Title: Leveraging Machine Learning in Input-Output Economic Modeling

Authors: Kehan HE[1], Zhifu MI*[2], Dabo GUAN*[2, 3, 4]

1. Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong SAR, China

2. The Bartlett School of Sustainable Construction, University College London, London, WC1E 6BT, UK

3. Department of Earth System Sciences, Tsinghua University, Beijing 100080, China

4. Institute for Carbon Neutrality, Tsinghua University, Beijing 100084, China

Abstract

This study investigates the use of machine learning to enhance the Input-Output (IO) model for analyzing economic interdependencies between industries and thus structural changes. Traditional IO models face challenges such as time lags and inaccuracies due to the labor-intensive nature of compiling IO tables. By employing machine learning, specifically Random Forest algorithm, this research aims to estimate and forecast economic structures more efficiently. Using aggregated OECD annual IO tables and World Bank Development Indicators, the study predicts the sector-to-sector production ratios, constructing a 9-by-9 IO structure based on 829 socio-economic indicators. The results demonstrate the method's potential, with 36% of the data points in the tested IO tables achieving relatively accurate predictions that have error levels less than ±30%. As Vietnam, an important emerging economy, transitions to more renewable energies, a sensitivity analysis of its economic structure, in response to varying renewable energy ratios, reveals non-linear effects on sector contributions. Simulation of this research suggests that increasing the renewable energy ratio from 15% to 2018's level leads to a decrease of 4.5% in the contribution of Chemical sector to its self-contribution, but a further increase of energy ratio to 25% will only decrease the same item by 0.4%. This approach shows promise as a cost-effective alternative to traditional IO economic models, offering policymakers an additional dimension of analysis to consider.

Introduction

The Input-Output (IO) model is a well-established economic framework used to analyze the interdependencies between industries (Thomassin, 2018). Originally developed by Wassily Leontief (Leontief, 1936), the model describes how the output of one sector serves as an input for another, facilitating a comprehensive understanding of economic flows (Rose and Miernyk, 1989). National statistics bureaus periodically compile IO tables to support economic analysis and policymaking. These tables have been instrumental in various applications, including environmental stress accounting that quantifies the environmental impacts of economic activities(Leontief, 1970, Zhao et al., 2009, He et al., 2021). They are also used in social footprint analysis to assess the social impacts of production and consumption (Wiedmann and Lenzen, 2018, Wiedmann et al., 2006) , disaster impact assessment to evaluate the economic consequences of natural and man-made disasters (Mendoza-Tinoco et al., 2020, Huang et al., 2022), and global supply chain studies to understand the complexities of international trade and production networks (Haddad et al., 2023).

Despite its utility, the IO model relies heavily on statistical surveys to compile the necessary IO tables, especially the intermediate interindustry Z matrx, which are prone to time lags, inaccuracies, and errors. Typically, national IO tables are updated approximately every three to five years, which can result in outdated information and necessitates extrapolation and interpolation for more frequent updates (Tukker et al., 2018, Eurostat, 2012). The labor-intensive nature of compiling these tables further complicates timely revisions. While developed countries have established standardized protocols for IO table construction (United Nations Statistical Division, 1999), less developed nations often struggle with issues like data scarcity (Antille, 1990, Eleish, 1963) and unreliable statistics (Singh, 1972). These challenges have driven researchers to seek more cost-effective alternatives for compiling IO tables, aiming to enhance both efficiency and accuracy (Lahr, 1993, Leblanc and Queyranne, 1980).

Alternative approaches employ econometric techniques to forecast changes in economic structures by analyzing socio-economic factors (Rodríguez-Pose, 1998) such as education levels (Rentería et al., 2016, Cuaresma and Mishra, 2011), population dynamics (Malmberg, 1994), and labor market trends (Castelló-Climent, 2019). Despite their utility, these methods often struggle to accurately capture the complex interlinkages between sectors Complex process-based economic models, such as the Computational General Equilibrium (CGE) model, attempt to quantify these interdependencies but face significant challenges due to data constraints (Adkins et al., 2003). Additionally, CGE models rely on simplified assumptions (Zhou and Chen, 2021) and require specialized expertise (An et al., 2023), which can compromise their predictive accuracy (Faehn et al., 2020).

Economists have explored hybrid approaches that integrate socio-economic parameters with economic models to address the challenges of capturing sectoral interlinkages (Merciai and Schmidt, 2018, Ayres, 1995). However, establishing direct theoretical linkages between these parameters and individual economic sectors remains difficult due to current limitations in data processing and computational capabilities (Faehn et al., 2020). In response, machine learning (ML) has emerged as a promising alternative for analyzing economic structural changes (Ghoddusi et al., 2019, Mullainathan and Spiess, 2017). Unlike theory-driven economic models, ML methods bypass stringent assumptions and instead detect empirical relationships among variables, allowing for the discovery of hidden interdependencies between socio-economic factors and economic structures (Vrontos et al., 2021, Guoa et al.). Although ML models do not explicitly articulate the underlying economic mechanisms, they facilitate rapid analysis within an IO framework, offering a powerful tool for understanding complex economic dynamics (He et al., 2025).

