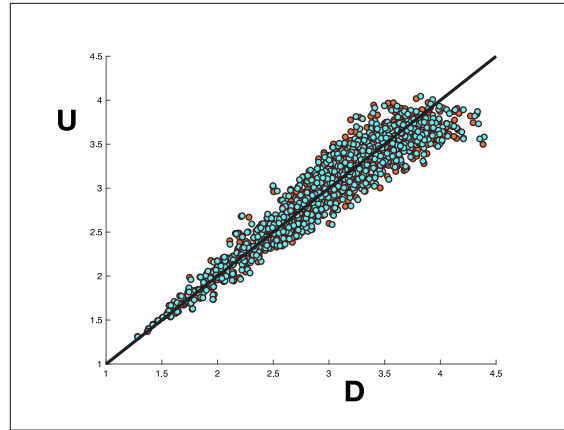


Fig. 9. Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 400$ random model with log-normal disorder with parameters $\mu' = 1$, $\mu'_F = 6.67$. Each of the 1000 light blue [orange] points is obtained from one instance of the random “source” matrix A , and represents the pair of values $((\mathbf{U}_1)_i, (\mathbf{D}_1)_i)$ [$(\tilde{U}_i, \tilde{D}_i)$, respectively]. The thick black line has slope $+1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. 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 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.⁵

From the matrices A_U and A_D so generated, we constructed the vectors of upstreamness $\mathbf{U}_1$ and downstreamness $\mathbf{D}_1$ 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. (25) and (26) respectively.

We first show in Figs. 6 and 7 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 downstreamness of sector $i = 7$, obtained from averaging over $m = 10000$ numerically generated instances of our random model, against the values of $C_N(\mu, \mu_F)$ analytically computed, and we observe an excellent agreement between the two.

In Fig. 8 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.

Moreover, in Figs. 9 and 10 we provide the same scatter plots as in Fig. 8, but this time for the “original” matrix A (and similarly for the vector 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.

In other experiments, we have further tested the effect of (random) “sparsification” of the matrix A , where elements of the random matrix A were set to zero with a certain probability $1 - p$. Computing the Pearson correlation coefficient between U and D

⁵ For A_U substochasticity is guaranteed by construction. For A_D this is true with overwhelming likelihood provided that $\mu_F \ll \mu$.

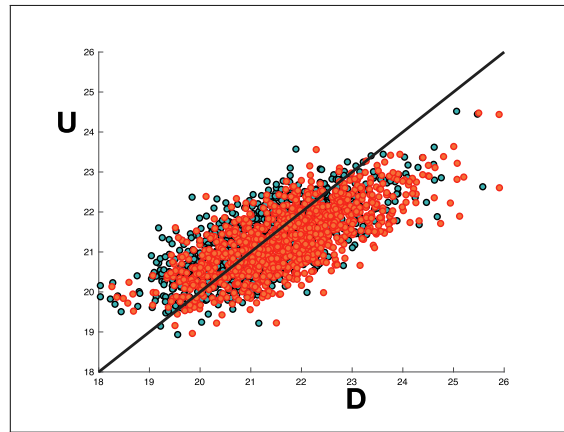


Fig. 10. Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the $N = 400$ random model with uniform disorder with parameters $\mu = 1$, $\mu_F = 0.05$. Each of the 1000 light blue [orange] points is obtained from one instance of the random “source” matrix A , and represents the pair of values $((U_1)_i, (D_1)_i)$ [$(\tilde{U}_i, \tilde{D}_i)$, respectively]. The thick black line has slope $+1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

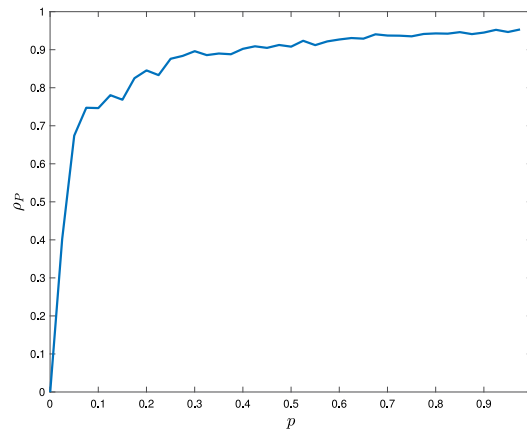


Fig. 11. For the “sparsified” exponential model, Pearson correlation coefficient between U and D of the same sector as a function of the sparsity parameter p (matrix A gets sparser as p gets lower).

