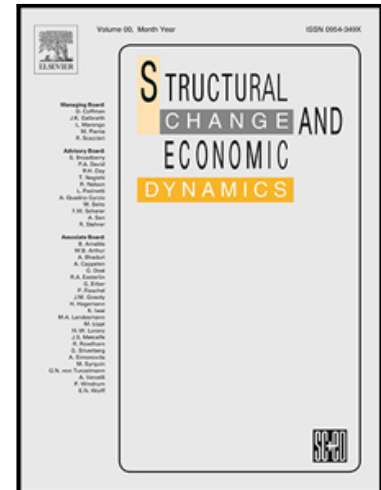


Journal Pre-proof

Using a Linear Regression Approach to Sequential Interindustry Model for Time-Lagged Economic Impact Analysis

Kehan He , Zhifu Mi , D'Maris Coffman , Dabo Guan

PII: S0954-349X(22)00052-2
DOI: <https://doi.org/10.1016/j.strueco.2022.03.017>
Reference: STRECO 1190



To appear in: *Structural Change and Economic Dynamics*

Received date: 5 October 2021
Revised date: 25 February 2022
Accepted date: 27 March 2022

Please cite this article as: Kehan He , Zhifu Mi , D'Maris Coffman , Dabo Guan , Using a Linear Regression Approach to Sequential Interindustry Model for Time-Lagged Economic Impact Analysis, *Structural Change and Economic Dynamics* (2022), doi: <https://doi.org/10.1016/j.strueco.2022.03.017>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier B.V.

Highlights

- We propose an innovative algorithm for a derivative of input-output model towards chronologically dynamic analysis.
- Simulated calculations are used to demonstrate the feasibility of the proposed algorithm.
- The innovation provides a powerful method to conduct economic structural analysis using data science tools.

Journal Pre-proof

Using a Linear Regression Approach to Sequential Interindustry Model for Time-Lagged Economic Impact Analysis

Kehan He¹ 0000-0002-8553-0048, Zhifu Mi^{1,*} z.mi@ucl.ac.uk 0000-0001-8106-0694, D'Maris Coffman¹ 0000-0003-3792-4744, Dabo Guan^{1,2} 0000-0003-3773-3403

¹The Bartlett School of Sustainable Construction, University College London, London, WC1E 7HB, UK

²Department of Earth System Science, Tsinghua University, Beijing, 10084, China

*Corresponding Authors:

Abstract

The input-output (IO) model is a powerful economic tool with many extended applications. However, one of the widely criticized drawbacks is its rather lengthy time lag in data preparation, making it impossible to apply IO in high-resolution time-series analysis. The conventional IO model is thus unfortunately unsuited for time-series analysis. In this study, we present an innovative algorithm that integrates linear regression techniques into a derivative of the IO method, the Sequential Interindustry Model (SIM), to overcome the inherent shortcomings of statistical lags in conventional IO studies. The regressed relationship can thus be used to predict, in the short term, the accumulated chronological impacts induced by fluctuations in sectorial economic demands under disequilibrium conditions. A simulated calculation is presented to serve as an illustration and verification of the new method. In the future, this application can be extended beyond economic studies to broader problems of system analysis.

Keywords: Input-Output analysis; Sequential Interindustry Model; Economic System Modelling; Time Series; Impact Analysis

Word count: 4141 words

Introduction

The input-output (IO) model is a powerful tool that was first proposed by Wassily Leontief in the 1930s (Leontief, 1936). It builds a bridge between the understanding of micro-level production and macro-level economic flows, thus providing policy makers with a tool to account for the

operation of past economies and devise more targeted macro-economic strategies. The concept of the IO model is a consolidation of the classical economic conceptualisation of the *tableau économique* proposed by François Quesnay in 1758 to depict the relationship between the agricultural/mining sector and the artisanal/manufacturing sector in eighteenth-century France (Leontief, 1953, Coffman, 2021). Since World War II, the statistical bureaus of world governments have taken initiatives to construct national IO tables to facilitate policymaking. Recent efforts by international scholars include constructing multiregional IO tables and ancillary emission inventories (Stadler et al., 2018, Timmer et al., 2015, Lenzen et al., 2013, Mi et al., 2017). Many recent studies also extend economic IO analysis to environmental stress in accounting for embodied carbon emissions at the macro level (Dietzenbacher et al., 2020, Hertwich and Peters, 2009, Weber et al., 2008). Other scholars have suggested using input-output based hybrid life cycle assessment models in building and infrastructure projects potentially employing Building Information Modelling (BIM) data (Sharrard et al., 2008, Coffman and Kelsey, 2019).

In addition to the efforts in extending IO applications, some efforts are also devoted in methodological improvement of the IO modelling. To encompass micro-level perspectives, researchers naturally consider the time domain as an avenue for the methodological improvement of the IO model. In fact, Leontief himself delved into the realm of the time domain of IO by introducing the dynamic IO model (Leontief, 1970, Leontief, 1953), attempting to explain the interaction between capital investment and production output on a chronological basis. Even though some researchers are still working towards the perfection and application of the dynamic IO model (Aulin-Ahmavaara, 2000, Rocco, 2019) to better factor in the impact of capital formation on productivity as a solution for economic prediction, such an approach has been criticized by other researchers for a number of reasons (Kurz and Salvadori, 2000). Besides the classic demand driven IO model, attempts have also been made to construct a supply-driven IO model in the 1980s (Oosterhaven, 1988), but few developments have been fulfilled until its recent emphasis in disaster event analysis (Galbusera and Giannopoulos, 2018, Yagi et al., 2020), reigniting the discussion on supply-driven IO (Reyes and Mendoza, 2021).

Under a similar intention, a number of other analysis tools build on the idea of chronological impact in IO model, such as dynamic inoperability (DIIM), supply bottlenecking (ARIO), and hybridization with Computational General Equilibrium (CGE) model, to analyse the impact of disasters in the long term (Mendoza-Tinoco et al., 2017, Zeng et al., 2019, Zeng and Guan, 2020, Guan et al., 2020). Such developments normally include non-linear characteristics, assuming a final steady state will be reached given a long-term general equilibrium. Building on this idea, the most well developed and widely applied superset of IO model is the CGE model. As discussed by Koks et al. (2015), CGE models focuses on the long-run future equilibrium state, contrasting to the nearer future focus of IO models. However, the expensive human capital investment required for the learning, building, and using of CGE remains a significant entry barrier for many scholars. The accuracy of CGE model in economic predictions is as well questioned by many (Polo and Viejo, 2015, Dixon and Rimmer, 2009, Zhou and Chen, 2020). A simpler version of chronological IO variant that focuses on the nearer future event would thus be a great complement to the existing toolbox of economic modelling.

