Jing YANG, PhD Candidate
E-mail: jaynescu@gmail.com
Business School, Sichuan University, Chengdu, China
Jiahao LIU, Master
E-mail: 18380122790@163.com
Business School, Sichuan University, Chengdu , China
Professor Zhaoxia GUO, (Corresponding Author)
E-mail: zxguo@scu.edu.cn
Business School, Sichuan University, Chengdu , China
Siquan ZHANG, PhD
E-mail: shiquanzhang@scu.edu.cn
School of Mathematics, Sichuan University, Chengdu, China
Lei GAO, PhD
E-mail: lei.gao@csiro.au
CSIRO, Glen Osmond SA, Australia
Michalis HADJIKAKOU, PhD
E-mail: m.hadjikakou@deakin.edu.au
School of Life and Environmental Sciences, Deakin University
Burwood VIC, Australia
Professor Brett BRYAN
Email: b.bryan@deakin.edu.au
School of Life and Environmental Sciences, Deakin University
Burwood VIC, Australia

A MULTIPLE NONLINEAR REGRESSION MODEL FOR ESTIMATING INPUT-OUTPUT VALUES WITH PARTIAL INFORMATION

Abstract. *The paper aims to find an effective method for estimating input-output (I-O) values in the target year when the original I-O table is incomplete. By examining the input and output relationship hidden in the RAS method, we find a multiple nonlinear regression model embedded in this method to achieve this well. Specifically, each element value in the target I-O table is closely related to its corresponding element in the original I-O table and the input and output growth rates of each sector. The performance of the proposed model is validated by extensive experiments based on simulated and real I-O datasets.*

Keywords*: Input-output analysis, RAS approach, input-output matrix, economic modelling, industry analysis.*

JEL Classification: C02, C51, C53, C67

DOI: 10.24818/18423264/56.3.22.11

1. Introduction

Input-output (I-O) analysis was initially developed as an effective tool for regional, national, and international economic analyses based on the interdependence of economic sectors or regions (Oosterhaven, 2019). A sector could be a region or an industry. Nowadays, it has been widely used in various areas such as economic analysis (Haddad et al., 2021), urban carbon transformations (Chen et al., 2017), energy trade and consumption analysis (Cui et al., 2015), and logistics forecasting (Hwang, 2014).

I-O analysis is based on an I-O table capturing flows among sectors or industries in different regions or nations (Miller and Blair, 2009; Oosterhaven, 2019). The I-O table is fundamental to I-O analysis, and it is crucial to building a reliable I-O table for the base year using data from a variety of economic sectors. The collection of these data requires considerable effort, resources, and time (Mahajan et al., 2018). Statistical agencies usually collect thorough I-O statistics at regular intervals and compute detailed I-O tables for different time periods. However, due to the dynamic nature of economic activities, the I-O relationship in the base year (i.e., the original I-O) may not reflect the input and output of subsequent years. Given some element values in the original I-O table and some constraints, this paper investigates how to obtain corresponding element values in the target I-O table.

Researchers have developed new methods to update an I-O table in a target year (i.e., the target I-O) so that it is more representative of the situation in the target year (Clark and Chenery, 1959; Miller and Blair, 2009; Stone and Brown, 1962). Methods for updating I-O tables can be classified as optimization methods such as the RAS method and its variants (Lenzen et al., 2009) and biproportional techniques (Lahr and De Mesnard, 2004), extrapolation methods for estimating trends and prediction coefficients (Miller and Blair, 2009), and expert methods such as cross-impact analysis (Weimer-Jehle, 2006).

Since an I-O table describing the total input and output volumes of a target year can be obtained relatively cheaply with the RAS method, it has received extensive attention from both researchers and practitioners (Hwang, 2014; Lahr and De Mesnard, 2004; Lenzen et al., 2009; Miller and Blair, 2009; Weimer-Jehle, 2006). Given an original I-O table and the total input and output volumes of each economic sector for the target year, the RAS method derives a target I-O table that satisfies the I-O constraints of the target year by iteratively updating the direct technical coefficients in the original I-O table. The total input or output volumes refer to the row or column sums of the desired I-O table, respectively. Lecomber (1975) proposed the Modified RAS (MRAS) method, which uses traditional minimum distance optimization to update the I-O table. The Generalized RAS (GRAS) method proposed by Junius and Oosterhaven (2003) allows for some

negative elements in the I-O table, which further extends the scope of the RAS method. While Jackson and Murray (2004) claimed that minimizing squared differences of the elements in GRAS leads to a smaller information gain, Oosterhaven (2005) pointed out that this conclusion is theoretically impossible. Valderas-Jaramillo and Rueda-Cantuche (2021) extended the generalized RAS (GRAS) technique to a multi-regional or multi-national setting for the estimation of multiple matrices in an integrated framework. Lenzen et al. (2009) proposed Konfliktfreies RAS (KRAS), which combines the characteristics of MRAS, and GRAS and allows the application of the RAS method in the case of data conflicts.