Building on this concept, our research introduces a ML-based method to estimate and predict economic interlinkages under IO structure, hence approximating the Z matrix in a cost-effective way. This approach aims to provide reliable estimations for missing IO tables and forecast structural changes under changing socio-economic conditions. The results demonstrate that the ML-based method can fairly predict inter-industry production within an IO framework. By using Vietnam as a case study, we show how the model effectively forecasts non-linear structural changes in response to variations in renewable energy ratios. Given the findings, this research suggests that the ML-based approach is a cost-effective alternative to traditional process-based and survey-reliant economic models. It can offer valuable policy insights for stakeholders on economic structure estimations and impact assessments amid evolving socio-economic conditions.

Method and Data

The central philosophy of this research posits that a nation's development follows a unique trajectory shaped by socio-economic indicators, which are essential for forecasting economic transformations. Socio-economic theories indicate that the relationships between economic indicators are complex and non-linear. A notable example is the Environmental Kuznets Curve (EKC), which proposes that as an economy expands, environmental degradation initially escalates but eventually diminishes once a certain income per capita threshold is surpassed. However, like many other economic theories, the EKC is a theoretical construct that has been criticized for its oversimplification and is increasingly challenged by recent empirical studies (Wang et al., 2024). This highlights the issue that, despite the complexity of non-linear theories, uncovering the relationships between socio-economic parameters remains difficult with process-based economic modeling. Consequently, there have been limited efforts to extrapolate changes in economic structures and to make predictions and estimations based on socio-economic indicators.

To address these challenges, this research employs machine learning techniques to analyze socio-economic indicators, enabling the estimation and understanding of economic development trajectories without the need for labor-intensive IO surveys and rigid economic theories. By leveraging comprehensive quantitative data—such as labor structure, natural resources, education levels, and urbanization rates—this approach can determine a country's stage of development and predict its economic structure as depicted by the IO table. This approach allows researchers to generate rapid estimations of an economy's IO-based structure, bypassing the need for traditional consensus surveys. Eventually, the goal of this algorithm is to advance the methodological development of IO models, thereby enhancing the understanding, estimation, and forecasting of economic structures within the IO framework.

Data Source

This study leverages a comprehensive dataset to examine the economic trajectories of various countries. The primary data source is the OECD annual IO tables (OECD, 2023), which span from 1995 to 2018 and cover 66 countries and economies. These tables provide detailed information across 45 economic sectors. However, the original 45-by-45 sector resolution necessitates training 2,045 models for each sector-to-sector interaction, posing a significant computational challenge to hardware devices. To address this, we aggregated the sectors into 9 broader sectors, namely Food, Chemical, Mining, Transport, Heavy Industry, Manufacturing, EHGW (Electricity, Heating, Gas, and Water), Construction, and Service. It effectively balances model complexity with available computational resources. Despite this reduction in computational demand, training with the Random Forest algorithm in this study still requires over 72 hours to complete a single round (using a system with 64.0GB RAM and an Intel Core i9-13900HX 2.2GHz processor).

In addition to the IO tables, we incorporate socio-economic indicators from the World Development Indicators (World Bank, 2024). Initially, the dataset includes 1478 indicators, but due to missing data for certain countries and years, we narrowed it down to 154 complete indicators. Recognizing the need for a more robust dataset, we supplemented these with 675 additional parameters derived from the IO tables. These parameters include sectoral value added, exports, imports, final demand, and capital formation, among others. Thus, the resulting dataset comprises a total of 829 indicators, which are used to train the Random Forest model. This model aims to determine the 9x9 intermediate input table $Z = \begin{bmatrix} z_{1,1} & \cdots & z_{1,9} \\ \cdots & \cdots & \cdots \\ z_{9,1} & \cdots & z_{9,9} \end{bmatrix}$, hence providing insights into the economic structural change and development trajectory of the countries studied. A complete list of the indicators of World

Bank Database and the aggregation concordance matrix for OECD IO tables are available in the Supplementary Information.

Random Forest Algorithm

Random Forest (Breiman, 2001) is an ensemble ML method used for and regression task in this study. It operates by constructing a multitude of decision trees during training and outputting the mode of the mean prediction of the individual trees. The process begins by generating several bootstrap samples from the economic indicators and IO table dataset. For each of these samples, a decision tree is constructed. The overall prediction of the Random Forest model is achieved by aggregating the predictions of the individual trees, leveraging the "wisdom of crowds" to improve accuracy and robustness.

In this study, the Random Forest regression process can be summarized as the following equation (1).

$$z_{ij} = f(k_1, k_2, \dots, k_n) \tag{1}$$

In equation (1), z_{ij} is the element of the intermediate production Z table that describes the input from sector i to produce output in sector j. k_i is the ith social-economic indicator provided by the World Bank database. In addition, other elements in the IO tables, such as final demand, value added, export/import etc., are also taken out to be considered as a social-economic indicator in k_i. As explained in the beginning of this section, this study assumes that countries are bounded to development trajectory, so that time becomes an invariant in this process. If country A in year $t1$ has exactly the same social-economic parameters as country B in year $t2$, the simulation will conclude that country A in year $t1$ and country B in year $t2$ will have the same intermediate production ratio. Hence, the annual IO tables of 66 countries/regions from 1995 to 2018 as given by OECD database provides 1584 data points. For this study, there are 81 (9-by-9) elements in the Z matrix, so 81 Random Forest models need to be trained based on the same 829 indicators.

