Factors Influencing the Accuracy of Heterogeneous Input-Output Models

Jialu Sun, Jian Xu, Chuan Li, Qi Su, Zhijian Jiang

(School of Economics and Management, University of Chinese Academy of Sciences, Beijing, 100190, China)

Abstract: This paper is an extension of the Alghero conference paper, continuing the use of initial values for intermediate flow matrices generated through Monte Carlo simulations and employing the TRAS technique to create control references that satisfy the structural characteristics of heterogeneous input-output tables. Using the 2016 non-competitive input-output table for China, distinguishing between domestic and foreign-funded enterprises, as a case study, this research explores various factors influencing the accuracy of heterogeneous input-output models. The findings are as follows: (1) Elements in the Leontief matrix that are more susceptible to error are paradoxically more accurate; the position of elements within the matrix and the specific structure of the heterogeneous model both impact the accuracy of the Leontief matrix elements. (2) The error in the Leontief matrix is linearly reduced to 30% in the direct consumption coefficient matrix, and when calculating results for individual sectors or total outputs, the errors in the Leontief matrix elements offset each other. These two mechanisms jointly ensure the overall accuracy of the heterogeneous input-output model. (3) Reducing the sector resolution increases the error in the Leontief matrix, diminishes the ability of element and sector errors to offset each other, and thus reduces the accuracy of model results such as export value-added.

Keywords: Heterogeneous Input-Output Model; Monte Carlo Simulation; Factors Influencing Accuracy; Sector Resolution

1 Introduction

Heterogeneous input-output models, by disaggregating industry sectors from specific dimensions, can depict differentiated production technologies within sectors and reveal significant conclusions that are obscured by the assumption of sector homogeneity. However, the increase in model dimensions implies a surge in the demand for intermediate flow data. Existing statistical data often falls short of supporting the accurate accounting of such detailed inter-sector flows. Therefore, the construction of heterogeneous input-output models has to rely on proportional assumptions and mathematical optimization methods to fill in missing data. This leads to an inherent paradox in heterogeneous input-output models: while increasing the model's dimensions enriches the information it covers and enhances its ability to explain the differentiated production technologies of real-world enterprises, the vast amount of intermediate flow data required for constructing heterogeneous models mainly comes from estimates. These estimates contain substantial errors compared to actual inter-sector flows, thus affecting the accuracy of the model results.

To assess the accuracy of non-survey compilation techniques for heterogeneous input-output tables and to address the research gap on the accuracy of these models, I proposed a method at last year's conference. This method generates initial values for the intermediate flow matrix based on Monte Carlo simulations and employs the TRAS technique to produce control references that meet the structural characteristics of heterogeneous input-output tables. Using the 2016 non-competitive input-output table for China, distinguishing between domestic and foreign-funded enterprises, adapted from the ICIO-AMNE database as an example, I performed 10,000 simulations under two scenarios where elements of the intermediate flow matrix follow normal and log-normal distributions. I measured the distribution of the Leontief matrix, output multipliers, and export value-added calculated from the simulated tables to explore the transmission pattern of errors from the intermediate flow matrix to the model's final results. The findings showed that the uncertainties in the Leontief matrix, output multipliers, and export value-added exhibit a decreasing trend. Errors in the intermediate flow matrix are gradually neutralized during the input-output analysis process, and the model demonstrates high overall accuracy and strong error self-correction capability.

This paper extends and deepens last year's conference paper by exploring factors influencing the accuracy of heterogeneous input-output models. It identifies which elements in the model are likely to have significant errors, suggesting that targeted surveys of these elements can enhance the model's accuracy at the lowest cost. The study of the mechanisms ensuring the overall accuracy of heterogeneous input-output models explains why the model's sectoral and total results are highly accurate. It also proves the reliability of existing non-survey compilation techniques in studying sectoral and total-level issues and demonstrates the improvement in model result accuracy brought about by increasing model resolution.

2 Literature Review

Although extensive research has explored the accuracy of input-output models from various perspectives by selecting different measurement indicators, few studies have delved into the factors influencing the accuracy of input-output models. This is primarily because the focus of most research has been on whether non-survey compilation methods or updates to input-output tables are reliable, and whether the underlying data for compiling input-output tables are trustworthy. Researchers only need to draw binary conclusions to achieve their research objectives.

With the increasing richness of economic data and the rapid development of computational capabilities, input-output analysis is evolving towards finer granularity. One manifestation of this evolution is the finer classification of sectors in input-output tables, prompting scholars to pay attention to the impact of sector classification granularity (resolution) on analysis accuracy. Some scholars have expressed concerns about constructing more detailed input-output models for two main reasons: first, the uncertainty of intermediate flow matrix elements in input-output tables increases as elements are further subdivided by industry; second, subdividing input-output tables requires additional information, introducing more conflicting economic data and posing challenges to balancing input-output tables, thus increasing compilation costs (Lenzen et al., 2012; Andrew and Peters, 2013). However, more research results suggest that by making reasonable assumptions and fully utilizing economic data as much as possible, refining the classification of sectors and regions in input-output tables can enhance the accuracy of input-output models (Lenzen et al., 2004; Tukker et al., 2009; Su et al., 2010a, 2010b; Lenzen, 2011; Weinzettel et al., 2014; Koning et al., 2015). These seemingly contradictory views actually reflect differences in the perspectives of accuracy evaluation. From the perspective of input-output tables themselves, constrained by the limitations of economic statistics, finer sector classification implies more assumptions and speculative data in the compilation process, inevitably increasing errors in input-output matrices. From the perspective of calculating results of input-output models, finer sector classification enhances the resolution of model results while ensuring high accuracy in detailed results.

However, existing literature has not explored how high-resolution input-output models effectively control the adverse effects of inferred data errors and achieve higher measurement accuracy. Neither has it systematically studied other factors influencing the accuracy of input-output models nor discussed the accuracy of heterogeneous input-output models, which heavily rely on inferred data. Building upon existing research, this paper attempts to make the following contributions: focusing on heterogeneous input-output models, it explores factors influencing accuracy from two dimensions—the elements of input-output matrices and the calculation results of models. It analyzes the mechanism by which sector resolution affects the accuracy of heterogeneous input-output models and investigates the mechanisms ensuring the overall accuracy of heterogeneous input-output models.