Hence, jumping out of the general equilibrium assumption, Romanoff and Levine (1977) propose the sequential interindustry model (SIM) as another strain of IO model innovation to pollinate the time domain analysis into the IO model in a culminative linear interaction. Similar to the dynamic

material flow analysis (B. Müller, 2006), the SIM hybridises the time lag in demand propagation with the IO model by building a direct linear linkage between future and past economic activities. Some real-world SIM applications have also been made by later researchers. For instance, Okuyama et al. (2004) and Okuyama et al. (2000) use the SIM to assess the economic impact of natural disasters in Japan by using quarterly disaggregated hypothetical data. However, past attempts with the SIM proposed technical coefficient changes based only on hypothetical estimations, which greatly aggravates the uncertainties of modelling outcomes and limits further applications. This is largely because the data demands of the SIM are not easily met (Levine et al., 2007), thus hindering the SIM from fulfilling its deeper potential. It is also the reason why many IO researches often merely mention and discuss the SIM in literature reviews, but resolve to other IO variations as the tool for IO related modelling involving chronological analysis (Barker and Santos, 2010, Malik et al., 2014, Mendoza-Tinoco et al., 2017, Yu et al., 2013, Avelino and Hewings, 2019, Avelino, 2017).

In our opinion, the reasons that there is hardly any advancement in the SIM model towards high time-resolution innovation are as follows. 1) Researchers are too convinced that the IO model is an economic model that obeys the principle of general equilibrium, so efforts are concentrated on introducing more economic concepts, such as inoperability (Yu et al., 2013, Barker and Santos, 2010), to build a “perfect economic model”, but the fundamental interactions between economic in a physical way are sadly overlooked. If we apply SIM in a short run disequilibrium setting under which macroeconomic structural changes are less influential factors to be considered, we could exploit SIM’s potential in regression with big data. 2) The data requirement for high time-resolution analysis is costly. National IO tables are normally produced every 3-5 years. It takes great effort from IO scholars to increase the time resolution on a yearly basis (Avelino, 2017) to match the annually or quarterly updated economic indicators. Data unavailability has disincentivized model builders from working on a theoretical model with limited applications. Thus, limited methodological advancement has been made based on the concept proposed by SIM. As Leontief pointed out at his later stage of research, the IO model is not simply an economic method but can also be understood from the technical/engineering perspective (Leontief, 1991). In this research, we share the same spirit of Coluzzi et al. (2011) to look at the IO model from the perspective of data science. Through an innovative algorithm, we creatively simplified the complicated interactions among economic sectors proposed by SIM into a solvable linear model. Then, we pollinated linear regression technique into the new algorithm, so that the best fitted technical coefficient from sectorial demands and outputs observations can be reversely calculated across the discrete time intervals, instead of hypothetically and less accurately ‘divide’ production coefficients along the time domain like past SIM modellers did. A simulation based on dummy input has been conducted in this research to show that the algorithm proposed can accurately calculate the unknown chronological production coefficients based on simply output observations. The innovated algorithm will thus offer potential applications both in data science and in econometric interdisciplinary research. It opens the “black box” to reveal the relationships that cannot be told by conventional big data tools such as machine learning algorithms. Thus, scholars could have a new perspective to describe the structure of an economy in a chronological way. If the production coefficient of SIM is to be changed, we will be capable to investigate the impact of an external shock, such as nature disaster, on the economy in a timely dynamic way. On the other hand, if inputted with adequate proxy data, the proposed algorithm may extensively alleviate the

cost associated with IO table compilation, revolutionizing the way of conducting large scale economic statistics.

Methodology and Data

Sequential Interindustry Model (SIM)

First, let us recall that the classic Leontief input-output model can be written as follows:

$$\begin{aligned} x &= y + Ax \\ x &= (I - A)^{-1}y \end{aligned} \quad (1)$$

In equation (1), x is the total output in vector form. y is the final demand in vector form. A is the production coefficient matrix. I is the identity matrix. Using the Taylor expansion, the equation above can be rewritten as follows:

$$x = Iy + Ay + A^2y + \cdots + A^ny \quad (n \rightarrow \infty) \quad (2)$$

The physical meaning of equation (2) is that the output induced by demand takes effect through layers of intermediate productions. Both equations (1) and (2) are input-output fundamentals that anyone in the IO field can easily recall.

SIM then introduces the time domain by splitting x and y into finite discrete time intervals as time series, meaning that x and y can be rewritten as $x_{(t)}$ and $y_{(t)}$, respectively. Equations (3) and (4) illustrate the principle of this conversion.

$$\begin{aligned} x &= \sum_{t=1}^{m+n} x_{(t)} \\ x &= x_{(0)} + x_{(1)} + x_{(2)} + \cdots + x_{(m)} + \cdots + x_{(m+n)} \quad (n \rightarrow \infty) \end{aligned} \quad (3)$$

$$\begin{aligned} y &= \sum_{t=1}^m y_{(t)} \\ y &= y_{(0)} + y_{(1)} + y_{(2)} + \cdots + y_{(m)} \end{aligned} \quad (4)$$

For each of the terms, the subscript denotes the specific demand and output at discrete time t . m is the number of discrete time interval investigated. n is the number of propagation layers that ideally approach infinity (the same as demonstrated by Taylor expansion in equation (2)). It means that for m observations of $x_{(t)}$ and $y_{(t)}$, due to productional propagations, changes on output $x_{(t)}$ will theoretically extend into the infinitely distant future. Hence, considering the time lag feature of $x_{(t)}$ and $y_{(t)}$, equation (1) will then be changed to equation (5) below.

$$x_{(t)} = y_{(t)} + Ax_{(t-1)} \quad \forall t > 1, \quad y_{(t)} = 0 \text{ if } t > m \quad (5)$$

In equation (5), a recursive algorithm is introduced to obtain the output $x_{(t)}$ at time t based on

two variables: the current demand $y(t)$ at time t and total output $x_{(t-1)}$ at the previous discrete time of $(t - 1)$. The rationale is that what has been produced “today” will signal intermediate production “tomorrow” and propagate into the more distant future recursively. For instance, a hundred cars are manufactured and consumed “today”, so that four hundred tyres are used in the inventory of car making factories. Thus, receiving the market signal, tyre manufacturers will produce four hundred tyres “tomorrow” to respond to the consumption signal sent out “today”. Since SIM discusses the state of economic structure in a very short period of time, we assume that the economy will not be able to respond by investing in capital equipment or similar means to change its structure, but to change production level in the face of differed market signal. Although the length of time duration investigated is m , n more discrete time is extended into the future as the residue of the outputs induced by demand induced by outputs before time m takes effect. The following equations illustrate equation (5) in its expanded form to help readers understand the mathematics and the underlying idea.