In existing applications, the RAS method and its variants operate based on the original I-O table in the base year and this I-O table is iterated to obtain the target I-O table. However, in some real-world applications, it is difficult to obtain a complete original I-O table (Lenzen et al., 2009) and easier to obtain the flow between one sector and its associated sectors. In certain cases, such as when sectors are disaggregated or where input recipes are updated to reflect new technologies (Wolfram et al., 2016), we may only be interested in I-O data for a certain sector or certain elements of the target year. Similar cases widely exist in the estimation of freight, population and capital flow movement between certain regions or sectors. For example, in logistics sector, the logistics authority of a city needs to estimate the freight volumes in a target year between this city and its associated cities so that the city can make more effective logistics and traffic planning (Liu et al., 2018). In this case, the city can only obtain the flow between this city and its associated regions, but cannot obtain and does not care about the flows between associated regions. This leads to an open and interesting research problem: when the original I-O table only contains data from certain elements, can we estimate their corresponding I-O values in the target year? This is the original motivation of this paper.

We first model the mathematical relationship reflected by the RAS method and then use this relationship to estimate the values of specific elements in the target I-O table in the absence of information on other elements. We then assess the performance of our method by using experiments and comparing our method with the conventional RAS method based on both simulated and real-world I-O data

2. Materials and methods

2.1. Traditional RAS method

Consider a closed economic system with N sectors (regions). All element values in the I-O table in the base year are given, as shown in Figure 1. The first N rows and N columns correspond to sectors 1 to N. Let x_{ij} denote the flow from origin sector i to destination sector j, which is the output volume generated by sector i for sector j. Let u_i stand for the intermediate total output from sector i, defined as the sum of the flow from sector i to N sectors (i.e., $u_i = \sum_{j=1}^{N} x_{ij}$). Let v_j

denote the intermediate total input to sector j, defined as the sum of the flow from N sectors to sector j (i.e., $v_j = \sum_{i=1}^{N} x_{ij}$). Here, u_i and v_j are placed in the last column and the last row of the I-O table, respectively. The total output of all sectors is equal to the total input (i.e., $\sum_{i=1}^{N} u_i = \sum_{j=1}^{N} v_j$).

We use the superscripts 0 and 1 to represent the base year and the target year, respectively, so M^0 and M^1 correspond to the I-O tables of the base year and the target year, respectively. Given the original I-O table $M^0 = \{x_{ij} | 1 \le i \le N, 1 \le j \le N\}$, the set $U^1 = \{u_i^1 | 1 \le i \le N\}$ of row sums, and the set $V^1 = \{v_j^1 | 1 \le j \le N\}$ of column sums in the target year, the traditional RAS method adopts the iterative procedure (Procedure 1) below to obtain the element values in the target I-O table M^1, where the sum of the i^{th} row (or j^{th} column) is equal to or as close as possible to the corresponding desired row sum u_i (or column sum v_i) in the target I-O table.

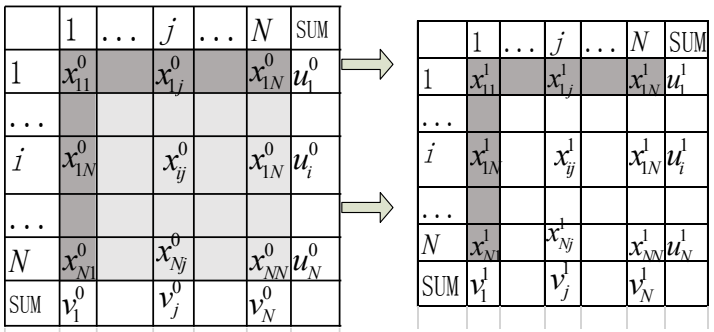

Figure 1. Original and target I-O tables and their symbols: a) I-O table M^0 in the base year, and b) I-O table M1 in the target year. In M^0 and M^1, some cells are shaded in dark or light gray, which will be further explained in section 2.2.

Procedure 1:

Define $r_i = \frac{u_i^1}{\sum_{j=1}^{N} x_{ij}}, 1 \le i \le N$ and $s_j = \frac{v_j^1}{\sum_{i=1}^{N} x_{ij}}, 1 \le j \le N$

While $\{|r_i - 1| > \varepsilon$ or $|s_j - 1| > \varepsilon\}$,

{

Set $x_{ij} \leftarrow x_{ij} \cdot r_i,\ 1 \le i \le N,\ 1 \le j \le N,$

Update $s_j = \dfrac{v_j^1}{\sum_{i=1}^{N} x_{ij}},\ 1 \le j \le N,$

Set $x_{ij} \leftarrow x_{ij} \cdot s_j,\ 1 \le i \le N,\ 1 \le j \le N,$

Update $r_i = \dfrac{u_i^1}{\sum_{j=1}^{N} x_{ij}},\ 1 \le i \le N$

}

In the traditional RAS procedure above, we let ε to denote the error tolerance, which is a small positive number (e.g., 0.001 or 0.005).