Concretely, the Random Forest model is based on decision tree algorithm. A decision tree is a supervised learning algorithm that splits data into subsets based on feature values to predict an output, forming a tree-like model of decisions. Each internal node represents a test on a feature, each branch corresponds to an outcome of the test, and each leaf node holds a prediction value.

Let K be the input indicator containing n features in vector form, z be the output intermediate production ratio in scalar form. So that the training set containing m observations can be expressed in equation (2)

$$S = \{(K_1, z_1), (K_2, z_2), \dots, (K_m, z_m)\}, K \in \mathbb{R}^n, Z \in \mathbb{R}$$

(2)

For each of the Random Forest model, a bootstrap aggregation, or bagging, is performed by randomly sampling on the training dataset S to generate bootstrap samples $S^{(b)}$. From the bootstrap samples $S^{(b)}$, a regression decision tree $T^{(b)}$ is constructed by recursively partitioning the feature space. At each node of the regression decision tree $T^{(b)}$, a random subset of the p features ($p < n$) is selected. Among the p features, the algorithm chooses the best split by minimizing the Mean Squared Error (MSE) of the target variable in the resulting child node.

After training B trees, the ensemble prediction for a new input K is made by averaging the predictions from all individual trees as shown in equation (3)

$$\widehat{z_{ij}} = \frac{1}{B} \sum_{b=1}^{B} T^{(b)}(K)$$

(3)

Here, $T^{(b)}(K)$ is is the predicted output of the b-th tree for input K, and $\widehat{z_{ij}}$ is the predicted input from sector i to produce output in sector j in the Z matrix of the IO table.

Scenario Setting

In this study, Vietnam was chosen as a case study to demonstrate the application of our method for deriving policy implications. Vietnam is an ideal case due to its rapid economic growth and increasing emphasis on sustainable development. The country is at a critical juncture where it is essential to balance economic expansion with environmental sustainability from an early stage of development. Vietnam has committed to achieving net-zero emissions by 2050 through a swift transition to renewable energy generation. With a substantial renewable energy potential of 1,000 GW, Vietnam significantly surpasses other developing nations in the region (Deffarges et al., 2023). Therefore, this study specifically varied the development indicator "Renewable Energy Ratio" to explore its impact on Vietnam's economic structure, as reflected by the estimated changes to the IO table. With reference to equation (1), the sensitivity analysis can be described by equation (4) as follows.

$$\Delta z_{ij} = f(k_1, k_2, \dots, \Delta k_l, \dots, k_n)$$

(4)

In equation (4), only the investigated indicator Δk_l, in this case "Renewable Energy Ratio", is varied to show how each of the cells in the intermediate production matrix will change, show by Δz_{ij} in equation (4).

To ensure the accuracy and consistency of the estimated IO table, we applied the RAS algorithm, a well-established method for balancing IO tables (Jackson and and Murray, 2004). This step is crucial for maintaining the balancing relationships dictated by the IO model. By examining these scenarios, we aim to provide insights into how shifts toward renewable energy can influence sectoral interactions and overall economic dynamics in Vietnam. Additionally, the Vietnam scenario demonstrates how the proposed Random Forest algorithm, combined with IO modeling, can offer real-world policy implications in development analysis.

Result

Effectiveness Analysis

This research considers a prediction model to be relatively accurate if the prediction falls within 70% to 130% of the true value, or has an error level of ±30%. Figure 1 illustrates the effectiveness of the proposed framework in constructing an IO intermediate production matrix. At a 68% confidence interval, 15.5% to 59.0% of the predictions for the 81 cells are relatively accurate.

Among production sectors, the Service sector achieves the highest average accuracy rate at 60.4%, while the Construction sector has the lowest at 21.6%. This may be because the Service sector, which relies more on labor input of educated population, is closely linked to socio-economic factors like education level that are recorded in more detail by the World Bank database used in this research. In contrast, the Construction sector involves more informal economic activities that are less accurately represented in socio-economic indicators, such as catering and transportation provided to construction workers. Without effective socio-economic indicators, the forecast capability of the Random Forest algorithm is thus deteriorated, shown by the less accurate predictions by the Construction sector.

As for consumption sectors, the Manufacturing sector shows the highest average accuracy rate at 54.0%, whereas the Mining sector has the lowest at 26.0%. The Manufacturing sector's reliance on upstream consumption and its well-documented records submitted to statistical bureaus strengthen its relationship with other production sectors, thus improving the prediction accuracy rate. Conversely, the limited upstream connections to the Mining sector weaken this relationship, reducing the relative accuracy rate.