of the same sector as a function of the sparsity parameter p for the “sparsified” exponential model described above, we find that the correlation between U and D survives at $\sim 90\%$ level – unless the sparsity increases to unrealistic values ($p \sim 0.2$) – and remains well above $\sim 60\%$ for a sparsity level as high as $p \sim 0.05$, where the matrix A is essentially entirely dominated by zeros and it is not even guaranteed that the corresponding GVC remains connected (see Fig. 11 below). This finding is consistent with the “chain” geometry of Section 3: decreasing the sparsity of the chain, up to the extreme case of a fully connected economy, has generally the effect of increasing the U-D correlation.

Our analyses so far have focused on the properties of an individual economy. To test whether our results hold more generally, we extended our model to consider the simple case of two interacting countries, represented by a 2×2 block matrix. The diagonal blocks represent interactions within sectors of the same economy, while the off-diagonal blocks represent interactions between sectors across the two countries.

We assume the diagonal blocks are fully populated, while in the off-diagonal blocks, each link is present with probability p , similar to the sparse case considered earlier. The two extreme cases of $p = 0$ and $p = 1$ correspond to two isolated economies and a ‘fully globalized’ economy, respectively. The results we discussed above apply to these extreme cases: when $p = 1$, the system behaves as a single, larger economy, whereas for $p = 0$, each country can be analyzed individually.

By tuning p , we can observe the effect of adjusting the level of integration between the economies, for instance, through trade barriers, as discussed in Antràs and Chor (2018). As shown in Fig. 12, where we present scatter plots of downstreamness and upstreamness for different values of p , we find that the model’s qualitative behavior and the ensuing U-D correlations remain largely independent of p .

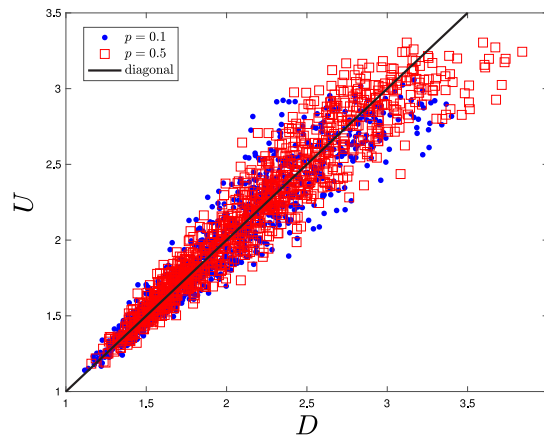


Fig. 12. Scatter plot between upstreamness and downstreamness of sector $i = 7$ for the 2×2 block model. The aggregate matrix is of size $N = 200$, with each block of size 100×100 . Blue dots refer to $p = 0.1$, while red squares to $p = 0.5$ for the off-diagonal blocks. The diagonal blocks are fully populated, and the non-zero entries of the matrix are drawn from an exponential distribution with parameter $\mu = 1$, while $\mu_F = 0.05$. The solid black line has slope $+1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. 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). First, in Fig. 13 we plot the empirical pdf of the entries of I–O matrices for three representative countries (Poland, France, and U.S.A.) across all available years. The empirical distributions are generally well-fitted by power laws with exponential cutoffs ($Y = aX^b \exp(-X/c)$), with best fitting parameters included inside the plot boxes. The essential power-law (fatter than exponential) nature of the entry distribution is partly due to a genuine “multi-scale” nature of the typical interactions among sectors, which span several orders of magnitude, but it is also partly affected by having to lump together matrices corresponding to different years to obtain a sufficiently meaningful statistics.

Next, for each country and each year, we have computed the upstreamness and downstreamness of that country using Eqs. (9) and (13) respectively, averaging over sectors.

In Fig. 14, we show the values of upstreamness for all countries in all years, spanning a period identical to that considered in the paper (Antràs and Chor, 2018). In Fig. 15, 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 Antràs and Chor (2018), namely that the two measures appear to be strongly correlated with a slope of the scatter plot very close to $+1$.