2.2. Using random values to replace missing values

To model the mathematical relationship reflected by the RAS method, we use random values to replace missing values in the base year first and examine how they affect the output values in the target year. For convenience of presentation, we assume that only the flow between sector 1 and its associated sectors are known and other values are missing in the base year. It is reasonable since comparing to obtaining the complete I-O table, it is much easier to obtain the flow between the sector we care about and its associated sectors in many real-world settings (e.g., regional I-O analysis, regional freight and trade projection). The total output of all sectors is equal to the total input (i.e., $\sum_{i=1}^{N} u_i = \sum_{j=1}^{N} v_j$). Let F^0 (F^1) denote the set consisting of the elements from both the first row and the first column of M^0 (M^1), and R^0 the unknown flows between the remaining $N-1$ sectors in the base year, as shown in Figure 1. In M^0, F^0 and R^0 are shaded in dark and light gray, respectively. In M^1, only F^1 is shaded in dark gray. We aim to obtain the element values in F^1 when only F^0 is known in M^0. The traditional RAS method cannot be used to handle this case since it does not work without a complete original I-O table. This case is realistically meaningful since sometimes it is hard to obtain a complete original I-O table or even a complete data of the corresponding row or column.

Let F^1 represents the flow between the first sector and its associated $N-1$ sectors in the target year, and the effects of R^0 element values on F^1 may be small. It is thus natural to assume that using random variables within a certain range based on historical data to replace missing values in R^0 could have little effects on the values in F^1. If the assumption holds, the values in F^1 can then be obtained by the traditional RAS method whichever random element values are used in R^0. We will verify this assumption based on a large number (K) of repetitive experiments

(indexed by k). In each experiment k, we randomly change the element values in R^0 and perform the RAS method to re-balance the I-O table. By so doing, we obtain the element values of F^1 in the target I-O table.

Let x_{ij}^{1k} denote the element value of row i and column j of M^1 in the k^{th} experiment. We take the set of element values in F^1 from the first experiment as the reference values. Let e_{ij}^k denote the relative error in the k^{th} experiment for the element of row i and column j of M^1. We have $e_{ij}^k = \left|x_{ij}^{1k} - x_{ij}^{11}\right| / x_{ij}^{11} \cdot 100\%$. Taking e_{ij}^k obtained in experiment k as an independent sample, we can obtain a total of $(N^2-1)\cdot(k-1)$ samples of relative error values in k repeated experiments. By analyzing the statistical results (e.g., mean and quantiles) of these samples, we can quantify the effects of changes in R^0 element values on F^1 element values.

The set V^1 of input totals and the set U^1 of output totals for all sectors are given in the RAS method. When most element values are missing in the original I-O table, we cannot directly obtain the values of V^1 and U^1. However, we can set V^1 (U^1) to V^0 (U^0) multiplied by specific growth rates as these growth rates are equivalently given by the RAS method. Let a_i (b_j) denote the growth rate of the output (input) total of sector i (j) from the base year to the target year. We have $a_i = (u_i^1/u_i^0 - 1) \times 100\%$ and $b_j = (v_j^1/v_j^0 - 1) \times 100\%$. The growth rate range of [−50%, 150%] covers the majority (>99.4%) of real-world applications of I-O analysis according to data in the World Input-Output Database (WIOD)[1]. Indeed, we analyzed all I-O datasets in WIOD in 2001–2014 and found that in 99.7% of cases, growth rates were less than 150%, while in 99.6% of cases, growth rates were greater than −50%.

Without the loss of generality, we set N to 19 and k to 10^5 in our validation experiments. Growth rates were randomly generated within [−50%, 150%] in each experiment. The random values in R^0 and the constant values in F^0 obeyed a uniform distribution within [1, 10^8]. We conducted extensive experiments based on various distributions, such as uniform, normal, and Poisson, with different distributions, which led to very similar results. For each element x_{1j}^{1k} or x_{j1}^{1k} ($2 \le j \le N$) in F^1, we recorded its 10^5-1 relative error values from the 10^5 experiments. The validation experiment is presented in the Supplementary Information[2].

Based on this experimental setting, we produced and recorded the average relative errors of all elements in F^1 in each experiment. We obtained 10^5-1 average relative error values from 10^5 experiments. As shown in Table 1, 90% of average relative errors were no greater than 2.20%, while 100% of average relative errors were no greater than 2.95%. The value of element x_{ij}^1 in F^1 obtained by the RAS method was almost independent of element values in R^0. This indicates that our

[1] http://wiod.org/database/wiots16
[2] https://www.dropbox.com/s/ab8pgas9o5v5vty/SupplementaryInfo.zip?dl=0

assumption of replacing element values of R^0 with random values is reasonable and has little effect on the accuracy of F^1 element values.

Table 1. Maxima of average relative errors for different data percentages in 105 experiments

Percentage of data	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Maximum average relative error	1.29%	1.40%	1.48%	1.55%	1.62%	1.69%	1.77%	1.87%	2.20%	2.95%

2.3. Multiple nonlinear regression model embedded in the RAS method

As the element value x_{ij}^1 in the target I-O table M^1 can be determined from the corresponding element value x_{ij}^0 in M^0 and the set $\{a_i\}$ ($\{b_j\}$) of growth rates of each sector's total input (output), this section explores the relationship between these variables with a multiple nonlinear regression model based on a large number of numerical experiments. The procedure for modeling the relationship involves two steps: variable selection and model construction.