$$\begin{aligned}
x_{(0)} &= y_{(0)} \\
x_{(1)} &= y_{(1)} + Ax_{(0)} = y_{(1)} + Ay_{(0)} \\
x_{(2)} &= y_{(2)} + Ax_{(1)} = y_{(2)} + Ay_{(1)} + A^2y_{(0)} \\
&\dots \dots \\
x_{(n)} &= y_{(n)} + Ax_{(n-1)} = y_{(n)} + Ay_{(n-1)} + A^2y_{(n-2)} + \dots + A^ny_{(0)} \\
&\dots \dots \\
x_{(m)} &= y_{(m)} + Ax_{(m-1)} = y_{(m)} + Ay_{(m-1)} + A^2y_{(m-2)} + \dots + A^ny_{(m-n)} \\
x_{(m+1)} &= Ay_{(m)} + A^2y_{(m-1)} + \dots + A^ny_{(m-n+1)} \\
x_{(m+2)} &= A^2y_{(m)} + A^3y_{(m-1)} + \dots + A^ny_{(m-n+2)} \\
&\dots \dots \\
x_{(m+n-1)} &= A^{n-1}y_{(m)} + A^ny_{(m-1)} \\
x_{(m+n)} &= A^ny_{(m)}
\end{aligned}$$

Since the row-sum of elements of production coefficients in A are all smaller than 1 as input must not exceed output for any productions, A^n converges to zero as n becomes sufficiently large. We can thus assume terms with A 's degree powers higher than n to be neglected.

Bearing that concept in mind, we next introduce the time domain into the production coefficient A in a similar manner to that of equations (3) and (4). A is thus converted into equation (6) as follows:

$$A = A_{(1)} + A_{(2)} + \dots + A_{(l)} \quad (6)$$

The reason for splitting A is to reflect the time lag that occurs between different sectors due to technical constraints (Romanoff and Levine, 1986). Hence, the modified production coefficients not only reflect the magnitude of intermediate inputs, but also the timing for the intermediate inputs to be fulfilled. For instance, tyres in the car manufacturing industry may be made quicker than cars' control chips due to reasons like the nature of engineering process and distances among different industry clusters. Such information cannot be captured by A in traditional IO models, but its spitted version in SIM as described by equation (6). An earlier simple attempt by Okuyama (2004) splits A into three stages. Equation (5) is thus further modified into equation (7), the general form of SIM as follows:

$$x_{(t)} = y_{(t)} + \sum_{i=1}^l A_{(i)} x_{(t-i)} \quad \forall i < t \quad y_{(t)} = 0 \text{ if } t > m \quad (7)$$

In an early work of Romanoff and Levine (1981), equation (7) is defined as the responsive production SIM where productions are responses to demand in the past. A further advancement in SIM modelling has been proposed to include anticipatory production of demand from the future. To avoid unnecessary complications, this research will only work with the general responsive production form of SIM.

For the purpose of illustration, the SIM process described in equation (7) is shown visually using the schematic diagram given in Figure 1 below. Each horizontal row shows how the output $x_{(t)}$ is comprised at the respective time t . Initially in the first row, it shows the state where $A_{(t)}$ and $y_{(t)}$ have not started interacting with each other. Until it moves to the second row of $t = 0$, $y_{(0)}$ is first multiplied by the identity matrix I to produce $x_{(0)}$, as highlighted in red on the left side. It suggests that output will match the market demand for the first time interval. In the next discrete time of $t = 1$, output at the previous discrete time, $x_{(0)}$, is multiplied by $A_{(1)}$ to produce one element that is to be added to the product of $y_{(1)}$ and I , according to the aforementioned assumption that output will match the instant demand. The product of $x_{(0)}$ and $A_{(1)}$ indicates that the production coefficient on the second discrete time will be applied to the output at first discrete time. It is to simulate the time-lagged response of intermediate productions to output of previous time discrete. The added sum then produces $x_{(1)}$ and is to be used in the next step of $t = 2$. Theoretically, this procedure is repeated for infinitely many processes. Essentially, readers may want to visualize the process as moving the series of $A_{(t)}$ and $y_{(t)}$ towards each other to perform a convolution-like operation, as demonstrated in Figure 1.

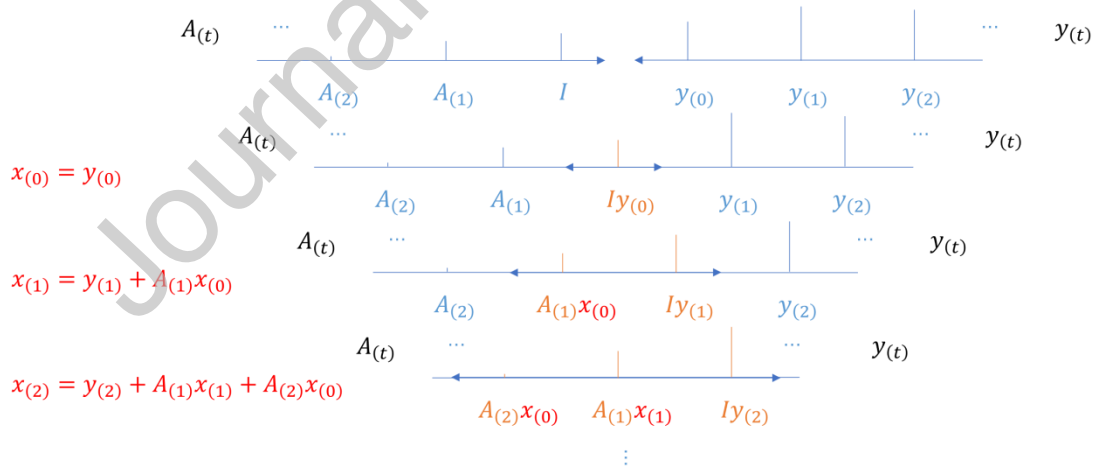


Figure 1 shows the algorithm process in a visual diagram. The entire process is very similar to the convolution operation and only different by introducing x (in red) in each step for the calculation of x for the next step. The orange variables represent the input data needed for each of the steps. The blue variables are the variables that are yet to participate in the operation.

For easier discussion, we constructed a SIM system with a dummy production coefficient matrix $A_{(t)}$ with two layers as shown in Table 1 below.