3 Factors Affecting the Accuracy of Leontief Matrix Elements

As the core indicator and crucial node of input-output analysis, the accuracy of the Leontief matrix directly affects the accuracy of model results. To ensure the robustness of research results, in our paper from last year, we set two scenarios for Monte Carlo simulations, where the elements of the intermediate flow matrix follow either a normal distribution or a log-normal distribution. In scenario one, we assumed that the elements of the intermediate flow matrix (z_{ij}) follow a normal distribution with a mean equal to the values in the heterogeneous input-output table (z_{ij}^0) and a relative standard deviation of 0.1 times (Rypdal and Winiwarter (2001); Wilting (2012); Moran and Wood (2014)).

$$z_{ij} \sim N(z_{ij}^0, (0.1z_{ij}^0)^2) \tag{1}$$

In scenario two, we assumed that the elements of the intermediate flow matrix follow a log-normal distribution. Referring to the research findings of Lenzen, Wood, and Wiedmann (2010) regarding the distribution of input-output matrix element errors, we set the relative standard deviation of the simulated intermediate flow matrix elements as:

$$\sigma z_{ij} = 0.393\left|z_{ij}^0\right|^{0.698} \tag{2}$$

Figure 1 illustrates the ratio of the coefficient of variation (cv) of Leontief matrix elements to the cv of intermediate flow matrix elements for the two scenarios. It can be observed that this ratio varies greatly among different elements and exhibits characteristics where the ratio of the cv is relatively small for main diagonal elements and relatively large for elements representing different heterogeneity types within the same sector. To explore the factors influencing the relative accuracy of Leontief matrix elements and identify the characteristics of Leontief matrix elements that are more susceptible to errors in intermediate flow matrix elements, this paper uses the ratio of cv values for elements in both scenarios as the dependent variable for regression analysis.

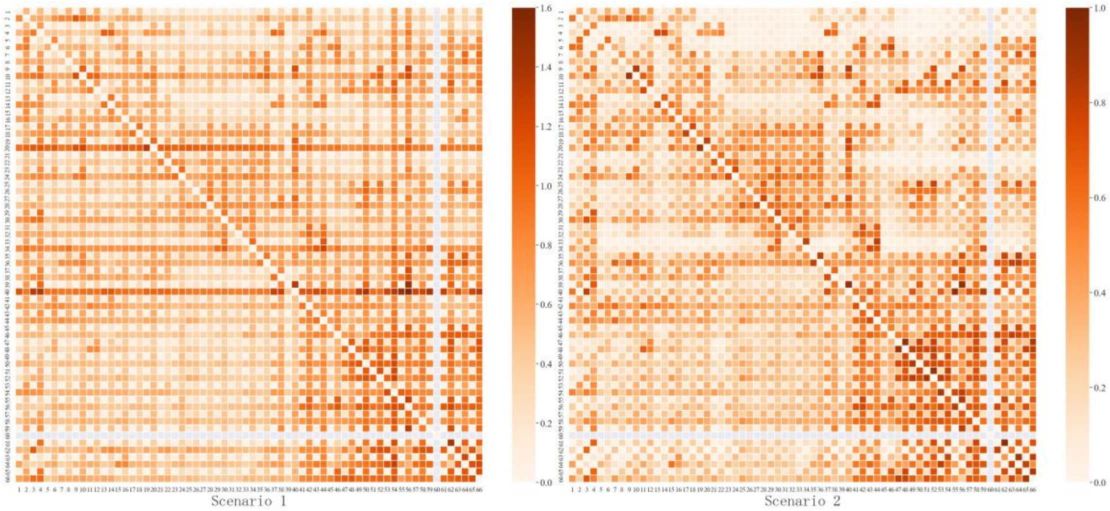

Figure 1: Ratio of Coefficients of Variation of Leontief Matrix Elements to Intermediate Flow Matrix Elements

According to the calculation formula of the Leontief matrix $L=(I-A)^{-1}$, it is known that the accuracy of a single element of the Leontief matrix is directly influenced by errors in all elements of the direct consumption coefficient matrix(A), and errors in a single element of the direct consumption coefficient matrix(A) affect the accuracy of all elements of the Leontief matrix. Therefore, it is reasonable to speculate that the accuracy of Leontief matrix elements is influenced by the following three effects: (1) the effect of the Leontief matrix element l_{ij} being influenced by errors in the corresponding direct consumption coefficient Δa_{ij}; (2) the effect of the Leontief matrix element l_{ij} being influenced by errors in other direct consumption coefficients $\Delta a_{rs(rs\neq ij)}$; (3) the effect of errors in the direct consumption coefficient Δa_{ij} on the other Leontief matrix elements $l_{rs(rs\neq ij)}$.

Based on the research by Sherman and Morrison (1949, 1950) on key coefficients of input-output models, the error in the direct consumption coefficient Δa_{ij} leads to an error in the Leontief matrix element Δl_{rs} as:

$$\Delta l_{rs}=\frac{l_{ri}l_{js}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}} \tag{3}$$

From equation (3), we can calculate the relative influence of the error in the direct consumption coefficient Δa_{ij} on the corresponding Leontief matrix element l_{ij}. This index can serve as a measure of the potential effect (1) influencing the accuracy of Leontief matrix elements, denoted as sel.

$$sel_{ij}=\frac{\Delta l_{ij}}{l_{ij}}=\frac{l_{ii}l_{jj}\Delta a_{ij}}{l_{ij}(1-l_{ji}\Delta a_{ij})} \tag{4}$$

Building upon equation (3) by interchanging r with i and s with j and letting r and s range from 1 to n, we obtain the total error in the Leontief matrix element l_{ij} caused by errors in other direct consumption coefficients $\Delta a_{rs(rs\neq ij)}$.