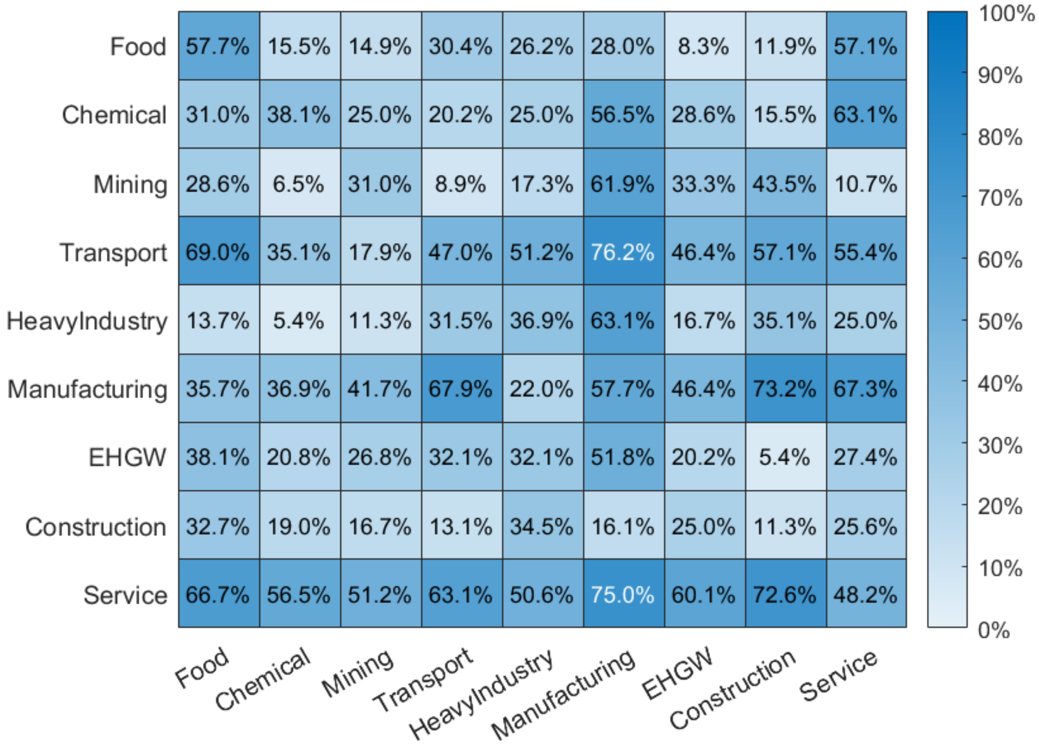

Figure 1 The distribution of relatively accurate predictions for the 9-by-9 intermediate production matrix in the training set. This research considers the prediction to be relatively accurate if the predicted value falls within the range of 70% to 130% of the true value, or an error level of ±30%. The color scale represents the number of relatively accurate predictions of the respective cells in the intermediate production matrix in percentage.

To evaluate the capability of the trained models in predicting unseen data, the training was repeated with the last seven countries (Sweden, Thailand, Tunisia, Turkey, USA, Vietnam, and South Africa) excluded from the original 66 countries, forming a verification set. This left 59 countries/regions for training in this round. Figure 2 illustrates the effectiveness of the trained models on the seven countries in the verification set. On average, these countries demonstrated a relative accuracy rate of 36% across the 81 trained models. Notably, Sweden exhibited the highest relative accuracy rate at 53% when reconstructing the intermediate production matrix. This suggests that the socio-economic indicators provided for Sweden in this study are more closely linked to its IO structure, resulting in better modeling accuracy. The variation in accuracy across countries may also reflect differing strengths in the relationship between economic structures and the socio-economic indicators used in the training.

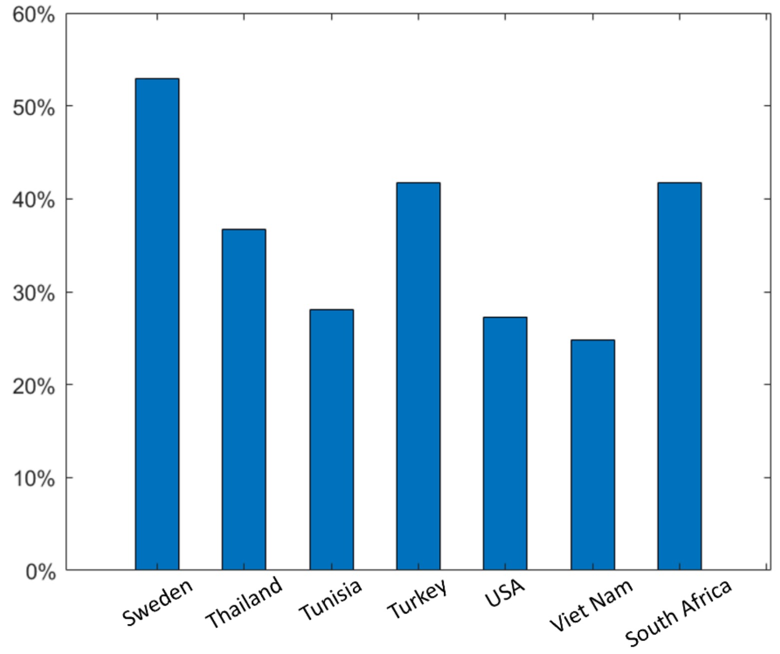

Figure 2 The distribution of relatively accurate predictions for the 7 countries in the verification set. The height of the bars represents the relative accuracy for the corresponding country in the intermediate production matrix in percentages.