	$A_{(1)}$	$A_{(2)}$
--	-----------	-----------

	Sector A	Sector B	Sector A	Sector B
Sector A	0.2	0	0	0.7
Sector B	0	0.375	0.25	0

Table 1 A simulated production coefficient matrix in time-series of two sectors and two layers, where $A_{(1)}$ and $A_{(2)}$ shows the interaction of the two sectors in $t=1$ and $t=2$ respectively.

Assuming two unitary demand of 1 happens for Sector A at $t = 0$ and for Sector B at $t = 1$ respectively, denoted as vectors $y_{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y_{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, the following output of $x_{(t)}$ can be then obtained according to equation (7):

$$x_{(0)} = y_{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

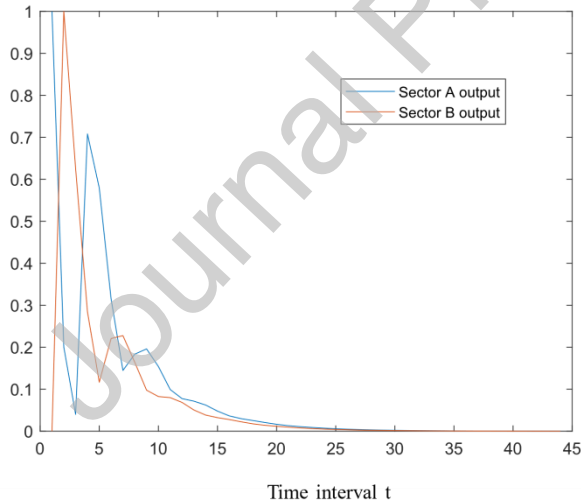
$$x_{(1)} = y_{(1)} + A_{(1)} x_{(0)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 \\ 0 & 0.375 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 1 \end{bmatrix}$$

$$x_{(2)} = A_{(1)} x_{(1)} + A_{(2)} x_{(0)} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.375 \end{bmatrix} \begin{bmatrix} 0.2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0.7 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.04 \\ 0.625 \end{bmatrix}$$

$$x_{(3)} = A_{(1)} x_{(2)} + A_{(2)} x_{(1)} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.375 \end{bmatrix} \begin{bmatrix} 0.04 \\ 0.625 \end{bmatrix} + \begin{bmatrix} 0 & 0.7 \\ 0.25 & 0 \end{bmatrix} \begin{bmatrix} 0.2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.708 \\ 0.2844 \end{bmatrix}$$

... ..

By organizing and plotting $x_{(t)}$ in Figure 2, it is easily observed that under SIM, the final output will decay towards zeros, validating the deductions in SIM we made previously.



	Sector A output	Sector B output
t=1	1.000	0.000
t=2	0.200	1.000
t=3	0.040	0.625
t=4	0.708	0.284
t=5	0.579	0.117
t=6	0.315	0.221
t=7	0.145	0.228
t=8	0.183	0.164
t=9	0.196	0.098
t=10	0.154	0.082
t=11	0.099	0.080
t=12	0.078	0.068
t=13	0.071	0.050
t=14	0.062	0.038
t=15	0.048	0.032
t=16	0.036	0.028
t=17	0.030	0.022
t=18	0.025	0.017
t=19	0.021	0.014
t=20	0.016	0.012
t=21	0.013	0.010
.....		

Figure 2 The decaying output over time as a response to the illustrative unitary demands in sector A and B and production coefficients $A_{(t)}$.

Model Innovation

The objective of our proposed algorithm is to obtain the production coefficient $A_{(t)}$ in a time series through ample observations of total output $x_{(t)}$ and final demand $y_{(t)}$. Although Levine et al. (2007) has proposed Z transformation to solve the convolution-like problem presented in SIM,

there is no mathematical solution to Z transformation for matrix variables into the frequency domain. Hence, no methodological advancement has been made to solve SIM model. It is natural to consider solving SIM in the time domain by utilizing the development of programming capacity developed in the recent decade.

The SIM model is thus to be solved within the time domain. In our proposed algorithm, we assume that productivity $A_{(t)}$ is unchanged, and production activity $x_{(t)}$ is only signalled by consumption activities $y_{(t)}$ as a responsive demand.

To reversely work out $A_{(t)}$, let us first write out equation (7) as follows:

$$\begin{aligned}
 x_{(0)} &= y_{(0)} \\
 x_{(1)} &= y_{(1)} + A_{(1)}x_{(0)} \\
 &= y_{(1)} + A_{(1)}y_{(0)} \\
 x_{(2)} &= y_{(2)} + A_{(1)}x_{(1)} + A_{(2)}x_{(0)} \\
 &= y_{(2)} + A_{(1)}y_{(1)} + (A_{(1)}^2 + A_{(2)})y_{(0)} \\
 x_{(3)} &= y_{(3)} + A_{(1)}x_{(2)} + A_{(2)}x_{(1)} + A_{(3)}x_{(0)} \\
 &= y_{(3)} + A_{(1)}y_{(2)} + (A_{(1)}^2 + A_{(2)})y_{(1)} + (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)})y_{(0)} \\
 &\dots \dots
 \end{aligned}$$

Which can be rewritten again into a linear system in matrix form, as shown in equation (8):

$$\begin{aligned}
 [I \quad A_{(1)} \quad (A_{(1)}^2 + A_{(2)}) \quad (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)}) \quad \dots] \times \begin{bmatrix} y_{(n)} & y_{(n+1)} & \dots \\ y_{(n-1)} & y_{(n)} & \dots \\ \vdots & \vdots & \dots \\ y_{(0)} & y_{(1)} & \dots \end{bmatrix} \\
 = [x_{(n)} \quad x_{(n+1)} \quad \dots]
 \end{aligned} \tag{8}$$

Let there be p sectors. If we take $(n+1) \times p$ observations of y to make the second matrix on the left a symmetric matrix, then the first term in equation (8) is solvable by taking an inverse of the second term, as shown in equation (9).

$$\begin{aligned}
 [I \quad A_{(1)} \quad (A_{(1)}^2 + A_{(2)}) \quad (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)}) \quad \dots] \\
 = [x_{(n)} \quad x_{(n+1)} \quad \dots] \times \begin{bmatrix} y_{(n)} & y_{(n+1)} & \dots \\ y_{(n-1)} & y_{(n)} & \dots \\ \vdots & \vdots & \dots \\ y_{(0)} & y_{(1)} & \dots \end{bmatrix}^{-1}
 \end{aligned} \tag{9}$$