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\Delta l_{ij(rs,rs\neq ij)}=\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{l_{ir}l_{sj}\Delta a_{rs}}{1-l_{sr}\Delta a_{rs}}-\frac{l_{ii}l_{jj}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}} \tag{5}$$

From equation (5), we obtain the sum of the relative influences of errors in other direct consumption coefficients $\Delta a_{rs(rs\neq ij)}$ on the Leontief matrix element l_{ij}. This index can serve as a measure of the potential effect (2) influencing the accuracy of Leontief matrix elements, denoted as *otl*.

$$otl_{ij}=\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{\Delta l_{ij(rs,rs\neq ij)}}{l_{ij}}=\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{l_{ir}l_{sj}\Delta a_{rs}}{l_{ij}(1-l_{sr}\Delta a_{rs})}-\frac{l_{ii}l_{jj}\Delta a_{ij}}{l_{ij}(1-l_{ji}\Delta a_{ij})} \tag{6}$$

Expanding equation (3), we can derive the total error in other Leontief matrix elements caused by errors in the direct consumption coefficient Δa_{ij}.

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\Delta l_{\mathrm{rs}(rs\neq ij)}=\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{l_{ri}l_{js}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}}-\frac{l_{ii}l_{jj}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}} \tag{7}$$

From equation (7), we obtain the average relative bias of other Leontief matrix elements caused by errors in the direct consumption coefficient Δa_{ij}. This index can serve as a measure of the potential effect (3) influencing the accuracy of Leontief matrix elements, denoted as *ato*.

$$ato_{ij}=\frac{\sum_{r=1}^{n}\sum_{s=1}^{n}\Delta l_{\mathrm{rs}(rs\neq ij)}}{\sum_{r=1}^{n}\sum_{s=1}^{n}l_{\mathrm{rs}(rs\neq ij)}}=\left(\sum_{r=1}^{n}\sum_{s=1}^{n}\frac{l_{ri}l_{js}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}}-\frac{l_{ii}l_{jj}\Delta a_{ij}}{1-l_{ji}\Delta a_{ij}}\right)/\sum_{r=1}^{n}\sum_{s=1}^{n}l_{\mathrm{rs}(rs\neq ij)} \tag{8}$$

Besides three explanatory variables related to key coefficient calculations, empirical results from former paper reveal a strong correlation between the position of elements in the heterogeneous input-output table and accuracy. Therefore, this paper introduces two dummy variables representing the position of elements: whether it is a main diagonal element (*diag*) and whether it corresponds to consumption relationships of the same sector but different heterogeneity types (*m*). Additionally, to examine whether the specific structure of the heterogeneous input-output model affects the relative accuracy of Leontief matrix elements, another dummy variable is introduced to indicate whether it represents inter-firm consumption relationships of heterogeneity (*het*), where 1 signifies 'yes' and 0 signifies 'no'.

On the basis of the aforementioned explanatory variables, it is easy to recognize that Leontief matrix elements themselves (l_{ij}) and elements of the direct consumption coefficient matrix (a_{ij}) are also key variables influencing the accuracy of Leontief matrix elements. However, these two variables are highly correlated with other explanatory variables, which may lead to multicollinearity issues in regression analysis. Hence, they are treated as control variables to ensure the robustness of regression results. They are respectively labeled as *l* and *a*.

In summary, this paper constructs the following regression model to investigate the factors influencing the accuracy of Leontief matrix elements:

$$cv=\beta_0+\beta_1 sel+\beta_2 otl+\beta_3 ato+\beta_4 diag+\beta_5 m+\beta_6 het+\beta Control+\varepsilon \tag{9}$$

In the empirical study, it is assumed that the error of all direct consumption coefficients is 10% of themselves, i.e., $\Delta a_{ij}=0.1a_{ij}(\Delta a_{rs}=0.1a_{rs})$, to calculate the variables sel, otl and ato. The dependent variable cv is the ratio of the coefficients of variation of Leontief matrix elements to the coefficients of variation of the intermediate flow matrix elements for scenarios 1 and 2 in the two regressions, respectively. The final regression results are presented in the table below.

Table 1: Regression Results for Factors Affecting Coefficient of Variation (cv) in Scenario 1

变量	(1)	(2)	(3)	(4)
sel	-13.47***	-11.91***	-12.79***	-11.54***
	(0.62)	(0.62)	(0.62)	(0.61)
otl	-4.59***	-4.28***	-4.49***	-4.06***
	(0.15)	(0.15)	(0.15)	(0.15)
ato	-39.30	250.57***	138.34***	115.71***
	(34.66)	(41.79)	(44.10)	(43.01)
diag	-1.77***	-0.35***	-1.71**	2.27***
	(0.07)	(0.14)	(0.67)	(0.27)
m	0.21***	0.27***	0.26***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)
het	0.21***	0.21***	0.21***	0.21***
	(0.01)	(0.01)	(0.01)	(0.01)
l		-1.31***		-3.94***
		(0.11)		(0.26)
a			-1.34***	5.43***
			(0.21)	(0.50)
Sample Size	4225	4225	4225	4225
Adj. R-squared	0.38	0.40	0.39	0.42

Table 2: Regression Results for Factors Affecting Coefficient of Variation (cv) in Scenario 2

变量	(1)	(2)	(3)	(4)
sel	-2.02***	-1.33***	-1.60***	-1.26***
	(0.30)	(0.30)	(0.30)	(0.30)
otl	-2.05***	-1.91***	-1.99***	-1.87***
	(0.08)	(0.08)	(0.08)	(0.08)
ato	-120.11***	7.42	-11.73	-17.86
	(16.83)	(20.38)	(21.37)	(21.19)
diag	-0.66***	-0.03	-0.62***	0.46***
	(0.03)	(0.07)	(0.03)	(0.14)
m	0.09***	0.11***	0.12***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)
het	0.18***	0.18***	0.18***	0.18***
	(0.00)	(0.00)	(0.00)	(0.00)
l		-0.58***		-1.07***
		(0.05)		(0.13)
a			-0.82***	1.02***
			(0.10)	(0.24)
Sample Size	4225	4225	4225	4225
Adj. R-squared	0.67	0.67	0.67	0.68

Note: *, **, *** denote significance levels of 10%, 5%, and 1%, respectively.