Case Study

As a vibrant emerging economy under rapid transition, Vietnam was selected as a case study to demonstrate the application of the proposed method in economic analysis. An uncertainty analysis was conducted to validate the method's effectiveness in accurately reconstructing the economic structure depicted by the intermediate IO matrix. In this training round, all observations in the dataset were used to train the models, except for Vietnam's IO table from 2018. By comparing the prediction results for Vietnam, this case study aims to further validate the effectiveness of the Random Forest algorithm specifically for Vietnam.

As shown in Figure 3 (a), when reconstructing the IO structure of Vietnam in 2017 using economic parameters that are included in the training set, the method designed in this research is capable to rather accurately reproduce the correct intermediate production ratio within an error range of -17.4% to 60.2% (95% confidence interval) for all 81 locations in the intermediate production matrix. If given a verification set of economic development indicators for the year 2018, as shown in Figure 3 (b), the uncertainty performance of the method developed deteriorates to an error range of -38.4% to +216.1% (95% confidence level). In addition, the accuracy of predictions for different intermediate production cells are distributed differently. Specifically, accuracy for input to Manufacturing sector and output from Chemical sector are relatively higher than other sectors. The reason may be that the social-economic indicators used in this training are more linked to these two sectors, thus yielding higher accuracy in the prediction made by the model.

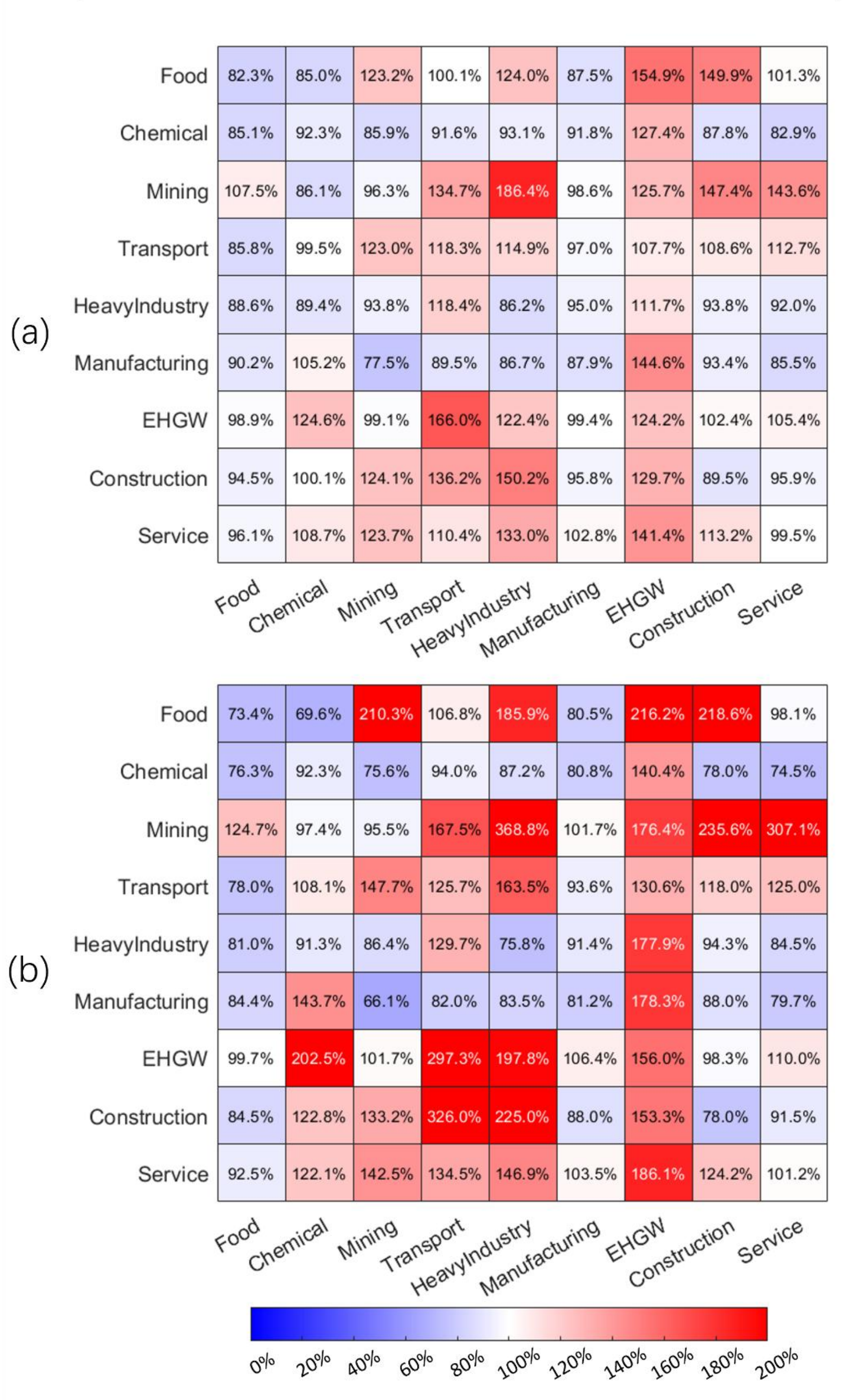

Figure 3 Effectiveness of the method in reconstructing the 9-by-9 intermediate production matrix of Vietnam for the (a) training year of 2017 and (b) verification year of 2018. The percentages in each cell represents the

ratio of the reconstructed data to real data. The blue and red colors show the extent of negative and positive variations of the reconstructed data from the real data respectively.