2.3.1. Variable selection

The value of the elements x_{ij}^1 in the target I-O table M^1 may be closely related to x_{ij}^0, a_i and b_j. For simplicity of presentation, we define two I-O ratio (IOR) variables α_i and β_j to replace a_i and b_j in model construction. Let α_i (β_j) denote the ratio of the output (input) total of sector i (j) in the target year to its corresponding value in the base year. That is, $\alpha_i = u_i^1/u_i^0 \times 100\% = a_i + 1$ and $\beta_j = v_j^1/v_j^0 \times 100\% = b_j + 1$. Let μ, σ, t_{max}, and t_{min} represent the mean, variance, maximum, and minimum of all I-O ratios, respectively. For example, we have $t_{max}=\max\{\alpha_1,...,\alpha_N,\beta_1,...,\beta_N\}$. We set $\gamma = t_{max} / t_{min}$. We selected seven variables (x_{ij}^0, α_i, β_j, μ, σ, t_{max}, and t_{min}) and their interaction terms as candidate variables for the multiple nonlinear regression model. To simplify the model, we first used Pearson correlation coefficients to calculate linear correlations between each candidate variable and the variable x_{ij}^1. The 15 candidate variables with correlations >0.4, listed in Table 2, were used as independent variables in the first iteration of model fitting.

Table 2. Candidate variables and their correlations with x_{ij}^1.

Variables	$x_{ij}^0\alpha_i\beta_j$	$x_{ij}^0\alpha_i\beta_j t_{min}$	$x_{ij}^0\alpha_i\beta_j\gamma$	$x_{ij}^0\alpha_i$	$x_{ij}^0\beta_j$	$x_{ij}^0\alpha_i t_{max}$	x_{ij}^0	$x_{ij}^0\beta_j t_{max}$
Correlation with x_{ij}^1	0.98	0.92	0.92	0.9	0.89	0.88	0.88	0.87
Variables	$x_{ij}^0 t_{max}$	$x_{ij}^0 t_{min}$	$x_{ij}^0\alpha_i t_{min}$	$x_{ij}^0\beta_j t_{min}$	$x_{ij}^0\alpha_i\gamma$	$x_{ij}^0\beta_j\gamma$	$x_{ij}^0\gamma$	
Correlation with x_{ij}^1	0.78	0.76	0.65	0.64	0.63	0.58	0.45	

2.3.2. Model construction

In a closed economic system with N sectors, we assumed that the element x_{ij}^0 in the original I-O table M^0 satisfies a normal distribution within [x_{min}, x_{max}] and the I-O ratio variables (α_i and β_j) of each sector satisfy a normal distribution within [t_{min}, t_{max}]. Based on these settings, we randomly generated different sets of M^0, α_i, and β_j in S repeated experiments. Let (x_{ij}^0, x_{ij}^1) denote an element pair consisting of the element value x_{ij}^0 in M^0 and its corresponding value x_{ij}^1 in M^1. In each experiment, given the set of randomly generated M^0, α_i, and β_j, we used the RAS method to obtain the corresponding M^1 and randomly selected one element pair (x_{ij}^0, x_{ij}^1). We then recorded the corresponding parameters (α_i, β_j, μ, σ, t_{min}, t_{max}, and γ). We obtained one input and output sample pair in each experiment, where these parameters and x_{ij}^0 made up the sample input and x_{ij}^1 represented the sample output. Accordingly, we obtained S input and output sample pairs in S repeated experiments. If S is sufficiently large, these sample pairs can represent the input and output relationship between the input variables and x_{ij}^1. We used these samples as training and test samples for model construction and validation.

A large number of variables can lead to the "curse of dimensionality" in the model-fitting process. To avoid this problem, we used stepwise bidirectional elimination (Jennrich and Sampson, 1968) to construct the multivariate nonlinear regression model. This approach starts with no variables in the model. In terms of F-tests, we added the variable (if any) whose inclusion improved model fitting the most (e.g., the variable with the maximum partial sum of squares) and then deleted the variable (if any) whose loss deteriorated model fitting the most. This process was repeated until variables included in the model could not be eliminated and no new variables could be introduced. We selected the model with the maximum coefficient of determination (R-squared). The obtained set of independent variables and the corresponding coefficients constituted the final multivariate nonlinear regression model.

We considered a closed economy with 20 sectors (N = 20) and set x_{min} = 0, x_{max} = 10^7, t_{min} = 50%, t_{max} = 250%, and S = 60,000. Using stratified sampling,

50,000 sets of simulated input and output sample pairs were obtained as training samples for model fitting, while the remaining 10^4 simulated sample pairs were used as test samples. Let $\hat{x}_{ij}^1$ denote the estimate of x_{ij}^1. For stepwise bidirectional elimination, we used a variable entrance probability of 0.05 and a removal probability of 0.1. The following multiple nonlinear regression (MNR) model was obtained:

$$\hat{x}_{ij}^1 = 0.974x_{ij}^0\beta_j + 0.56x_{ij}^0\alpha_i + 0.212x_{ij}^0\alpha_i\beta_j - 0.483x_{ij}^0 t_{\min} + 0.046x_{ij}^0\alpha_i\beta_j\gamma - 0.242x_{ij}^0\beta_j t_{\max} - 0.088x_{ij}^0\alpha_i\gamma \quad (1)$$

Equivalently, we have

$$\hat{x}_{ij}^1 = 0.974x_{ij}^0(b_j+1) + 0.56x_{ij}^0(a_i+1) + 0.212x_{ij}^0(a_i+1)(b_j+1) - 0.483x_{ij}^0 t_{\min} + 0.046x_{ij}^0(a_i+1)(b_j+1)\gamma - 0.242x_{ij}^0(b_j+1)t_{\max} - 0.088x_{ij}^0(a_i+1)\gamma \quad (2)$$