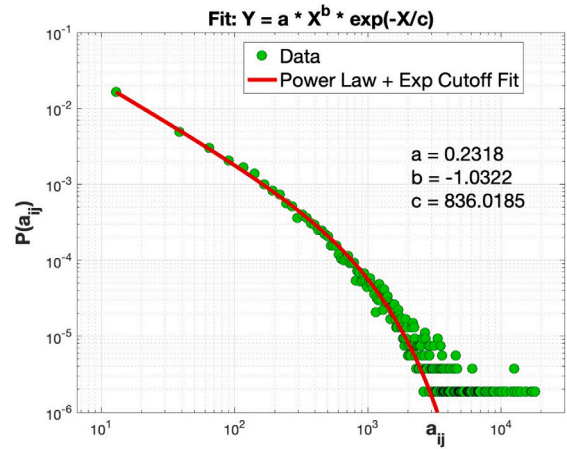
In Fig. 16, though, we took the same I–O matrices for the entire period 1995–2011 and we randomly reshuffled the columns of such matrices according to a random permutation of the set $\{1, \dots, N\}$. The resulting matrices satisfy the same aggregate constraints (namely, the row sums) of the original, 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 .

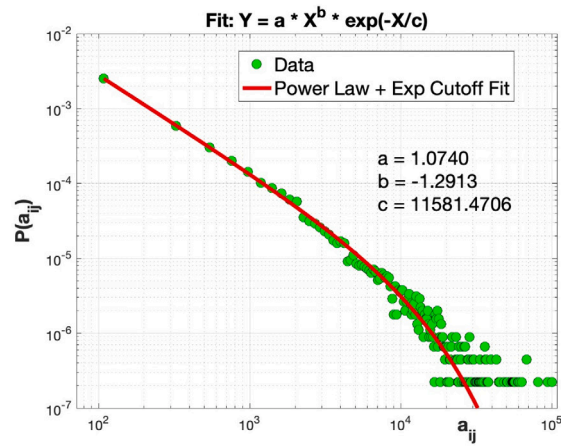
While our work provides an intuition on the origin of correlations between the measures of upstreamness and downstreamness introduced in the literature, the challenge of how to devise a better measure of downstreamness that (i) is not trivially correlated to the upstreamness, and (ii) complies with the more intuitive notion of measuring the positioning of a sector along the production chain has been thoroughly addressed in Wang et al. (2017), to which we refer for details (see also (Branger et al., 2023)).

8. Discussion

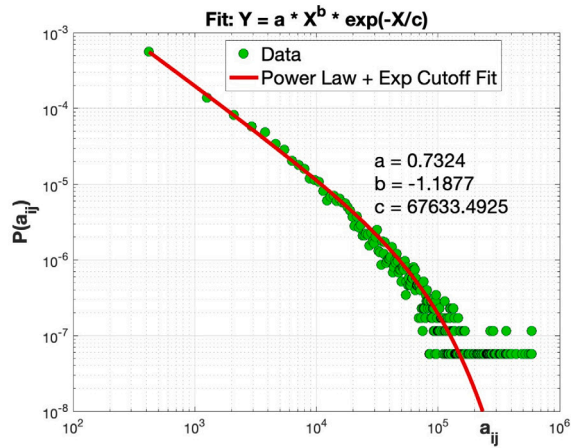
In summary, we have considered two classes of *random* Input/Output matrices A to test whether the “puzzling” correlation detected between upstreamness and downstreamness at the sector and country level (Antràs and Chor, 2018) would survive even if the underlying economic forces and the inter-sectorial dependencies had nothing to do with the real world.



(a) Poland



(b) France



(c) U.S.A.

Fig. 13. Probability density function in log-log scale of the entries of WIOD matrices for three countries, obtained by histogramming all the entries of the I–O matrices across the available 16-years span. In red, the line of best fit with a power-law with exponential cutoff. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

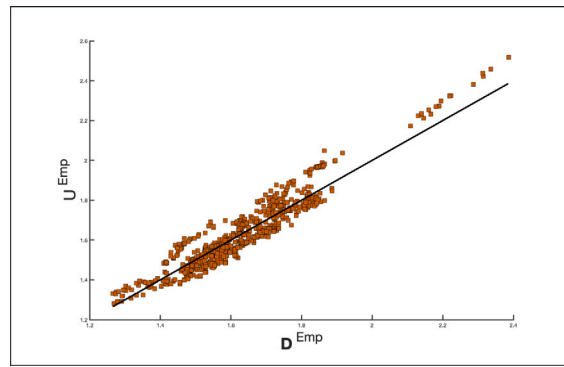


Fig. 14. Empirical upstreamness vs. empirical downstreamness (averaged over 35 sectors) for 39 countries for the years 1995–2001.