By observing the term on the left-hand side of the equation, we can see that the time subscripts and power degrees of $A_{(t)}^q$ in each of the elements all sum up to be the same. For instance, the element $(A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)})$ has its terms' time subscripts and power degrees all sum up to 3. Physically, this can be explained by the different paths taken by production operations to reach the current layer; i.e., in $(A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)})$, the production sectors that are "closer" to each other and linked by $A_{(1)}$ will be induced three times compared to only once by $A_{(3)}$. Such an algorithm telling the combinations that satisfy such a requirement is called a partition in number theory, for which plenty of well-programmed functions in multiple programming platforms for its realization are available. Hence, knowing $A_{(1)}$ in the beginning, we can first obtain $A_{(2)}$ according to the expression of $A_{(2)}$ shown in the second row of equation set (10). Having $A_{(1)}$ and $A_{(2)}$, it is then feasible to obtain $A_{(3)}$ according to the third row of equation set (10). Thus, any $A_{(t)}$ for $t > 1$ can be obtained by recursive induction step by step.

$$\begin{aligned}
 A_{(1)} &= A_{(1)} \\
 A_{(2)} &= (A_{(1)}^2 + A_{(2)}) - A_{(1)}^2 \\
 A_{(3)} &= (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)}) - (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)}) \\
 &\dots\dots
 \end{aligned} \tag{10}$$

In equations (8) to (10), we have described a mathematic algorithm to linearly solve for an ideal SIM system. To apply SIM into the real world, we will need to combine statistical and econometric research techniques with SIM in order to compensate for uncertainties. We thus introduce linear regression to minimize the summation of square errors, so as to find the best fitted solution to $A_{(t)}$ based on observations of $y_{(t)}$ and $x_{(t)}$.

First, let the variables in equation (8) become:

$$\begin{aligned}
 B &= [I \quad A_{(1)} \quad (A_{(1)}^2 + A_{(2)}) \quad (A_{(1)}^3 + A_{(1)}A_{(2)} + A_{(2)}A_{(1)} + A_{(3)}) \quad \dots] \\
 U &= \begin{bmatrix} y_{(n)} & y_{(n+1)} & \dots \\ y_{(n-1)} & y_{(n)} & \dots \\ \vdots & \vdots & \dots \\ y_{(0)} & y_{(1)} & \dots \end{bmatrix} \\
 V &= [x_{(n)} \quad x_{(n+1)} \quad \dots]
 \end{aligned}$$

So that equation (8) becomes:

$$BU = V \tag{11}$$

The left-hand side of equation (11) is thus the estimated output based on $A_{(t)}$ and $y_{(t)}$, while the right-hand side is the real output. In equation (11), all variables are asymmetric matrices, unlike required equation (9), to include all observations of $x_{(t)}$ and $y_{(t)}$ available. The error in estimation is thus easily obtained as $BU - V$. Since all variables are matrices, the sum of squared errors is thus given by a simple matrix operation as:

$$f(x) = (BU - V)^T (BU - V) \tag{12}$$

Where the superscript T means transpose matrix. To find the value of B that minimizes $f(x)$, we first need to take the first derivative of $f(x)$ with respect to B , so that equation (12) becomes:

$$\begin{aligned}
 \frac{\delta f}{\delta B} &= \frac{\delta (BU - V)^T (BU - V)}{\delta B} \\
 &= \frac{\delta (U^T B^T BU - 2V^T BU + V^T V)}{\delta B} \\
 &= 2U^T B^T U - 2V^T U
 \end{aligned} \tag{13}$$

Then, we can set equation (13) to zero and find the expression of B at which minimizes the squared error sum given in equation (12). The following matrix operations are thus performed.

$$\begin{aligned}
 0 &= 2U^T B^T U - 2V^T U \\
 U^T B^T U &= V^T U \\
 BUU^T &= VU^T \\
 BUU^T (UU^T)^{-1} &= VU^T (UU^T)^{-1} \\
 B &= VU^+
 \end{aligned} \tag{14}$$

In equation (14), U^+ denotes the Moore-Penrose pseudoinverse matrix of U , widely used for

linear regression problems (MacAusland, 2014). The best fitted $A_{(t)}$ can thus be easily obtained.

Simulation Results

To verify the efficacy of the innovative algorithm proposed, we simulated a $x_{(t)}$ series based on the SIM interaction of a randomly generated 2-sector $y_{(t)}$ series of 200 discrete time intervals and the 2-layer production system $A_{(t)}$ described in Table 1. We first arbitrarily set propagation layers to be 40, so that the influence from demands at $y_{(0)}$ will be negligible after output $x_{(42)}$, i.e. after 40 layers of propagation and 2 layers of production by $A_{(t)}$. Starting to take observations from $x_{(43)}$, 158 observations of $x_{(t)}$ are used to construct the V matrix in equation (11). U matrix and its pseudoinverse are also constructed using all 200 observations of $y_{(t)}$ according to equation (14). Thus, the preliminary form of $A_{(t)}$ shown as B in equation (11) is easily obtained. Performing the partition algorithm illustrated in equation (10) will give us the regressed production coefficient $A'_{(t)}$.