Based on the regression results in Tables 1 and 2, several patterns can be summarized:

First, the coefficients of the variables *sel*, representing the effect of Leontief matrix elements (l_{ij}) influenced by corresponding errors in direct consumption coefficients (Δa_{ij}), and otl, representing the effect of Leontief matrix elements influenced by errors in other direct consumption coefficients ($\Delta a_{rs(rs\neq ij)}$), are consistently negative across all regression results. This indicates that elements of the Leontief matrix that are more susceptible to errors in corresponding direct consumption coefficients (Δa_{ij}) and errors in other direct consumption coefficients ($\Delta a_{rs(rs\neq ij)}$) tend to be more accurate. This seemingly counterintuitive phenomenon demonstrates that the accuracy of Leontief matrix elements relies on the mutual neutralization of different errors in direct consumption coefficients during the matrix inversion process. Elements in the Leontief matrix that more effectively neutralize the effects of errors in direct consumption coefficients exhibit higher accuracy.

Second, the variable *ato*, representing the effect of errors in direct consumption coefficients (Δa_{ij}) on other Leontief matrix elements ($l_{rs(rs\neq ij)}$), shows a positive significant coefficient in the first three regressions of the first scenario but a negative significant coefficient in the first regression of the second scenario. This suggests heterogeneity in the impact of ato on the accuracy of Leontief matrix elements due to differences in the distribution of intermediate flow matrix elements. When the intermediate flow matrix elements approximate a normal distribution, the greater the impact of errors in direct consumption coefficients (Δa_{ij}) on other Leontief matrix elements ($l_{rs(rs\neq ij)}$), the more likely larger errors are to occur in the corresponding Leontief matrix element (l_{ij}).

Third, the dummy variable *het*, representing whether it represents inter-firm heterogeneous consumption relationships, has a consistently positive significant coefficient across all regression results. This indicates that compared to Leontief matrix elements representing homogeneous consumption relationships, those representing heterogeneous inter-firm consumption relationships are more susceptible to errors.

Fourth, except for the regression model that simultaneously includes Leontief matrix elements (*l*) and direct consumption coefficients (*a*), the regression coefficients of the dummy variable *diag*, representing whether it is a diagonal element, are consistently negative in most regression results. This demonstrates that diagonal elements in the Leontief matrix exhibit higher accuracy. The change in the sign of the regression coefficient in model (4) is due to the high correlation between the variable *diag* and the control variables *a* and *l*, which affects the regression coefficient of the variable *diag* due to multicollinearity. The dummy variable *m*, representing whether it corresponds to elements of the same sector but different heterogeneity types, has consistently positive coefficients in both sets of regressions. This suggests that even with the inclusion of the variable *het*, which controls for the structural characteristics of heterogeneous input-output tables, elements in the Leontief matrix representing sectors with completely different heterogeneity types may not only lose the high accuracy of diagonal elements but also be more prone to larger biases compared to other elements.

4 Safeguard Mechanisms for the Overall Accuracy of Heterogeneous Input-output Models

To elucidate the mechanism behind the overall accuracy of heterogeneous input-output models, this study employs three matrix distance metrics: STPE, Theil's U, and MAPE. These metrics measure the distance between intermediate flow matrix elements and the original data calculated from direct consumption coefficient matrices, Leontief matrices, output multipliers, sectoral value-added exports, and the original data distinguishing between domestic and foreign non-competitive input-output tables, at different error levels. This approach aims to explore the reasons for the progressive reduction in error in heterogeneous input-output models.

The calculation methods for the three matrix distance metrics are as follows:

$$STPE = 100\frac{\sum\sum|m_{ij}-m_{ij}^{0}|}{\sum\sum m_{ij}^{0}} \tag{10}$$

$$Theil's\ U = \sqrt{\frac{\sum\sum(m_{ij}-m_{ij}^{0})^2}{\sum\sum(m_{ij}^{0})^2}} \tag{11}$$

$$MAPE = 100\frac{\sum\sum\frac{|m_{ij}-m_{ij}^{0}|}{m_{ij}^{0}}}{n} \tag{12}$$

Here, m_{ij} represents the matrix elements obtained from each simulation, m_{ij}^{0} represents the matrix elements calculated based on the original data of the input-output table distinguishing between domestic and foreign non-competitive sectors, and n represents the number of elements in the matrix or vector.

Figure 2 depicts the relationship between the distances of the direct consumption coefficient matrix A calculated using the three metrics and the Leontief matrix distances. To display the scatter plots based on the three metrics on a single graph, this study calculates the ratio of the maximum distance of the A matrix distance measured by STPE to the maximum distance of the A matrix distance measured by Theil's U and MAPE. Then, the distances of the A matrix and Leontief matrix measured by Theil's U and MAPE are multiplied by these ratios, respectively. This ensures that the results of the three metrics have the same horizontal axis scale and do not alter the relative relationship between the distances of the A matrix and Leontief matrix measured by each metric.

From Figure 2, several patterns can be observed: Firstly, regardless of the distance metric used, the Leontief matrix distance relative to the A matrix distance exhibits a clear linear relationship, diminishing approximately in proportion. Secondly, the reduction ratio of the Leontief matrix distance relative to the A matrix distance varies depending on the metric used. The reduction ratio of the Leontief matrix distance measured by MAPE is approximately 45% of the A matrix distance, while for Theil's U, it is only 24%. Compared to STPE, Theil's U is more sensitive to large absolute distances, whereas MAPE is more sensitive to errors in small coefficients. From this, it can be inferred that larger elements in the Leontief matrix are more accurate, and these larger elements play a more critical role in input-output analysis, thereby ensuring the accuracy of the model.