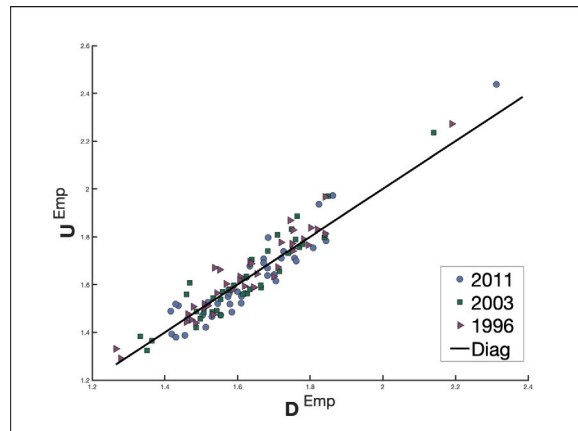


Fig. 15. Empirical upstreamness vs. empirical downstreamness (averaged over 35 sectors) for 39 countries in selected years (1996–2003–2011).

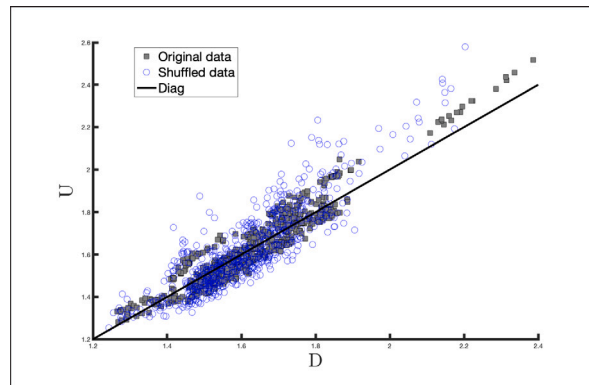


Fig. 16. Upstreamness vs. downstreamness (averaged over 35 sectors) calculated on the empirical and reshuffled matrices for 39 countries for the years 1995–2001. Blue circles represent values calculated on I–O matrices where columns have been randomly reshuffled. Grey squares are upstreamness/downstreamness pairs calculated on the original data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

First, we constructed a random model for the matrix A that mimics a closed economy composed of N economic sectors. We showed analytically that 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 Antràs and Chor (2018). 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 downstreamness 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.

By further analyzing a simplified geometry for the economy (i.e. a linear chain) in Section 3, we identified further subtleties with the standard definitions of upstreamness and downstreamness, which are jointly unable to support – even in such a simple geometry – the intuitive interpretation in terms of positioning of a sector along the chain with respect to final demand and primary factors of production, respectively. We were able to trace such less intuitive features of the U–D measures to the input–output identities that any I–O matrix must satisfy. This “curse of the input–output identities” effectively introduces additional links into the I–O table as extra value must be added at each node, which in turn leads to uncontrolled effects on the numerical values taken by the U and D indicator, blurring their natural economic interpretations and pushing one indicator to essentially align with the other (with the exclusion of special cases constructed *ad hoc*, e.g. unphysical levels of sparsity, or vanishing value added at each node).

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 Antràs and Chor (2018) and the puzzling correlation highlighted thereof for the following 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) 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.

In fact, we carried out simulations for the case when the random matrix A has a 2×2 block structure to mimic the case of two interacting countries. We considered fully dense diagonal blocks to model interactions of the sectors within each country, and sparse off-diagonal blocks to model interactions between countries. In this simple model, the sparsity of off-diagonal blocks can be changed to study the effect of different levels of integration between the countries. We found that the positive correlation between upstreamness and downstreamness we documented for the close economy persists for this more structured economy, pretty much irrespective of the level of integration among the two countries of our simplified setting.