An ANOVA test was performed showing that the variables in the MNR model were statistically significant at a level of 0.05. The adjusted R-squared of our MNR model was 0.991. The MNR model included multicollinearity since its terms are dependent, but multicollinearity was acceptable in this model because: (i) it did not decrease predictive reliability and only affected coefficient estimates and (ii) we only cared about model performance and not regression coefficients (Stock and Watson, 2015). We examined the Pearson correlation coefficients between the five variables (x_{ij}^0, α_i , β_j, $t_{\min}$, and $t_{\max}$) used in the MNR model and found that these variables were not significantly correlated, indicating that multicollinearity was only caused by the interaction of the five variables. This was not an issue as *p*-values for interaction terms are not influenced by multicollinearity.

2.3.3. Model validation

Both simulated and real-world data were used to perform performance validation on our MNR model. We used the 10^4 sets of simulated test samples obtained from stratified sampling (described in sub-section 2.3.2) as simulated data. For each test sample, we recorded the relative error of the estimated output $\hat{x}_{ij}^1$ in the MNR model compared to the actual output x_{ij}^1 generated by the RAS method. The actual output of the RAS method is generated based on a complete original I-O table according to the method described in section 2.2. A statistical analysis was performed on the relative errors of the MNR model's outputs for all test data.

We used real-world I-O data to further validate the accuracy and generalization of the MNR model in real-world applications. We used all available I-O data from 1995 to 2012 (a total of eight years) from the official website of the China Input-Output Association[3]. Between 1995 and 2012, the classification of

[3] http://www.stats.gov.cn/ztjc/tjzdgg/trccxh/zlxz/trccb/201701/t20170113_1453448.html

China's domestic economic sectors has changed due to institutional reforms and restructuring, with the number of economic sectors increasing from 33 in 1995 to 42 in 2005 and decreasing to 39 in 2012. We adjusted the I-O tables to make them comparable. First, rows and columns with similar industrial characteristics were aggregated. Second, we removed I-O tables from years with relatively many missing values (i.e., years 2000 and 2012). We thereby obtained six I-O tables with 28 economic sectors. On the basis of the six tables, we generated five validation datasets (i.e., I-O table pairs (M^0, M^1)) by grouping pairs of I-O tables from adjacent years. The six I-O tables and five validation datasets are shown in the Supplementary Information[4].

The independent variables required in the MNR model were obtained from M0 and M1. In the five validation datasets (1–5), all element values were within [1, 106] and their I-O growth rates ranged within [−72%, 268%], [−24%, 136%], [35%, 490%], [24%, 255%], and [23%, 127%], respectively. The average growth rates of each sector were ~50%. Each validation dataset (M0, M1) included 784 (28 × 28) validation sample pairs. For each validation sample pair, we took as inputs x_{ij}^0 in M0 and related independent variables and used the MNR model to calculate corresponding model outputs $\hat{x}_{ij}^1$. We calculated the relative error of each model output.

For both simulated and real-world data, given the I-O table in the base year and growth rates, we use the RAS method to generate the target I-O table directly, and use the MNR model to generate the element values in the target I-O table in turn.

2.4. Sensitivity analysis

We conduct the sensitivity analysis to examine the effects of different parameter settings on the performance of the MNR model from three perspectives, including different I-O values, different growth rate ranges, and different I-O data dimensions. We considered four I-O value ranges ([0, 10^4], [0, 10^6], [0, 10^8], and [10^4, 10^7]), four growth rate ranges ([−50%, 150%], [−50%, 200%], [−50%, 300%], and [−70%, 300%])[5], and seven I-O data dimensions (10, 20, 30, 50, 80, 100 and 200), respectively. For each range or each data dimension, with other parameter settings fixed as described in sub-section 2.3.2, we randomly generated 10^5 test samples according to the method described in the same sub-section, and calculated the relative error of the MNR model output for each test sample.

[4] https://www.dropbox.com/s/ab8pgas9o5v5vty/SupplementaryInfo.zip?dl=0
[5] Four corresponding IOR value ranges are [50%, 250%], [50%, 300%], [50%, 400%] and [30%, 400%].

3. Results

3.1. Validation results for simulated data

For the 10^4 simulated test samples, the mean of relative errors of model outputs was 4.77%. As shown in Table 3, 90% of relative errors were no greater than 9.18% and the maximum relative error was 12.51% for all test samples. These results show that the MNR model can map the input and output relationships of sample data and approximate the RAS method under the parameter settings, including I-O value ranges [0, 10^7]) and growth rate range [−50%, 150%].