As a demonstration of the potential of this model as a viable alternative to complex economic models, this research has conducted a sensitivity analysis for economic structure change given changes in renewable energy ratio for Vietnam. It clearly shows the complex and nonlinear change of economic structures given changes in the economic indicators of a country.

The simulation shown in Figure 4 suggests a nonlinear pattern of economic structure change given an increased ratio of renewable energy. For instance, it shows that when the renewable energy ratio increased from 15% to the level of 2018 (21.8%), the contribution of Chemical sector to itself dropped tremendously by 4.5%, while the contribution of Chemical sector to EHGW increased by 0.4%. It reveals that at this stage of development, Vietnam's chemical industries, more importantly the petroleum industry, will tremendously improve its efficiency by reducing contribution from petroleum industry and switch its intermediate demand to cleaner energy source of electricity and gas. On the other hand, if the renewable energy ratio is further increased to a hypothetical level of 25% with all other economic parameters unchanged, the decrease in Chemical sectors' contribution to itself will be reduced to only 0.4%, although still the most significant change among all intermediate inter-industry contributions. It reveals the diminishing improvement of chemical industry's efficiency given increased renewable energy ratio. If the ratio of renewable energy further increased to 30%, the changes of chemical sector to all the other sectors are not as large and limited to only less than 0.1%. On the contrary, contribution from the Mining sector to EHGW sector will significantly increase to 0.5%, suggesting the development of Vietnam will be faced with a bottle neck in the extraction of raw materials, which leads to higher cost and increased intermediate contribution in the mining sector to clean energy generation.

(a)

	Food	Chemical	Mining	Transport	HeavyIndustry	Manufacturing	EHGW	Construction	Service
Food	-0.0%	-0.1%	-0.0%	-0.0%	-0.5%	-0.1%	0.0%	-0.0%	-0.0%
Chemical	-0.0%	-4.5%	-0.0%	0.2%	-0.0%	-0.0%	0.4%	-0.0%	-0.1%
Mining	-0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%
Transport	-0.0%	0.0%	0.0%	0.1%	-0.0%	-0.0%	0.0%	0.0%	0.0%
HeavyIndustry	-0.1%	-0.0%	0.0%	0.0%	-0.0%	-0.0%	0.2%	0.0%	-0.0%
Manufacturing	-0.0%	0.2%	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%	0.0%
EHGW	0.0%	-0.0%	-0.0%	0.1%	0.0%	0.0%	-0.1%	0.0%	-0.0%
Construction	-0.0%	0.0%	0.1%	0.0%	-0.0%	0.0%	0.1%	0.1%	-0.0%
Service	-0.0%	0.0%	-0.0%	0.0%	0.0%	-0.0%	0.0%	0.0%	0.0%

(b)

	Food	Chemical	Mining	Transport	HeavyIndustry	Manufacturing	EHGW	Construction	Service
Food	-0.0%	-0.1%	-0.1%	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%	-0.0%
Chemical	-0.2%	-0.4%	0.0%	-0.0%	-0.0%	-0.0%	0.0%	-0.0%	-0.0%
Mining	-0.1%	-0.0%	0.0%	0.2%	0.0%	-0.0%	0.2%	-0.0%	0.0%
Transport	-0.0%	0.0%	0.0%	0.0%	0.0%	-0.0%	0.0%	0.0%	0.0%
HeavyIndustry	-0.0%	-0.0%	0.0%	0.1%	-0.0%	-0.0%	0.0%	-0.0%	-0.0%
Manufacturing	-0.0%	0.2%	-0.0%	-0.0%	-0.0%	-0.0%	0.0%	-0.0%	0.0%
EHGW	-0.0%	0.0%	0.0%	0.1%	-0.0%	0.0%	-0.1%	0.0%	-0.0%
Construction	-0.0%	-0.0%	0.1%	0.0%	-0.0%	-0.0%	0.0%	-0.0%	-0.0%
Service	-0.0%	0.0%	-0.0%	0.0%	-0.0%	-0.0%	0.0%	0.0%	0.0%

(c)

	Food	Chemical	Mining	Transport	HeavyIndustry	Manufacturing	EHGW	Construction	Service
Food	0.0%	-0.1%	-0.0%	0.0%	-0.4%	0.0%	-0.1%	0.0%	0.0%
Chemical	0.0%	-0.1%	0.0%	-0.0%	-0.0%	-0.0%	0.0%	-0.0%	0.0%
Mining	-0.0%	0.0%	0.0%	0.1%	-0.2%	0.0%	0.0%	-0.0%	0.0%
Transport	0.0%	0.0%	-0.0%	0.0%	0.0%	0.0%	0.0%	-0.0%	-0.0%
HeavyIndustry	-0.0%	0.0%	0.0%	-0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Manufacturing	0.0%	0.1%	0.0%	0.0%	-0.0%	0.0%	-0.0%	0.0%	0.0%
EHGW	-0.0%	-0.0%	0.5%	-0.0%	-0.0%	0.0%	-0.0%	0.0%	0.0%
Construction	0.0%	0.1%	0.0%	-0.0%	-0.1%	0.0%	0.1%	0.2%	0.0%
Service	0.0%	0.0%	0.2%	0.0%	-0.0%	0.0%	0.0%	-0.0%	0.0%

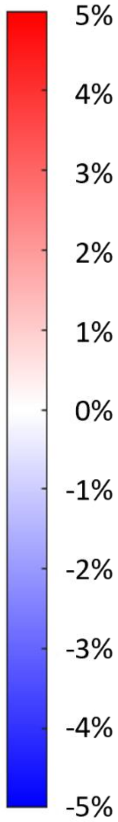

Figure 4 The percentage change in each cell of the intermediate production matrix of Vietnam, given the renewable energy ratio of Vietnam changes from (a) 15% to 21.8% (actual value in 2018), (b) 21.8% (actual value in 2018) to 25%, and (c) 25% to 30%.