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 Bartolucci et al. (2025, 2023). This assumption was tested on empirical I–O data in Bartolucci et al. (2025, 2023), and on the random model here in Figs. 6 and 7, 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 Bartolucci et al. (2023, 2021), Crumpton et al. (2022)), a situation that does not often materialize in practice. By considering national I–O tables available from the 2013 release of the WIOD (Timmer et al., 2015), 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. (24) 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 (Bartolucci et al., 2025).

Financial disclosure

The work of F. Caravelli was carried out under the auspices of the NNSA of the U.S. DoE at LANL under Contract No. DE-AC52-06NA25396. F. Caravelli was also financed via DOE LDRD grant 20240245ER. P.V. acknowledges support from UKRI Future Leaders Fellowship Scheme (No. MR/X023028/1).

Declaration of competing interest

The authors declare no competing interests.

Appendix. Derivation of Theorem 1 and Corollary 1

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

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

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

$$\frac{1}{\xi} = \int_0^\infty ds e^{-\xi s} \quad \xi > 0 \quad (\text{A.2})$$

we have

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

where

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

from Gradshteyn and Ryzhik (2000), formula 3.197.5 (pag. 335) with $\lambda = 1$, $\nu = -1$, $\alpha = \mu_F/\mu$ and $\tilde{\mu} = -k$. Here, $B(\cdot, \cdot)$ is the Beta function, and ${}_2F_1$ is a hypergeometric function.

Similarly

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

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

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

where

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

and

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

Finally

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

where

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

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

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

where

$$\begin{aligned}K_1 &= \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N ds dt e^{-\mu \sum_k x_k - \mu \sum_{k \geq 2} y_k - \mu_F F_1} e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} x_i^2 = \\ &= \frac{1}{\mu^{N-1}} \int ds dt \frac{1}{\mu_F + s + t} \left[\int dx e^{-\mu x - (s+t)x} \right]^{N-1} \int dx x^2 e^{-\mu x - (s+t)x} \\ &= \frac{2}{\mu^{N-1}} \int_0^\infty ds dt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+2}} = \\ &= \boxed{\frac{2}{\mu^{N-1}} L(N+2)}.\end{aligned}\quad (\text{A.11})$$

$$\begin{aligned}K_2 &= \int dx_1 \cdots dx_N dF_1 dy_2 \cdots dy_N ds dt e^{-\mu \sum_k x_k - \mu \sum_{k \geq 2} y_k - \mu_F F_1} x_i \sum_{k \geq 2} y_k e^{-s(\sum_k x_k + F_1) - t(\sum_k x_k + F_1)} = \\ &= (N-1) \int ds dt \frac{1}{\mu_F + s + t} \left[\int dx e^{-\mu x - (s+t)x} \right]^{N-1} \int dx x e^{-\mu x - (s+t)x} \left[\int dy e^{-\mu y} \right]^{N-2} \int dy y e^{-\mu y} \\ &= \frac{N-1}{\mu^N} \int_0^\infty ds dt \frac{1}{\mu_F + s + t} \frac{1}{(\mu + s + t)^{N+1}} = \\ &= \boxed{\frac{N-1}{\mu^N} L(N+1)}.\end{aligned}\quad (\text{A.12})$$

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

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

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.17](#)).

To prove [Corollary 1](#), we need to further compute $\mathbb{E}[r^2]$ and then simplify the resulting expression for the slope (31), yielding a slope = +1 for any N .