Table 3. Results of maximum relative errors of model outputs for different data percentages in test samples

Percentage of data	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Maximum relative error	0.30%	0.61%	0.98%	1.47%	2.15%	3.02%	4.23%	5.99%	9.18%	12.51%

3.2. Validation results for real-world data

Figure 2 shows the comparison results of the maximum relative errors generated by the MNR model and the RAS method for datasets 1-5. For dataset 5, the relative errors of 90% of validation samples generated by the MNR model were no greater than 27.59% and the means of these relative errors were 14.02%. However, 90% of the relative errors generated by the MNR model for datasets 1 to 4 were less than 173.88%, 155.73%, 85.33%, and 44.76%, respectively, and the means of these relative errors were 93.74%, 67.84%, 62.52%, and 22.69%, respectively. As for the relative errors of the corresponding outputs generated by the RAS method, for datasets 1 to 5, 90% of the relative errors were less than 231.27%, 165.41%, 93.66%, 61.44%, and 28.63%, respectively, and the means of these relative errors were 111.5%, 68.9%, 68.92%, 31.26%, and 12.48%, respectively. Our MNR model and the RAS method exhibited similar trends in relative errors over different percentages of validation samples for each dataset (Figure 2). The key advantage of the MNR model over the RAS method is that the MNR model does not need the complete I-O table to generate target values, which is much simpler to use. Comparing to the relative errors for the simulated dataset in section 3.1, both the RAS method and the MNR model led to larger relative errors for datasets 1 to 5. The reasons for this are twofold. First, the real-world datasets involve a multi-year period and usually show the larger growth rates than the simulated dataset, which could lead to the larger relative errors. Second, the real-world datasets may contain some information that cannot be handled well by the classical RAS method. Even so, the MNR model performs better than the RAS

method for nearly all datasets over different percentage distributions (Figure 2). This is understandable because there is no 100% match between the MNR model and the RAS method, although it is constructed based on the RAS method. On the other hand, it is worth noting that the focus of this paper is not to develop a method that can estimate the I-O values in the target year with a perfect accuracy, but to develop an effective alternative to the RAS method for real-world I-O analysis applications where only partial I-O values are available in the original table. In these cases, an estimation accuracy comparable to the RAS method is reasonable and acceptable.

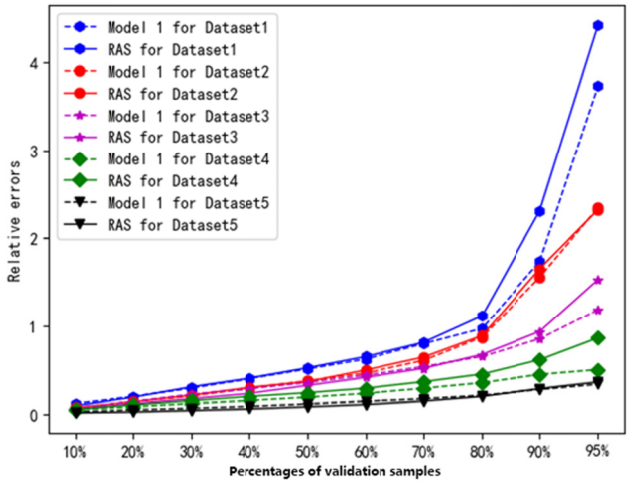

Figure 2. Maximum relative errors against different percentages of validation samples for both our MNR model and RAS in five real-world datasets

3.3. Sensitivity analysis

This section presents the sensitivity analysis results of the effects of different parameter settings, described in section 2.4, on the performance of our MNR model.

Table 4 shows the statistical results of relative errors for the test samples from four different I-O value ranges, which shows the maximum relative errors for different percentages of test samples for the 4 datasets with different value ranges. It can be found from this table that different value ranges produce very similar results. For the 4 value ranges, the relative errors of 95% of test samples are not greater than 15.74%, 15.79%, 15.77% and 15.79%, respectively. The average relative errors of all test samples in the 4 test datasets are 7.53%, 7.56%, 7.55% and 7.56%, respectively.

Table 4. Results of maximum relative errors of model outputs for different data percentages in test samples under different I-O value ranges

Percentages	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
$[0, 10^4]$	1.07%	2.15%	3.3%	4.51%	5.87%	7.33%	9.11%	11.62 %	15.74 %	20.05 %
$[0, 10^6]$	1.07%	2.17%	3.3%	4.53%	5.84%	7.37%	9.2%	11.66 %	15.79 %	20.19 %
$[0, 10^8]$	1.06%	2.14%	3.28%	4.52%	5.84%	7.34%	9.19%	11.68 %	15.77 %	20.28 %
$[10^4, 10^7]$	1.06%	2.17%	3.3%	4.5%	5.82%	7.33%	9.17%	11.66 %	15.79 %	20.3%

Figure 3 shows the maximum relative errors of the MNR model for test samples from different growth rate ranges. It can be found that the accuracy of the MNR model decreased as the growth rate range increased, with the model under-fitting when the growth rate range was large. For the largest growth rate range ([−70%, 300%]), the maximum relative error was around 37% and 15% of test samples showed relative errors >20%. These results indicate that the value of x_{ij}^1 in M^1 is less related to x_{ij}^0 and is affected by other elements in the original I-O table when the growth rate range is large, consistent with the results from section 3.2. However, growth rate ranges larger than [−50%, 150%] seldom occur in real-world applications.

Figure 4 shows the statistical results of the relative errors. It can be seen that the MNR model performs well across test samples of different I-O data dimensions. The relative errors of 90% test samples are less than 10%, and the maximum relative error is less than 17.5%. We can thus conclude that the dimension of the I-O table (i.e., the number of sectors) has a negligible effect on the MNR model. That is, our MNR model can be applied to I-O tables with various dimensions.