Conclusion

In this study, we demonstrated the potential of leveraging machine learning, specifically the Random Forest algorithm, to predict IO structures effectively. Our findings indicate that this approach can achieve varying levels of accuracy across different sectors and countries, with the service sector showing a notable accuracy rate of 44%. The case study on Vietnam further highlighted the model's capability to simulate economic structure changes in response to shifts in renewable energy ratios, revealing complex, nonlinear patterns of sectoral interdependencies.

The integration of machine learning into IO modeling offers significant advantages over traditional methods, particularly in terms of cost-effectiveness and the ability to uncover complex relationships without relying on intricate process-based economic models. This approach shows promise for broader applications in economic forecasting, providing a viable alternative to more complex models in analyzing changes in economic structure. It offers an efficient tool for assessing policy sensitivity.

However, the study also faced challenges, primarily related to computational constraints and data availability. The limited computational capacity of the research team's devices restricted the inclusion of more parameters, which could enhance model accuracy. Additionally, the quality and completeness of available data remain critical factors influencing the model's performance. In the World Bank database used for Random Forest training, many datasets were incomplete for the period from 1995 to 2018, leading to the exclusion of valuable indicators such as healthcare metrics and development indices. Future research should focus on overcoming these limitations by incorporating additional parameters from diverse sources and exploring the model's applicability. There is also potential for integrating this machine learning approach with other economic modeling techniques to further enhance its predictive capabilities and add explanatory features to the analysis of the model.

Overall, this research serves as a first attempt to the cross-over between IO modeling and machine learning techniques. We encourage further exploration and development of machine learning methodologies to advance economic modeling and forecasting.

References

ADKINS, L. C., RICKMAN, D. S. & HAMEED, A. 2003. Bayesian Estimation of Regional Production for CGE Modeling. *Journal of Regional Science,* 43**,** 641-661.

AN, K., ZHANG, S., ZHOU, J. & WANG, C. 2023. How can computable general equilibrium models serve low-carbon policy? A systematic review. *Environmental Research Letters,* 18**,** 033002.

ANTILLE, G. 1990. Input–Output Tables and Scarcity of Data. *Economic Systems Research,* 2**,** 147-156.

AYRES, R. U. 1995. Thermodynamics and process analysis for future economic scenarios. *Environmental and Resource Economics,* 6**,** 207-230.

BREIMAN, L. 2001. Random Forests. *Machine Learning,* 45**,** 5-32.

CASTELLó-CLIMENT, A. 2019. The Age Structure of Human Capital and Economic Growth. *Oxford Bulletin of Economics and Statistics,* 81**,** 394-411.

CUARESMA, J. C. & MISHRA, T. 2011. The role of age-structured education data for economic growth forecasts. *Journal of Forecasting,* 30**,** 249-267.

DEFFARGES, J., VU, H., DELTEIL, B., HATA, T., PHAM, V., SENGUPTA, S. & TARA, K. 2023. *Putting renewable energy within reach: Vietnam's high-stakes pivot* [Online]. McKinsey. Available: https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/putting-renewable-energy-within-reach-vietnams-high-stakes-pivot [Accessed].

ELEISH, G. E. 1963. USES OF THE INPUT-OUTPUT MODEL IN DEVELOPMENT PLANNING IN UNDERDEVELOPED COUNTRIES. *Review of Income and Wealth,* 1963**,** 286-317.

EUROSTAT 2012. *Eurostat-OECD Methodological Manual on Purchasing Power Parities (2012 Edition)*, OECD Publishing.

FAEHN, T., BACHNER, G., BEACH, R., CHATEAU, J., FUJIMORI, S. I., GHOSH, M., HAMDI-CHERIF, M., LANZI, E., PALTSEV, S. V. & VANDYCK, T. 2020. Capturing key energy and emission trends in CGE models: Assessment of status and remaining challenges. Discussion Papers.

GHODDUSI, H., CREAMER, G. G. & RAFIZADEH, N. 2019. Machine learning in energy economics and finance: A review. *Energy Economics,* 81**,** 709-727.

GUOA, Y., LIA, J., LOA, T., ZHUA, Z., LEEA, G. & TOSCASA, P. Discovering Relationships Among Economic Variables using Machine Learning Techniques.

HADDAD, E. A., ARAúJO, I. F., ROCHA, A. & SASS, K. S. 2023. Input–output analysis of the Ukraine war: A tool for assessing the internal territorial impacts of the conflict. *Regional Science Policy & Practice,* 15**,** 8-56.