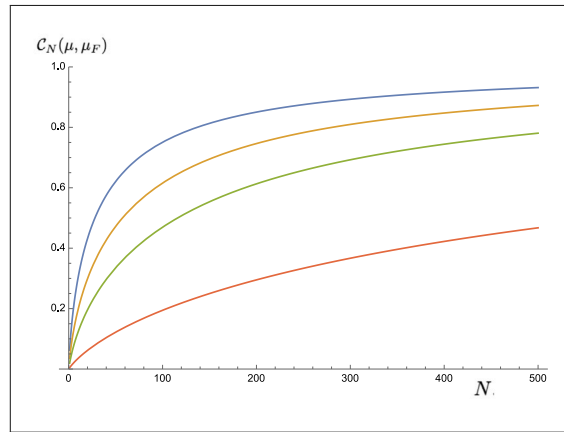


Fig. A.17. Covariance $C_N(\mu, \mu_F)$ between the approximate upstreamness and downstreamness for the random model (exact formula in Theorem 1). The parameters (μ, μ_F) are (1, 0.1) (blue), (2, 0.1) (orange), (2, 0.05) (green), (2, 0.01) (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\begin{aligned}
 \mathbb{E}[r^2] &= \int_0^\infty \prod_{i=1}^N da_i p(a_i) \int_0^\infty dF p_F(F) \left[\frac{\sum_k a_k}{\sum_k a_k + F} \right]^2 \\
 &= \mu^N \mu_F \int d\mathbf{a} dF ds \, se^{-\mu \sum_k a_k - \mu_F F} \left(\sum_k a_k \right)^2 e^{-s(\sum_k a_k + F)} \\
 &= \mu^N \mu_F \left[N \int d\mathbf{a} dF ds \, se^{-\mu \sum_k a_k - \mu_F F} a_1^2 e^{-s(\sum_k a_k + F)} + (N^2 - N) \int d\mathbf{a} dF ds \, se^{-\mu \sum_k a_k - \mu_F F} a_1 a_2 e^{-s(\sum_k a_k + F)} \right] \\
 &= \mu^N \mu_F N \int_0^\infty ds \, s \left[\int dx e^{-\mu x - sx} \right]^{N-1} \int dy \, y^2 e^{-\mu y - sy} \int dF e^{-\mu_F F - sF} + \mu^N \mu_F (N^2 - N) \int_0^\infty ds \, s \times \\
 &\quad \times \left[\int dx e^{-\mu x - sx} \right]^{N-2} \left[\int dy \, y e^{-\mu y - sy} \right]^2 \int dF e^{-\mu_F F - sF} \\
 &= 2\mu^N \mu_F N \int_0^\infty ds \frac{s}{(\mu + s)^{N+2}} \frac{1}{\mu_F + s} + \mu^N \mu_F (N^2 - N) \int_0^\infty ds \frac{s}{(\mu + s)^{N+2}} \frac{1}{\mu_F + s} \\
 &= \boxed{\mu^N \mu_F (N^2 + N) L(N + 2)} \blacksquare
 \end{aligned}$$

Data availability

The datasets analysed in this study are publicly available at <https://www.rug.nl/ggdc/valuechain/wiod/wiod-2013-release?lang=en>. The codes written for the analysis will be made available upon request to the corresponding author.

References

- Antràs, P., Chor, D., 2012. Organizing the global value chain. *Econometrica* 81 (6), 2127–2204.
- Antràs, P., Chor, D., 2018. On the Measurement of Upstreamness and Downstreamness in Global Value Chains. Working Paper 24185, <http://www.nber.org/papers/w24185>.
- Antràs, P., Chor, D., Fally, T., Hillberry, R., 2012. Measuring the upstreamness of production and trade flows. *Am. Econ. Rev.: Pap. Proc.* 102 (3), 412–416.
- Bartolucci, S., Caccioli, F., Caravelli, F., Vivo, P., 2021. Spectrally gapped random walks on networks: a mean first passage time formula. *SciPost Phys.* 11 (088).
- Bartolucci, S., Caccioli, F., Caravelli, F., Vivo, P., 2023. Ranking influential nodes in networks from aggregate local information. *Phys. Rev. Res.* 5, 033123.
- Bartolucci, S., Caccioli, F., Caravelli, F., Vivo, P., 2025. Upstreamness and downstreamness in input-output analysis from local and aggregate information. *Sci. Rep.* 15 (1), 2727.
- Branger, N., Flacke, R.M., Meyerhof, P., Windmüller, S., 2023. Stock returns in global value chains: The role of upstreamness and downstreamness. *J. Empir. Financ.* 74, 101437.
- Burford, R.L., 1977. Regional input-output multipliers without a full IO table. *Ann. Reg. Sci.* 11 (3), 21–38.
- Burford, R.L., Katz, J.L., 1985b. Shortcut ‘input-output’ multipliers, alive and well: Response to Jensen and Hewings. *Environ. Plan. A* 17 (11), 1541–1549.
- Colon, C., et al., 2020. Fragmentation of production amplifies systemic risks from extreme events in supply-chain networks. *PLoS One* 15 (12), e0244196.
- Crumpton, M.J., Fyodorov, Y.V., Vivo, P., 2022. Statistics of the largest eigenvalues and singular values of low-rank random matrices with non-negative entries. Preprint [arXiv:2208.09430](https://arxiv.org/abs/2208.09430).
- Drake, R.L., 1976. A short-cut to estimates of regional input-output multipliers: methodology and evaluation. *Int. Reg. Sci. Rev.* 1 (2), 1–17.