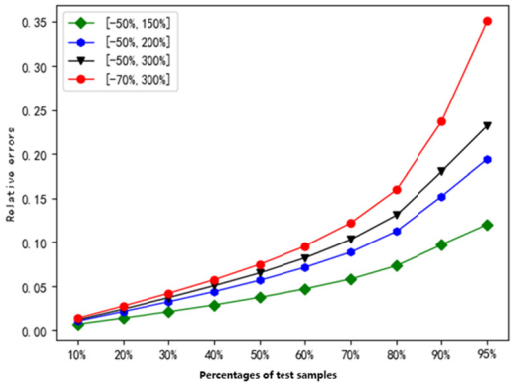

Figure 3. Maximum relative errors of the MNR model against different percentages of test samples across growth rate ranges

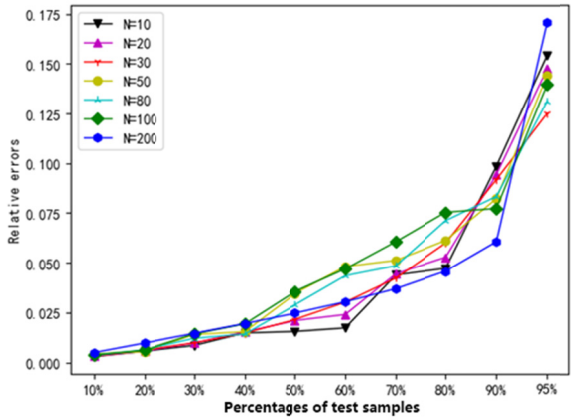

Figure 4. Maximum relative errors of the MNR model against different percentages of test samples for different I-O data dimensions

4. Discussion

4.1 Advantages of the MNR model

We discovered a multivariate nonlinear regression model hidden in the RAS method proposed by Stone and Brown (1962). The proposed model is able to effectively estimate the value of any specific element in the target I-O table based on its corresponding value in the base year and the I-O constraints in the target year. Like the RAS method, the proposed model guarantees a balanced I-O table

once all the elements of the target IO table are estimated, according the results presented in Section 4.

The RAS method and its variants, such as MRAS (Lecomber (1975), GRAS (Junius and Oosterhaven (2003) and KRAS (Lenzen et al. (2009), rely a complete original I-O table to calculate the I-O values in the target table, and their calculations are usually based on a nonlinear optimization or iterative procedure (Bacharach, 1970). Compared to these existing models, the proposed model has a clear and direct mathematical relationship, which is easier to code and has a better computational performance because this model does not undergo the iterative process.

4.2 Potential uses of the MNR model

For the most majority of growth rate ranges (e.g., [-50%, 150%]) in real-world I-O applications, the proposed model performs similarly to the classical RAS method, for any I-O table dimension and any I-O value ranges. The model can thus be used as an effective alternative to the RAS method and applied widely in real-world applications of I-O table rebalancing due to its ease-of-use. Relevant applications include projecting the volumes of freight movement, estimating population movement and capital flow between certain regions or economic sectors, and so on.

It is not easy to obtain the complete original I-O tables in many real-world applications, especially for the construction of international and interregional I-O tables. Unfortunately, it is hard to use traditional I-O analysis methods in these cases with only partial original data (Lenzen et al., 2009). The proposed model can handle these cases effectively. Our MNR model is capable of effectively estimating the element value x_{ij}^1 of the target I-O table based on only its corresponding element x_{ij}^0 in the original table along with the input and output growth rates of each sector. Compared with existing RAS methods and variants, the proposed model should be the first choice for estimating the elements in the target table when only partial elements are available (concerned) in the original (target) I-O table.

4.3 Limitations and future research

As presented in sub-section 3.3, the growth rate ranges of I-O values could have effects on approximating the performance of the proposed model to the RAS method. The proposed model cannot fit the RAS method well when the three growth rate ranges wider than [-50%, 150%] are considered, although these ranges seldom occur in real-world I-O applications. The reason is that with the increase of growth rate range, the element value of the target I-O table modelled by the classical RAS method will be related to more elements instead of only the

corresponding element value in the original I-O table. Even so, the real data-based validation presented in section 3.2 has showed that the MNR model is able to provide the estimation performance close to those generated by the RAS method. Our future research will use I-O tables from other commonly used I-O databases (e.g., EXIOBASE and WIOD) to further examine the performance of the MNR model, and also use this model in more real-world I-O applications.

We obtain the MNR model by fitting the I-O relationship of the traditional RAS method. However, other RAS variants may also contain similar simple and effective mathematical relationships. A prospective future research direction is to examine the I-O relationships of other RAS variants (e.g., GRAS) and explore simple and effective alternatives for these models. Theoretical proof of the simple relationship formulated by the MNR model in the future research is also required, since the MNR model is constructed and validated based on extensive numerical experiments.