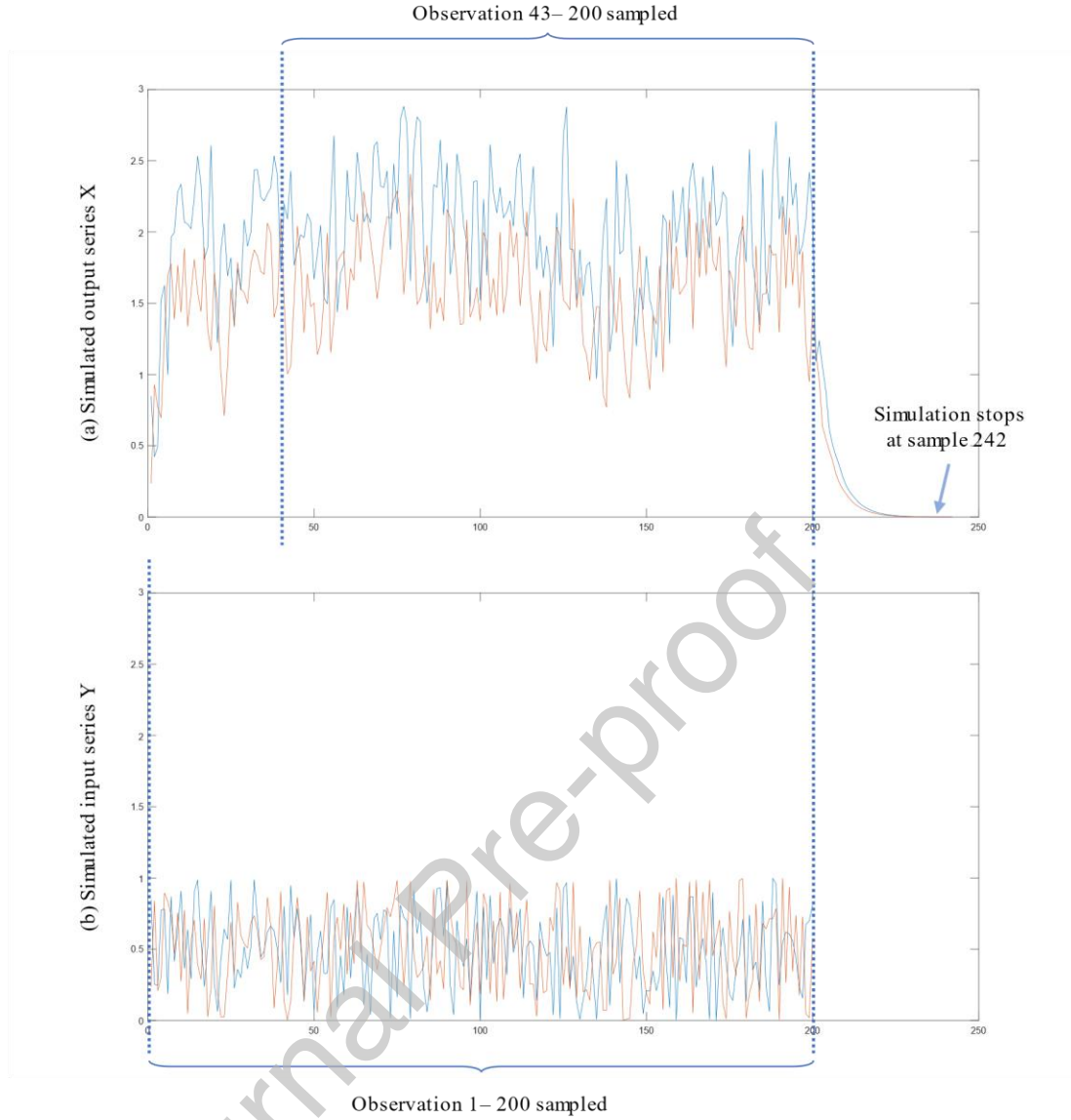


Figure 3 (a) Simulated total output $x(t)$ with 250 2×1 vectors in series obtained from the SIM interaction between demand $y(t)$ and production system $A(t)$. (b) Simulated final demand $y(t)$ with 200 randomized 2×1 vectors as input series.

Table 2 shows the errors between the $A'(t)$ obtained and the actual $A(t)$. Compared to $A(t)$, errors are minimal enough to be ignored and conclude that $A'(t)$ and $A(t)$ are basically identical, demonstrating the effectiveness of our innovated algorithm.

	$A'_{(1)} - A_{(1)}$		$A'_{(2)} - A_{(2)}$	
	Sector A	Sector B	Sector A	Sector B
Sector A	-1.43×10^{-6}	-0.11×10^{-6}	-9.96×10^{-6}	-5.75×10^{-6}
Sector B	-1.01×10^{-6}	-7.70×10^{-6}	-7.22×10^{-6}	-2.72×10^{-6}

Table 2 the errors between regressed production coefficient $A'(t)$ and the actual production coefficient $A(t)$

To quantitatively measure the squared mean error e , equation (15) below is used:

$$e = \frac{\sum_i^p \sum_j^p \sum_l^t (a_{ij}'^{(l)} - a_{ij}^{(l)})^2}{q} \quad (15)$$

In equation (15), $a_{ij}'^{(l)}$ and $a_{ij}^{(l)}$ is the i th column and j th row element at l th layer of $A'_{(t)}$ and $A_{(t)}$ respectively. p and t are the number of sectors and number of production layers respectively. In the simulation exercise, $p = 2$ and $t = 2$. q is the total number of elements in $A'_{(t)}$ and $A_{(t)}$, being $2 \times 2 \times 2 = 8$ in our case. The squared mean error between $A'_{(t)}$ and $A_{(t)}$ is thus 4.66×10^{-11} as calculated from Table 2, sufficiently small to conclude that $A'_{(t)}$ is a good enough approximation of $A_{(t)}$.

To further test the capability of the innovated algorithm for SIM, the squared mean errors are calculated under varied number of propagation layers. Figure 4 illustrates how the mean squared errors change with different number of propagation layers set. It is clearly seen that as the number of propagations increases, the mean square error between $A'_{(t)}$ and $A_{(t)}$ exponentially decreases, suggesting that $A'_{(t)}$ approaches the true value of $A_{(t)}$ as number of propagation increases. It makes sense as SIM models infinitely distant past demand to still have an impact on the present output, but minimal and negligible as propagation extends to the future. Increasing the number of propagations set in our algorithm factors in the increasingly minimal effect on output $x_{(t)}$ from demand $y_{(t)}$, thus producing a more accurate $A'_{(t)}$ as solution to the system.

Another important observation is that the squared mean error drastically increases to 1×10^{-2} after the number of propagation layers becomes more than 64, sufficiently large to invalidate the estimated $A'_{(t)}$. The reason is that insufficient number of sample available hinders the functioning of linear regression algorithm. When the number of propagation layers is 64, the dimension of U in equation (11) becomes 134-by-134, a symmetric matrix just sufficient to solve for B . As the number of propagation layers increases beyond 64, the system described in equation (11) becomes an underdetermined system with no unique solution. $A'_{(t)}$ thus deviates significantly from the true value $A_{(t)}$. If there are more observations available, more propagation layers can then be accommodated to improve the accuracy in the proposed algorithm further.

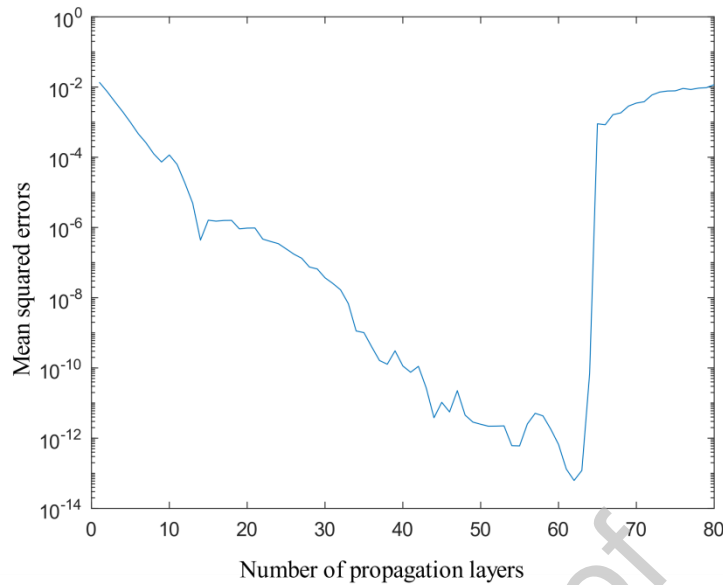


Figure 4 Mean squared errors of the elements in the timed production coefficient $A_{(t)}$ against varied number of propagation layers taken.