- Evans, W.D., 1954. The effect of structural matrix errors on interindustry relations estimates. *Econometrica* 22 (4), 461–480.
- Fally, T., 2012. Production staging: Measurement and facts (unpublished).
- Gradshteyn, I.S., Ryzhik, I.M., 2000. Table of Integrals, Series and Products, sixth ed. Academic Press.
- Jensen, R.C., Hewings, G.J.D., 1985. Shortcut 'input-output' multipliers: The resurrection problem (a reply). *Environ. Plan. A* 17 (11), 1551–1552.
- Jensen, R.C., Hewings, G.J.D., 1985a. Shortcut 'input-output' multipliers: A requiem. *Environ. Plan. A* 17 (6), 747–759.
- Katz, J.L., Burford, R.L., 1985. Shortcut formulas for output, income and employment multipliers. *Ann. Reg. Sci.* 19 (2), 61–76.
- Kogelschatz, H., 2007. On the Solution of Stochastic Input–Output-Models. Discussion Paper Series n. 447, University of Heidelberg.
- Kop Jansen, P.S.M., 1994. Analysis of multipliers in stochastic input–output models. *Reg. Sci. Urban Econ.* 24, 55–74.
- Kop Jansen, P., Ten Raa, T., 1990. The choice of model in the construction of Input-Output coefficients matrices. *Internat. Econom. Rev.* 31 (1), 213–227.
- Kozicka, M., 2019. Novel approach to stochastic Input-Output modeling. *RAIRO- Oper. Res.* 53, 1155–1169, <https://doi.org/10.1051/ro/2018046>.
- Leontief, W., 1936. Quantitative input-output relations in the economic system of the United States. *Rev. Econ. Stat.* 18, 105–125.
- Leontief, W., 1986. Input–Output Economics. Oxford University Press.
- McNerney, J., 2012. Applications of Statistical Physics To Technology Price Evolution (Ph.D. thesis). Boston University.
- McNerney, J., Savoie, C., Caravelli, F., Farmer, J.D., 2021. How production networks amplify economic growth. *Proc. Natl. Acad. Sci.* 119 (1), e2106031118.
- Miller, R.E., Temurshoev, U., 2015. Output upstreamness and input downstreamness of industries/countries in world production. *Int. Reg. Sci. Rev.* 40 (5), 443–475.
- Peterson, B., Olinick, M., 1982. Leontief models, Markov chains, substochastic matrices, and positive solutions of matrix equations. *Math. Model.* 3, 221–239.
- Phibbs, P.J., Holsman, A.J., 1981. An evaluation of the Burford Katz short cut technique for deriving input–output multipliers. *Ann. Reg. Sci.* 15 (3), 11–19.
- Quandt, R.E., 1958. Probabilistic errors in the Leontief system. *Nav. Res. Logist. Q.* 5, 155–170.
- Simonovits, A., 1975. A note on the underestimation and overestimation of the Leontief inverse. *Econometrica* 43, 493–498.
- 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.
- Timmer, M.P., Dietzenbacher, E., Los, B., Stehrer, R., de Vries, G.J., 2015. An illustrated user guide to the world input–output database: the case of global automotive production. *Rev. Int. Econ.* 23, 575–605.
- United Nations Department for Economic and Social Affairs Statistics Division, 1999. Handbook of Input-Output Table Compilation and Analysis.
- Wang, Z., Wei, S.J., Yu, X., Zhu, K., 2017. Characterizing Global Value Chains: Production Length and Upstreamness. Working Paper No. w23261, National Bureau of Economic Research.
- West, G.R., 1986. A stochastic analysis of an input–output model. *Econometrica: J. Econ. Soc.* 54, 363–374.