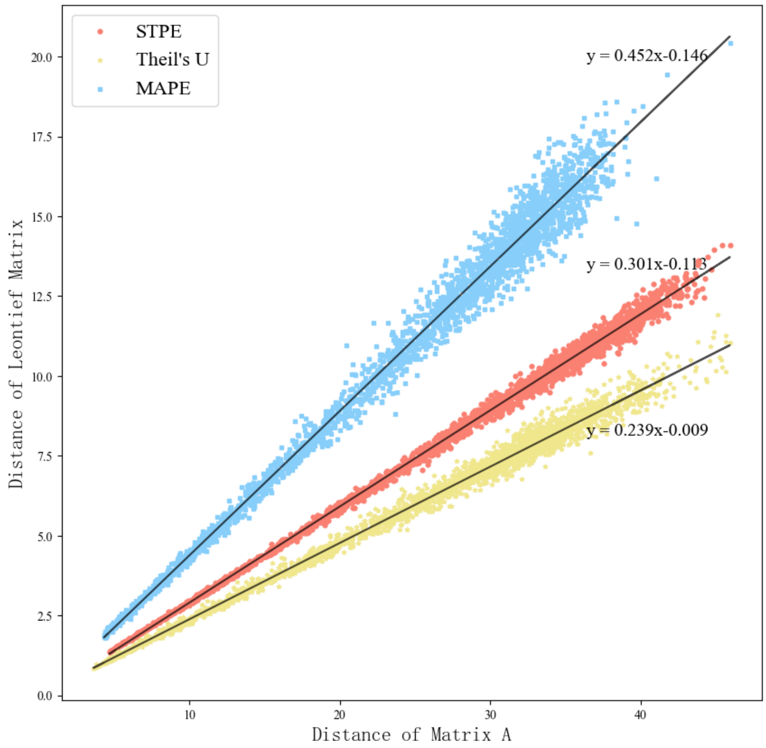

Figure 2: Direct Consumption Coefficient Matrix Distance Versus Leontief Matrix Distance

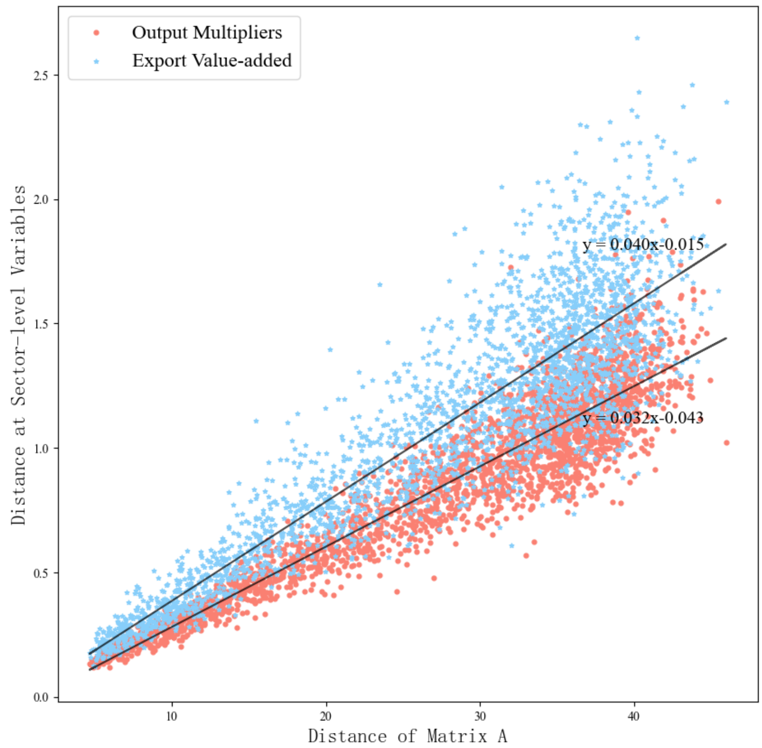

Figure 3: Direct Consumption Coefficient Matrix Distance in Relation to Output Multiplier Distance and Sectoral Export Value-added Distance

In this study, the most commonly used STPE metric was selected to measure the distances of output multipliers and sectoral export value-added. As shown in Figure 3, the distances for output multipliers and sectoral export value-added also exhibit a linear relationship with the A matrix distance. However, their distribution is more dispersed, and the distances are further reduced compared to the Leontief matrix distances. Specifically, the output multiplier distance is approximately 3.2% of the A matrix distance, while the sectoral export value-added distance is about 4.0% of the A matrix distance. This phenomenon indicates a significant error-canceling effect when the object of input-output analysis is aggregated from a matrix dimension to a vector dimension. The errors of the elements in the Leontief matrix cancel out along the rows or columns, resulting in higher accuracy for the output multiplier and sectoral export value-added vectors.

From Figure 4, it can be seen that when the object of input-output analysis is further compressed from vectors to a single value of gross export value-added, the error level is further reduced, but the linear relationship with the A matrix distance is essentially lost. This indicates that as the degree of aggregation of the calculation object increases, the errors in the elements of the Leontief matrix cancel each other out across the matrix dimension, resulting in a higher accuracy of the final calculation results. Even if there are significant errors in the direct consumption coefficient matrix, it is still possible to obtain precise total calculation results. However, this does not imply that a more accurate input-output table is meaningless for total calculations. An accurate intermediate flow matrix can effectively prevent abnormal deviations in total measurement results.

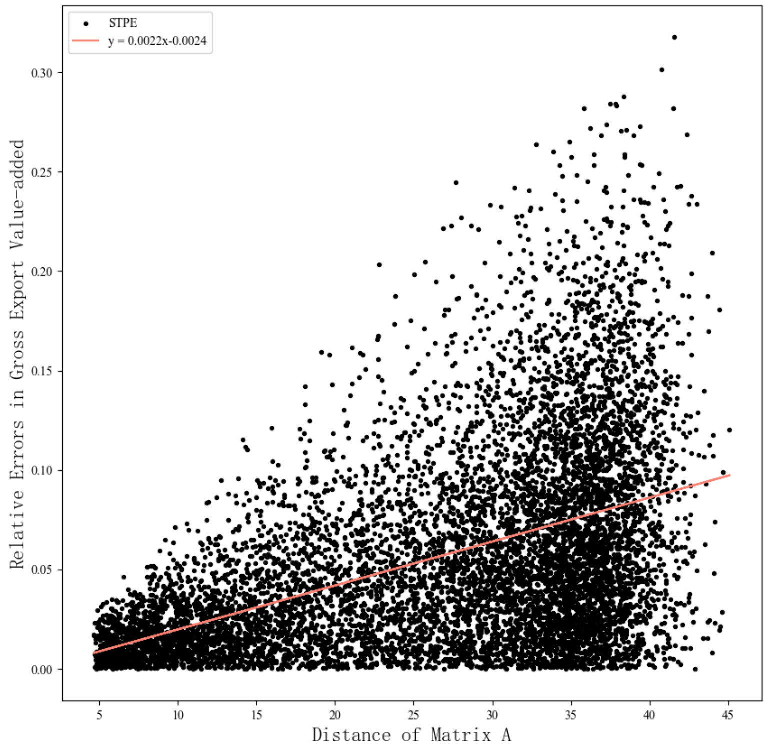

Figure 4: Direct Consumption Coefficient Matrix Distance Versus Relative Errors in Gross Export Value-added

5 The Impact Mechanism of Sectoral Resolution on the Accuracy of Heterogeneous Input-Output Models

As discussed in the literature review section, many scholars have explored the impact of sectoral resolution on the accuracy of input-output models. It is generally believed that making full use of economic data and improving the resolution of input-output models through reasonable assumptions can enhance the accuracy of model results. However, no scholars have explored the mechanism behind this phenomenon. Constructing a heterogeneous input-output model and increasing the number of sectors share the same objective, which is to provide more detailed and precise research results by improving the model's resolution. Therefore, it is necessary to analyze the mechanism of how sectoral resolution affects the accuracy of heterogeneous input-output models.