In the simulation exercise, we work on a system with known layers of production in $A_{(t)}$. In real world applications, the number of layers in production is normally unknown. It is thus sensible to vary both the number of production layers and number of propagation layers in real world applications to find the solution for best fitted $A_{(t)}$ with the minimum mean squared errors as the best approximation for chronological production coefficients.

Discussion & Conclusions

As a unique variation of IO model, SIM offers tremendous potential in economic system analysis. Unlike the dynamic IO model and its later development which keep general equilibrium as an underlying assumption, the SIM is by nature a disequilibrium model that focuses on the nearer future analysis. Unfortunately, disproportionally fewer efforts and achievement in SIM methodological advancement have taken place since its first proposal in 1980s. In this study, we propose an innovative algorithm as an important methodological advancement for the SIM. It complements with SIM by providing a way to integrate it with econometric linear regression analysis. Investigation into the chronologically extended production coefficients can provide temporal information into the interlinkages among economic sectors. For instance, although knowing the units of steel and the associated time needed to produce one unit of output in the manufacturing industry may be possible by conventional life cycle assessment methods, the cost of such an assessment will be tremendously high if comprehensively conducted at macro-level, not to mention the difficulty in constructing the sectorial interlinkages among all the industrial sectors in a similar manner as IO tables. The complementary algorithm to SIM provides a cheap and efficient way to quickly draw a comprehensive picture of chronological interlinkages based on economic activities in high time-resolution. Being more accurate than hypothetically constructed

in previous studies, the linearly regressed temporal production coefficients of SIM can be reversely used for short future predictions of sectorial outputs given unscheduled events such as natural disaster.

The proposed algorithm can also be deemed as a powerful tool to handle big data and better utilize them for the benefit of socio-economic symbiosis research. As illustrated in the simulation, the time resolution for demand and output observations must be high enough for the feasibility of linear regression. Daily or hourly demand and output observations should be preferably used in our algorithm. Although online transaction data across different sectors is ideally the best, it is highly unlikely that such data are available due to technical constraints at present. As a sensible compromise, proxy data with high chronological resolutions on economic activities, such as electricity consumption and satellite remote sensing, may be a solution to the stringent data quality requirement. Since there are limited modelling tools available for complex system symbiosis analysis, real world application and validation will be the desired next step forward for SIM and the proposed complement algorithm.

Extending the improved model beyond economic research, other complex networks similar to IO systems with delay characteristics, such as the human body metabolism system, can borrow the method presented in this study for their unique applications. Unlike macro econometric data, observation data on smaller systems are normally much easier to be obtained. Analysis similar to the previously proposed can be conducted to investigate how materials interact among themselves in a temporal manner to offer insight into the systems investigated.

However, algorithm proposed in this study unfortunately does not solve some fundamental limitation of SIM. Firstly, SIM does not address to the “bottlenecking” issue of IO, i.e., when demand exceeds the current production capacity, industries will not be able to linearly scale up production as a response. Secondly, SIM does not differentiate changes in capital and consumer goods. Unlike other model frameworks which have factored in the relationships between changes of capital goods and productivity, SIM is not able to well model the impact on capital changes as an endogenized parameter. Later research on SIM methodological improvement can take the momentum of this study to overcome the weakness of SIM by exploring its integration with other IO variant, such as dynamic IO analysis.

Data availability statement

Data are available upon request from the authors.

Declaration of interest statement

We wish to confirm that there are no known conflicts of interest associated with this publication and that there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

We would like to thank University College London for generously providing studentship to support K.H.'s postgraduate research. We are also thankful for the feedback provided by peer researchers during the 2020 International Input-Output Association Conference Young Researchers Session, which was hosted online due to the Covid-19 pandemic.

References

- AULIN-AHMAVAARA, P. Dynamic input-output and capital. Presentado en: 13th International Conference on Input-Output Techniques. University of Macerata, Italia.(Agosto 2000), 2000. Citeseer.
- AVELINO, A. F. & HEWINGS, G. J. 2019. The challenge of estimating the impact of disasters: Many approaches, many limitations and a compromise. *Advances in spatial and economic modeling of disaster impacts*. Springer.
- AVELINO, A. F. T. 2017. Disaggregating input–output tables in time: the temporal input–output framework. *Economic Systems Research*, 29, 313-334.
- B. MÜLLER, D. 2006. Stock dynamics for forecasting material flows—Case study for housing in The Netherlands. *Ecological Economics*, 59, 142-156.
- BARKER, K. & SANTOS, J. R. 2010. Measuring the efficacy of inventory with a dynamic input–output model. *International Journal of Production Economics*, 126, 130-143.
- COFFMAN, D. M. 2021. François Quesnay, Luigi Pasinetti and the historical contexts of economic theory. *Structural Change and Economic Dynamics*, 56, 64-73.
- COFFMAN, D. M. & KELSEY, J. 2019. *Applications of big data to construction*, Routledge.
- COLUZZI, B., GHIL, M., HALLEGATTE, S. & WEISBUCH, G. 2011. Boolean delay equations on networks in economics and the geosciences. *International Journal of Bifurcation and Chaos*, 21, 3511-3548.
- DIETZENBACHER, E., CAZCARRO, I. & ARTO, I. 2020. Towards a more effective climate