HE, K., COFFMAN, D. M., HOU, X., LI, J. & MI, Z. 2025. An electricity big data application to reveal the chronological linkages between industries. *Economic Systems Research,* 37**,** 76-94.

HE, K., MI, Z., CHEN, L., COFFMAN, D. M. & LIANG, S. 2021. Critical transmission sectors in embodied atmospheric mercury emission network in China. *Journal of Industrial Ecology,* 25**,** 1644-1656.

HUANG, R., MALIK, A., LENZEN, M., JIN, Y., WANG, Y., FATURAY, F. & ZHU, Z. 2022. Supply-chain impacts of Sichuan earthquake: a case study using disaster input–output analysis. *Natural Hazards,* 110**,** 2227-2248.

JACKSON, R. & AND MURRAY, A. 2004. Alternative Input-Output Matrix Updating Formulations. *Economic Systems Research,* 16**,** 135-148.

LAHR, M. L. 1993. A Review of the Literature Supporting the Hybrid Approach to Constructing Regional Input–Output Models. *Economic Systems Research,* 5**,** 277-293.

LEBLANC, D. & QUEYRANNE, M. 1980. Comparison of alternative methods of updating Input-Output tables: A note. *Technological Forecasting and Social Change,* 16**,** 143-146.

LEONTIEF, W. 1970. Environmental Repercussions and the Economic Structure: An Input-Output Approach. *The Review of Economics and Statistics,* 52**,** 262-271.

LEONTIEF, W. W. 1936. Quantitative Input and Output Relations in the Economic Systems of the United States. *The Review of Economics and Statistics,* 18**,** 105-125.

MALMBERG, B. 1994. Age structure effects on economic growth — Swedish evidence. *Scandinavian Economic History Review,* 42**,** 279-295.

MENDOZA-TINOCO, D., HU, Y., ZENG, Z., CHALVATZIS, K. J., ZHANG, N., STEENGE, A. E. & GUAN, D. 2020. Flood Footprint Assessment: A Multiregional Case of 2009 Central European Floods. *Risk Analysis,* 40**,** 1612-1631.

MERCIAI, S. & SCHMIDT, J. 2018. Methodology for the Construction of Global Multi-Regional Hybrid Supply and Use Tables for the EXIOBASE v3 Database. *Journal of Industrial Ecology,* 22**,** 516-531.

MULLAINATHAN, S. & SPIESS, J. 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives,* 31**,** 87–106.

OECD 2023. STAN Input Output Database (Edition 2022).

RENTERíA, E., SOUTO, G., MEJíA-GUEVARA, I. & PATXOT, C. 2016. The Effect of Education on the Demographic Dividend. *Population and Development Review,* 42**,** 651-671.

RODRíGUEZ-POSE, A. 1998. Social Conditions and Economic Performance: The Bond Between Social Structure and Regional Growth in Western Europe. *International Journal of Urban and Regional Research,* 22**,** 443-459.

ROSE, A. & MIERNYK, W. 1989. Input–Output Analysis: The First Fifty Years. *Economic Systems Research,* 1**,** 229-272.

SINGH, P. 1972. INPUT/OUTPUT ANALYSIS—AN APPRAISAL IN THE CONTEXT OF AS YET AN UNFINISHED EXPERIMENT IN KENYA. *Review of Income and Wealth,* 18**,** 393-398.

THOMASSIN, P. J. 2018. The Evolution of Input–Output Analysis. Springer

Singapore.
TUKKER, A., DE KONING, A., OWEN, A., LUTTER, S., BRUCKNER, M., GILJUM, S., STADLER, K., WOOD, R. & HOEKSTRA, R. 2018. Towards Robust, Authoritative Assessments of Environmental Impacts Embodied in Trade: Current State and Recommendations. *Journal of Industrial Ecology,* 22**,** 585-598.
UNITED NATIONS STATISTICAL DIVISION 1999. *Handbook of input-output table compilation and analysis*, UN.
VRONTOS, S. D., GALAKIS, J. & VRONTOS, I. D. 2021. Modeling and predicting U.S. recessions using machine learning techniques. *International Journal of Forecasting,* 37**,** 647-671.
WANG, Q., LI, Y. & LI, R. 2024. Rethinking the environmental Kuznets curve hypothesis across 214 countries: the impacts of 12 economic, institutional, technological, resource, and social factors. *Humanities and Social Sciences Communications,* 11**,** 292.
WIEDMANN, T. & LENZEN, M. 2018. Environmental and social footprints of international trade. *Nature Geoscience,* 11**,** 314-321.
WIEDMANN, T., MINX, J., BARRETT, J. & WACKERNAGEL, M. 2006. Allocating ecological footprints to final consumption categories with input–output analysis. *Ecological Economics,* 56**,** 28-48.
WORLD BANK 2024. World Development Indicators. *In:* BANK, W. (ed.).
ZHAO, X., CHEN, B. & YANG, Z. F. 2009. National water footprint in an input–output framework—A case study of China 2002. *Ecological Modelling,* 220**,** 245-253.
ZHOU, L. & CHEN, Z. 2021. Are CGE models reliable for disaster impact analyses? *Economic Systems Research,* 33**,** 20-46.