The sectors in the input-output model can be aggregated using an aggregation matrix G. Suppose the number of sectors before aggregation is K, and the number after aggregation is K^*, then the aggregation matrix G is a $K^* \times K$ binary matrix. Each column of G must have one and only one element equal to 1, and the rest equal to 0, with each row containing at least one element equal to 1. After sector aggregation, the intermediate flow matrix of the input-output model becomes $Z^* = GZG^T$, where G^T is the transpose of G. The value-added vector becomes $V^* = GV$, the output vector $x^* = Gx$, and the export vector $e^* = Ge$. To ensure the generality of the research results, we gradually and randomly aggregate the non-competitive input-output table distinguishing domestic and foreign capital in China to different resolution levels. The aggregation matrix G at each resolution level is randomly generated, meaning that any two or more sectors can be merged regardless of whether the industries they represent are similar in reality. However, the same aggregation matrix G is used for both the domestic and foreign capital parts to maintain the heterogeneous structure. Each aggregated model undergoes a Monte Carlo simulation, where the distribution form of intermediate flow matrix elements remains consistent across different sectoral resolution levels, with elements having a relative standard deviation of 0.1-1 times. The distance between the simulation results and the input-output table results after aggregation is measured using the methods outlined in Chapter 4.

From Figure 5, the following patterns can be observed: (1) The linear relationship between the A matrix distance and the Leontief matrix distance is maintained across different sectoral resolution levels; (2) The method of sector aggregation does not affect the linear relationship between the Leontief matrix distance and the A matrix distance, nor does it affect the slope of the regression line; (3) As the number of sectors decreases (resolution level decreases), the distribution range of A matrix distance and Leontief matrix distance expands, indicating that the error between the simulated A matrix and the Leontief matrix gradually increases, and the accuracy decreases; (4) As the number of sectors decreases, the slope of the regression line for the Leontief matrix distance measured by STPE relative to the A matrix distance slightly decreases, the slope for the distance measured by Theil's U significantly increases, and the slope for the distance measured by MAPE significantly decreases.

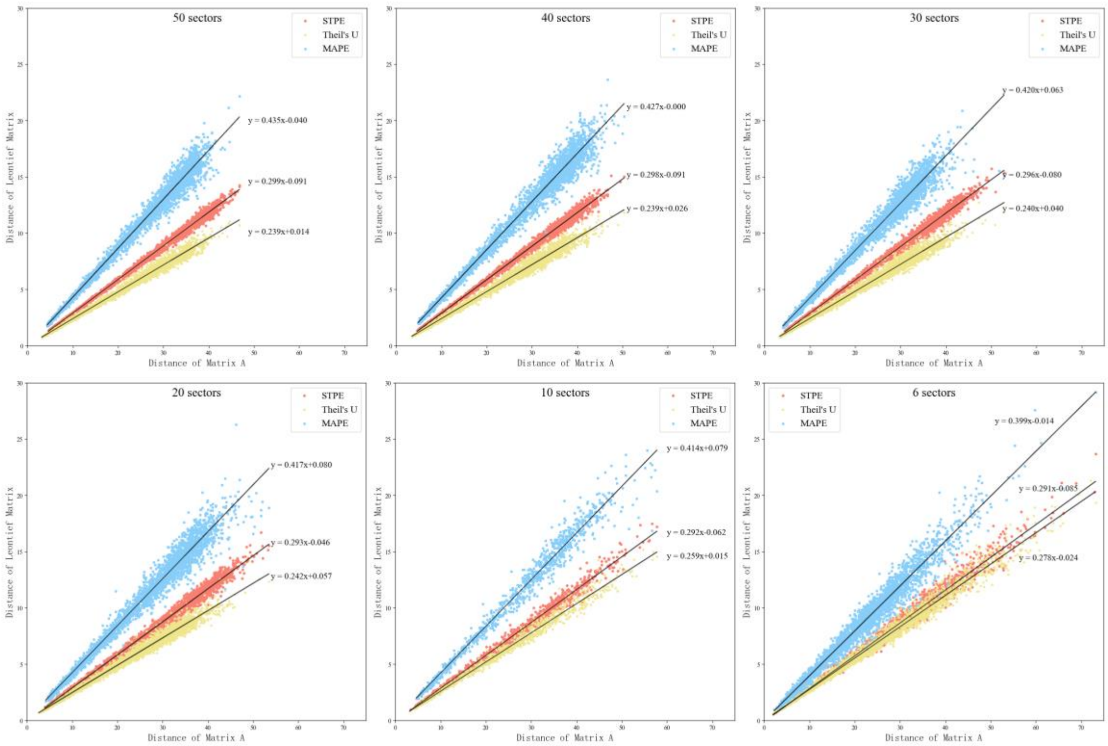

Figure 5: Direct Consumption Coefficient Matrix Distance vs. Leontief Matrix Distance for Different Resolution Levels

These phenomena indicate that the resolution level and the method of sector aggregation almost do not affect the convergence effect of the Leontief matrix on the errors of the direct consumption coefficient matrix or the intermediate flow matrix. The Leontief matrix error measured by STPE will linearly reduce to 30% of the direct consumption coefficient matrix error. The differences in slope changes for different indicators are mainly due to the increase in the elements of the Leontief matrix as the degree of sector aggregation increases, and different indicators have different sensitivities to this change.