- policy on international trade. *Nature Communications*, 11, 1130.
- DIXON, P. & RIMMER, M. T. 2009. *Forecasting with a CGE model: Does it Work?*, Centre of Policy Studies (CoPS).
- GALBUSERA, L. & GIANNOPOULOS, G. 2018. On input-output economic models in disaster impact assessment. *International journal of disaster risk reduction*, 30, 186-198.
- GUAN, D., WANG, D., HALLEGATTE, S., DAVIS, S. J., HUO, J., LI, S., BAI, Y., LEI, T., XUE, Q., COFFMAN, D. M., CHENG, D., CHEN, P., LIANG, X., XU, B., LU, X., WANG, S., HUBACEK, K. & GONG, P. 2020. Global supply-chain effects of COVID-19 control measures. *Nature Human Behaviour*, 4, 577-587.
- HERTWICH, E. G. & PETERS, G. P. 2009. Carbon Footprint of Nations: A Global, Trade-Linked Analysis. *Environmental Science & Technology*, 43, 6414-6420.
- KOKS, E. E., CARRERA, L., JONKEREN, O., AERTS, J. C. J. H., HUSBY, T. G., THISSEN, M., STANDARDI, G. & MYSIAK, J. 2015. Regional disaster impact analysis: comparing Input-Output and Computable General Equilibrium models. Copernicus GmbH.
- KURZ, H. D. & SALVADORI, N. 2000. The Dynamic Leontief Model and the Theory of Endogenous Growth. *Economic Systems Research*, 12, 255-265.
- LENZEN, M., MORAN, D., KANEMOTO, K. & GESCHKE, A. 2013. BUILDING EORA: A GLOBAL MULTI-REGION INPUT-OUTPUT DATABASE AT HIGH COUNTRY AND SECTOR RESOLUTION. *Economic Systems Research*, 25, 20-49.
- LEONTIEF, W. 1970. Environmental Repercussions and the Economic Structure: An Input-Output Approach. *The Review of Economics and Statistics*, 52, 262-271.

- LEONTIEF, W. 1991. The economy as a circular flow. *Structural Change and Economic Dynamics*, 2, 181-212.
- LEONTIEF, W. W. 1936. Quantitative input and output relations in the economic systems of the United States. *The review of economic statistics*, 105-125.
- LEONTIEF, W. W. 1953. Studies in the Structure of the American Economy.
- LEVINE, S. H., GLORIA, T. P. & ROMANOFF, E. 2007. A dynamic model for determining the temporal distribution of environmental burden. *Journal of Industrial Ecology*, 11, 39-49.
- MACAUSLAND, R. 2014. The Moore-Penrose inverse and least squares. *Math 420: Advanced Topics in Linear Algebra*, 1-10.
- MALIK, A., LENZEN, M., ELY, R. N. & DIETZENBACHER, E. 2014. Simulating the impact of new industries on the economy: the case of biorefining in Australia. *Ecological Economics*, 107, 84-93.
- MENDOZA-TINOCO, D., GUAN, D., ZENG, Z., XIA, Y. & SERRANO, A. 2017. Flood footprint of the 2007 floods in the UK: The case of the Yorkshire and The Humber region. *Journal of Cleaner Production*, 168, 655-667.
- MI, Z., MENG, J., GUAN, D., SHAN, Y., SONG, M., WEI, Y.-M., LIU, Z. & HUBACEK, K. 2017. Chinese CO₂ emission flows have reversed since the global financial crisis. *Nature Communications*, 8, 1712.
- OKUYAMA, Y. 2004. Modeling spatial economic impacts of an earthquake: input-output approaches. *Disaster Prevention and Management; Bradford*, 13, 297-306.
- OKUYAMA, Y., HEWINGS, G. J. & SONIS, M. Sequential interindustry model (SIM) and

impact analysis: application for measuring economic impact of unscheduled events. 47th north American meetings of the regional science association international, Chicago, IL, 2000.

OKUYAMA, Y., HEWINGS, G. J. D. & SONIS, M. 2004. Measuring Economic Impacts of Disasters: Interregional Input-Output Analysis Using Sequential Interindustry Model. In: OKUYAMA, Y. & CHANG, S. E. (eds.) *Modeling Spatial and Economic Impacts of Disasters*. Berlin, Heidelberg: Springer Berlin Heidelberg.

OOSTERHAVEN, J. 1988. On the plausibility of the supply-driven input-output model. *Journal of Regional Science*, 28, 203-217.

POLO, C. & VIEJO, R. 2015. On the accuracy of cge forecasts in expansion and recession: spain 1990–1997. *Economic Systems Research*, 27, 525-542.

REYES, F. A. & MENDOZA, M. A. M. 2021. Demand-Driven and Supply-Sided Input-Output Models. *Journal of Quantitative Economics*, 1-17.

ROCCO, M. V. 2019. Integrating Energy and Economy models based on the Dynamic Input-Output framework. *IIOA*. Glasgow, United Kingdom.

ROMANOFF, E. & LEVINE, S. H. 1977. Interregional sequential interindustry modeling: a preliminary analysis of regional growth and decline in a two region case. *Northeast Regional Science Review*, 7, 87-101.

ROMANOFF, E. & LEVINE, S. H. 1981. Anticipatory and Responsive Sequential Interindustry Models. *IEEE Transactions on Systems, Man, and Cybernetics*, 11, 181-186.

ROMANOFF, E. & LEVINE, S. H. 1986. CAPACITY LIMITATIONS, INVENTORY, AND

TIME-PHASED PRODUCTION IN THE SEQUENTIAL INTERINDUSTRY MODEL.

Papers in Regional Science, 59, 73-91.

SHARRARD, A. L., MATTHEWS, H. S. & RIES, R. J. 2008. Estimating Construction Project Environmental Effects Using an Input-Output-Based Hybrid Life-Cycle Assessment Model. *Journal of Infrastructure Systems*, 14, 327-336.

STADLER, K., WOOD, R., BULAVSKAYA, T., SÖDERSTEN, C. J., SIMAS, M., SCHMIDT, S., USUBIAGA, A., ACOSTA-FERNÁNDEZ, J., KUENEN, J. & BRUCKNER, M. 2018. EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. *Journal of Industrial Ecology*, 22, 502-515.

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. *Review of International Economics*, 23, 575-605.

WEBER, C. L., PETERS, G. P., GUAN, D. & HUBACEK, K. 2008. The contribution of Chinese exports to climate change. *Energy Policy*, 36, 3572-3577.

YAGI, M., KAGAWA, S., MANAGI, S., FUJII, H. & GUAN, D. 2020. Supply Constraint from Earthquakes in Japan in Input-Output Analysis. *Risk analysis*, 40, 1811-1830.

YU, K. D. S., TAN, R. R. & SANTOS, J. R. Impact estimation of flooding in Manila: An inoperability input-output approach. 2013 IEEE Systems and Information Engineering Design Symposium, 2013. IEEE, 47-51.

ZENG, Z. & GUAN, D. 2020. Methodology and application of flood footprint accounting in a hypothetical multiple two-flood event. *Philosophical Transactions of the Royal Society A*, 378, 20190209.

ZENG, Z., GUAN, D., STEENGE, A. E., XIA, Y. & MENDOZA-TINOCO, D. 2019. Flood footprint assessment: a new approach for flood-induced indirect economic impact measurement and post-flood recovery. *Journal of Hydrology*, 579, 124204.

ZHOU, L. & CHEN, Z. 2020. Are CGE models reliable for disaster impact analyses? *Economic Systems Research*, 1-27.

Journal Pre-proof