5. Conclusions

This paper contributes the literature by developing the first method for the element estimates of incomplete I-O tables. We proposed an effective multivariate nonlinear regression model (the MNR model) for updating I-O tables with only partial information by investigating the mathematical relationships within the RAS method based on extensive numerical experiments. To the best of our knowledge, this model is the first for estimating I-O values in I-O analysis when only incomplete original I-O table is available. Specifically, given reasonable growth rate ranges (e.g., [-50%, 150%]) as well as the input growth rate and the output growth rate of each sector, the element value x_{ij}^1 in the target I-O table M^1 is highly dependent on its corresponding element x_{ij}^0 but almost independent of the other element values in the original I-O table M^0. We also found that the growth rate ranges of I-O table have large effects on the estimation performance of our MNR model, which is similar to the RAS method. As the growth rate range increases, the estimation performance of the model gradually decreases, in a way that is comparable to the results of the conventional RAS Method. The value ranges and the dimension of I-O data have few effects on the estimation performance of the MNR model. The model can be used as an effective alternative to the RAS method and applied widely in real-world I-O table rebalancing due to its ease-of-use and comparable performance with the RAS method.

Data Availability Statement

The data that support the findings of this study are openly available at https://www.dropbox.com/s/ab8pgas9o5v5vty/SupplementaryInfo.zip?dl=0.

REFERENCES

[1]**Bacharach, M. (1970),** ***Biproportional Matrices and Input-output Change*****.** *Cambridge University Press*;

[2] **Chen, G., M. Hadjikakou and T. Wiedmann (2017),** ***Urban Carbon Transformations: Unravelling Spatial and Inter-Sectoral Linkages for Key City Industries Based on Multi-Region Input–Output Analysis*****.** *Journal of Cleaner Production*, 163:224-240;

[3] **Clark, P.G. and H. Chenery (1959),** ***Interindustry Economics*****.** *John Wiley & Sons*;

[4] **Cui, L.-B., P. Peng and L. Zhu (2015),** ***Embodied Energy, Export Policy Adjustment and China's Sustainable Development: A Multi-Regional Input-Output Analysis*****.** *Energy*, 82:457-467;

[5] **Haddad, E.A., F. Perobelli, I. Araújo & K. Bugarin (2021),** ***Structural Propagation of Pandemic Shocks: An Input–Output Analysis of the Economic Costs of COVID-19*****.** *Spatial Economic Analysis*, 16(3): 252-270;

[6] **Hwang, T.S. (2014),** ***Freight Demand Modeling and Logistics Planning for Assessment of Freight Systems' Environmental Impacts*****.** *Doctoral dissertation, University of Illinois at Urbana-Champaign*;

[7] **Jackson, R. and A. Murray (2004), A*lternative Input-output Matrix Updating Formulations*****.** *Economic Systems Research*, 16: 135-148;

[8] **Jennrich, R. and P. Sampson, (1968),** ***Application of Stepwise Regression to Non-linear Estimation*****.** *Technometrics*, 10:63-72;

[9] **Junius, T. and J. Oosterhaven (2003),** ***The Solution of Updating or Regionalizing a Matrix with Both Positive and Negative Entries*****.** *Economic Systems Research*, 15:87-96;

[10] **Lahr, M. and L. De Mesnard (2004),** ***Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis*****.** *Economic Systems Research*, 16:115-134;

[11] **Lecomber, J.R.C. (1975),** ***A Critique of Methods of Adjusting, Updating and Projecting Matrices. In: R.I.G. Allen & W.F*****.** *Gossling (eds) Estimating and Projecting Input-Output Coefficients*;

[12] **Lenzen, M., B. Gallego and R. Wood (2009),** ***Matrix Balancing under Conflicting Information*****.** *Economic Systems Research*, 21:23-44;

[13] **Liu, J., G. Ou and Z. Guo (2018),** ***Regional Freight Volume Forecasting with Incomplete Data of Origin/Destination Freight Volumes.*** *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE,* 839-843;

[14] **Mahajan, S., J. Beutel, S. Guerrero, et al. (2018),** ***Handbook on Supply, Use and Input-Output Tables with Extensions and Applications;*** *United Nations Publications, Department of Economic and Social Affairs, Statistics Division;*

[15] **Miller, R.E. and P.D. Blair, (2009),** ***Input-output Analysis: Foundations and Extensions*****.** *Cambridge university press*;

[16] **Oosterhaven, J. (2005),** ***GRAS versus Minimizing Absolute and Squared Differences: A Comment****. Economic Systems Research*, 17:327-331;
[17] **Oosterhaven, J. (2019),** ***Rethinking Input-Output Analysis: A Spatial Perspective.*** *Springer International Publishing*;
[18] **Stock, J.H. and M.W. Watson (2015),** ***Introduction to Econometrics.*** *Pearson*;
[19] **Stone, R. and A. Brown (1962),** ***A Programme for Growth, vol. 1: A Computable Model of Economic Growth.*** *London: Chapman and Hall*;
[20] **Valderas-Jaramillo, J.M. and J.M. Rueda-Cantuche (2021),** ***The Multidimensional Nd-GRAS Method: Applications for the Projection of Multiregional Input–Output Frameworks and Valuation Matrices.*** *Papers in Regional Science*, DOI: 10.1111/pirs.12625;
[21] **Weimer-Jehle, W. (2006),** ***Cross-impact Balances: A System-theoretical Approach to Cross-impact Analysis.*** *Technological Forecasting and Social Change* 73:334-361;
[22] **Wolfram, P., T. Wiedmann and M. Diesendorf (2016),** ***Carbon Footprint Scenarios for Renewable Electricity in Australia.*** *Journal of Cleaner Production* 124:236-245.