The impact mechanism of sectoral resolution on the accuracy of the Leontief matrix lies in the fact that increasing the resolution can reduce the error of the direct consumption coefficient matrix. This, through the proportional relationship between the Leontief matrix error and the direct consumption coefficient matrix error, enhances the accuracy of the Leontief matrix.

From Figure 6, it can be seen that as the sectoral resolution decreases, the slope of the regression line of the output multiplier distance to the A matrix distance gradually decreases, while the slope of the regression line of the sectoral export value-added distance to the A matrix distance gradually increases. This heterogeneity in response to resolution changes is due to the fact that the output multiplier is a direct sum of the elements of the Leontief matrix along the columns, where sector aggregation has a similar effect of neutralizing positive and negative errors. When calculating the sectoral export value-added, the elements of one row of the Leontief matrix need to be multiplied sequentially with the elements of the export column vector and summed. Sector aggregation increases the export weights of the errors in bigger Leontief matrix elements, thereby amplifying the errors in the export value-added. Additionally, the decrease in sectoral resolution also enlarges the distribution range of the sectoral multipliers and export value-added, thereby increasing the errors in the output multipliers, particularly the export value-added.

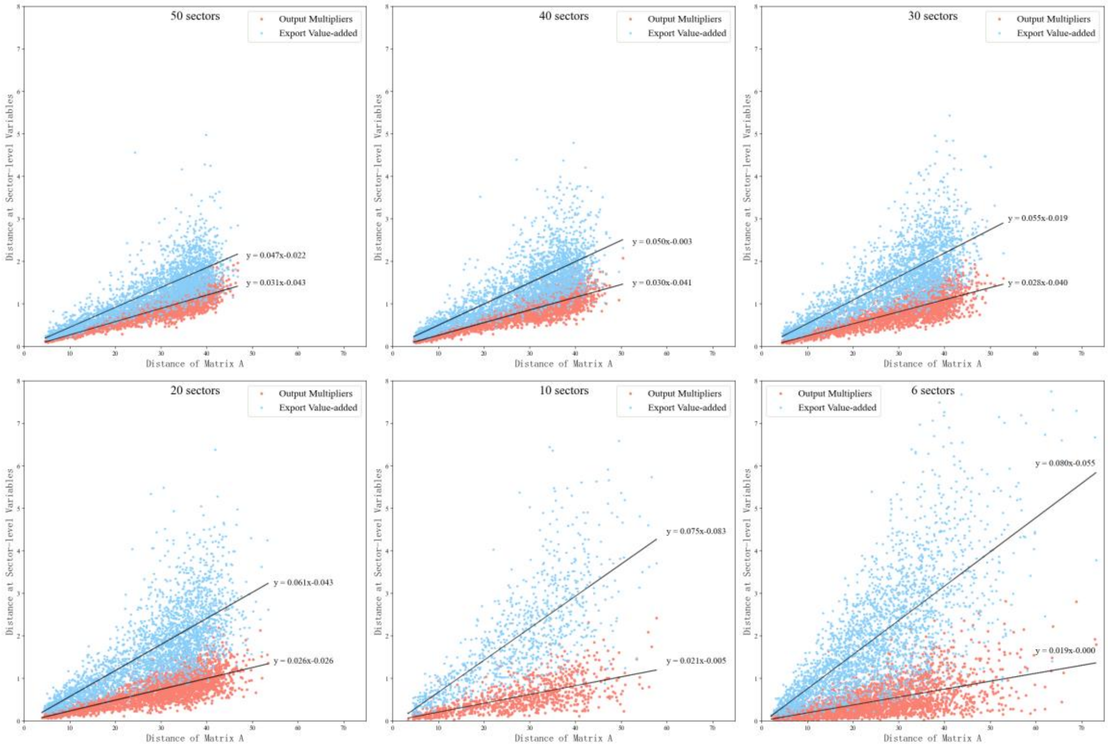

Figure 6: Direct Consumption Coefficient Matrix Distances at Different Resolutions Versus Output Multipliers and Sectoral Export Value-added Distances

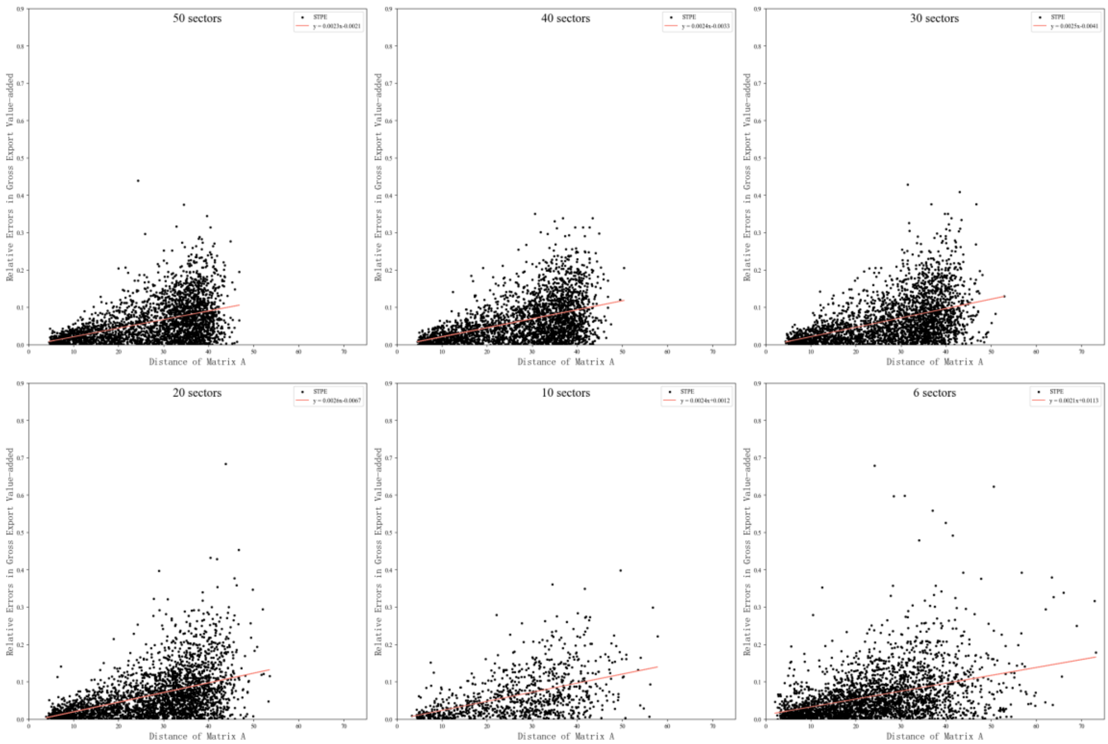

Figure 7: Relative Error of Direct Consumption Coefficient Matrix Distance to Gross Export Value-added at Different Resolutions

As shown in Figure 7, although the relative error of the gross export value-added does not significantly change the regression line slope of the direct consumption coefficient matrix distance as sectoral resolution decreases, the reduction in the number of sectors weakens the error-neutralizing effect between sectors. This leads to an expanded distribution range of the gross export value-added, thereby significantly increasing the probability of large errors when calculating the gross export value-added.

In conclusion, the change in resolution does not significantly affect the ability of Leontief matrix to reduce errors in the direct consumption coefficient matrix or intermediate flow matrix. However, it severely weakens the error-neutralizing ability between elements and sectors, leading to an expanded distribution range of errors. This increases the likelihood of significant biases in the model's calculated results.

6 Conclusion

Building upon the foundation established in the previous paper presented at the Alghero Conference, this study further explores various topics affecting the accuracy of heterogeneous input-output models. These include factors influencing the accuracy of Leontief matrix elements, mechanisms safeguarding the overall accuracy of heterogeneous input-output models, and the impact of sector resolution on model accuracy.

To explore the factors influencing the accuracy of heterogeneous input-output models, this study draws upon research findings on key coefficients in input-output models. It constructs a regression equation with the coefficient of variation of Leontief matrix elements as the dependent variable. The study discovers a phenomenon where elements in the Leontief matrix (l_{ij}) that are more susceptible to errors in direct consumption coefficients (Δa_{rs}) tend to be more accurate. Conversely, elements (l_{ij}) whose corresponding errors in direct consumption coefficients (Δa_{ij}) have a greater impact on other Leontief matrix elements ($l_{rs(rs\neq ij)}$) tend to be less accurate. After controlling for other factors, the position of elements in the Leontief matrix may also influence their accuracy, with elements on the main diagonal being more accurate. Elements representing entirely inter-sectoral relationships within the same sector may exhibit larger errors. Additionally, the heterogeneous structure of input-output models also affects the accuracy of Leontief matrix elements. Elements depicting consumption relationships among homogeneous-type enterprises are more accurate compared to those depicting relationships among heterogeneous-type enterprises.

The overall accuracy of heterogeneous input-output models is safeguarded by two effects: Firstly, the error in the direct consumption coefficient matrix measured by the STPE metric is linearly reduced to 30% when calculating the Leontief matrix. Secondly, when computing sector-level or aggregate data, errors in the Leontief matrix are mutually neutralized, ensuring the overall accuracy of heterogeneous input-output models.

Changes in sector resolution and the specific method of sector merging do not affect the linear reduction relationship between Leontief matrix errors and errors in the direct consumption coefficient matrix. However, reducing sector resolution increases the distribution range of both the direct consumption coefficient matrix and the Leontief matrix, thereby reducing their accuracy.

When calculating practical issues such as sectoral export value, reducing the number of sectors weakens the mutual neutralization ability of errors in each sector, resulting in a wider distribution range of total export value and reduced accuracy of the final model results.

References

[1] Lenzen M., Kanemoto K., Moran D,. et al. Mapping the Structure of the World Economy[J]. Environmental Science &Technology, 2012, 46(15): 8374-8381.

[2] Andrew R., Peters G. A Multi-region Input–output Table Based on the Global Trade Analysis Project Database (GTAP-MRIO)[J]. Economic Systems Research, 2013, 25(1): 99-121.

[3] Lenzen M., Pade L., Munksgaard J. CO_2 Multipliers in Multi-region Input-output Models[J]. Economic Systems Research, 2004, 16(4): 391-412.

[4] Tukker A., Poliakov E., Heijungs R., et al. Towards a Global Multi-regional Environmentally Extended Input–output Database[J]. Ecological Economics, 2009, 68(7): 1928-1937.

[5] Su B., Huang H., Ang B., et al. Input–output Analysis of CO_2 Emissions Embodied in Trade: the Effects of Sector Aggregation[J]. Energy Economics, 2010, 32(1): 166-175.

[6] Su B., Ang B. Input–output Analysis of CO_2 Emissions Embodied in Trade: the Effects of Spatial Aggregation[J]. Ecological Economics, 2010, 70(1): 10-18.

[7] Lenzen M. Aggregation Versus Disaggregation in Input–output Analysis of the Environment[J]. Economic Systems Research, 2011, 23(1): 73-89.

[8] Weinzettel J., Steen-Olsen K., Hertwich E., et al. Ecological Footprint of Nations: Comparison of Process Analysis, and Standard and Hybrid Multiregional Input–output Analysis[J]. Ecological Economics, 2014, 101: 115-126..

[9] Koning A., Bruckner M., Lutter S., et al. Effect of Aggregation and Disaggregation on Embodied Material Use of Products in Input–output Analysis[J]. Ecological Economics, 2015, 116: 289-299.

[10] Rypdal K., Winiwarter W. Uncertainties in Greenhouse Gas Emission Inventories—Evaluation, Comparability and Implications[J]. Environmental Science & Policy, 2001, 4(2-3): 107-116.

[11] Wilting H. Sensitivity and Uncertainty Analysis in MRIO Modelling; Some Empirical Results with Regard to the Dutch Carbon Footprint[J]. Economic Systems Research, 2012, 24(2): 141-171.

[12] Moran D., Wood R. Convergence Between the Eora, WIOD, EXIOBASE, and Open EU's Consumption-based Carbon Accounts[J]. Economic Systems Research, 2014, 26(3): 245-261.

[13] Lenzen M., Wood R., Wiedmann T. Uncertainty Analysis for Multi-region Input–Output Models–A Case Study of the UK's Carbon Footprint[J]. Economic Systems Research, 2010, 22(1): 43-63.