A Study on the Accuracy of Heterogeneous Input-Output Model Based on Monte Carlo Simulation

——An example of Distinguishing Domestic and Foreign Investment in Chinese Non-competitive Input-Output Table

Sun Jialu, Xu Jian, Li chuan, He Meng, Ji Kangxian

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

Abstract: The heterogeneous input-output model has found widespread application in research on trade value-added, energy, and environment in recent years. However, due to the lack of data on trade flows between various types of enterprises when compiling heterogeneous input-output tables, scholars have had to rely on proportionality assumption and optimization methods to construct intermediate flow matrices. Based on a thorough study of existing methods for compiling heterogeneous input-output tables, this study proposes a novel method based on Monte Carlo simulation for generating initial values for intermediate flow matrices, which are then adjusted using the TRAS method to ensure conformity with the structural characteristics of the heterogeneous input-output model. The study then measures the accuracy of the Leontief inverse matrix, output multipliers, and export value-added. By simulating the intermediate flow matrix elements 10,000 times under two scenarios, i.e., normal distribution and lognormal distribution, and varying the mechanism for forming the standard deviation of the intermediate flow matrix elements during the simulation, the study shows that the uncertainty of the intermediate flow matrix, Leontief inverse matrix, output multipliers, and export value-added of the Chinese non-competitive input-output model adapted from ICIO-DF(2016) exhibits a decreasing trend. The results of the study indicate that in the process of establishing a heterogeneous input-output model, as long as the total matrix, such as output, value-added, final demand, imports, and exports, are accurately estimated, even if the intermediate flow matrix obtained from proportionality assumptions and optimization methods is biased, the accuracy of the Leontief inverse matrix, output multipliers, and export value-added can still be maintained and improved in order, and the empirical research results obtained from the model can still maintain a high overall accuracy.

Key works: Heterogeneous input-output model; Distinguish between domestic and foreign investment; Monte Carlo simulation; TRAS; Accuracy

1. Introduction

The Input-Output (IO) model, as a type of general equilibrium model, has played an irreplaceable role in analyzing macroeconomic policies from a multi-sectoral perspective since its inception. However, traditional IO tables only differentiate industries or products, and explore the interrelationships between industry or product sectors, lacking the ability to model and analyze from other perspectives beyond industries or products. As a new research hotspot, the heterogeneous IO model increases the dimensions of the IO model by splitting economic sectors according to certain standards, and has the ability to analyze the differences in production technology, product distribution, and industrial connections of different types of enterprises within the same sector from a more micro perspective.

Based on the above functions, heterogeneous IO models are widely used in research fields such as employment, global value chains, and environmental economics. Scholars have compiled numerous heterogeneous input-output tables from dimensions such as trade patterns, enterprise ownership attributes, and enterprise scale, and have achieved fruitful results. Many economists have noted the uniqueness of processing trade in terms of production technology and value-added composition, they used methods such as quadratic programming to distinguish processing trade from general trade on the basis of a non-competitive input-output table, and constructed a heterogeneous input-output model that distinguishes trade patterns. Economic indicators such as trade added value, trade embodied carbon emissions, and vertical specialization rate were accurately measured, and the problem that the original measurement methods distorted trade value-added and trade status was solved (Chen et al., 2001; Lau et al., 2007; Koopman et al., 2012; Dietzenbacher et al., 2012; Yang et al., 2015; Chen et al., 2019). Due to significant differences in production technology, product distribution, and other aspects among enterprises with different ownership, the heterogeneous input-output model that distinguishes enterprise ownership has received widespread attention. Duan et al. (2013) and Ma et al. (2015) compiled a heterogeneous input-output table regarding both trade pattern and enterprise ownership, and calculated the contributions of enterprises with different ownership types or trade pattern to China's export value-added and national income. Jiang (2015a, 2015b) used the heterogeneous IO table compiled by Ma (2015) to calculate the differences in energy utilization efficiency among enterprises with different ownership types. In addition, some scholars have noted the heterogeneity of the scale of enterprises, Tang et al. (2016) found that Chinese state-owned enterprises and private small and medium-sized enterprises mainly engage in indirect exports, and the domestic part of export value-added is relatively high. Meng et al. (2018) used the heterogeneous IO table compiled by Tang et al. (2016) to measure the carbon emissions of enterprises of different sizes. Zhang Junrong et al. (2021) constructed a heterogeneous IO model that distinguishes the scale of enterprises, and used the structural decomposition analysis (SDA) method to analyze the employment-promoting effect and influencing factors of enterprises of different scales in China's internal and external circulation.

In recent years, heterogeneous input-output models have been combined with multi-regional input-output models (MRIO), Duan and (2018) pointed out that due to the uneven distribution of China's processing trade among regions, the heterogeneity of trade modes should be considered in Chinese regional input-output table, otherwise the share of local value-added in China's exports in each region will be overestimated to varying degrees. Based on the OECD's Global Input-Output Database (ICIO-DF) that distinguishes enterprise ownership from 2005 to 2016, Cadestin et al. (2018) found that multinational corporations have significant direct and indirect impacts on the output, value added, international trade, and employment of their subsidiaries in host countries, and the extent of their influence depends on the level of integration between the subsidiaries of multinational corporations and the host country's economy. Meng et al. (2020) found significant differences in the smile curves of domestic and foreign-funded information and communication technology (ICT) enterprises in the global value chain between China and the United States, proving that enterprises in the two countries participate in global value chain with different technological advantages, reflecting the full use of the relative competitive advantages of the two countries. Zhang et al. (2020) tracked the carbon footprint of foreign subsidiaries of multinational enterprises and found that due to a decrease in carbon intensity, the total amount of carbon emissions transferred through investment reached its peak in 2011. The carbon footprint of multinational companies from developed countries has decreased, while the carbon transfer from mainland China has significantly increased. Duan and Jiang (2021) simulated carbon emissions under the scenario of anti-globalization, and found that multinational companies have caused pollution haven effects in both high-income and low-income economies. The rise of anti-globalization may temporarily suppress the overall increase in global carbon emissions.

Due to more detailed classification within the sectors, the construction of the heterogeneous input-output model mentioned above lacks sufficient economic data support, and has to use proportionality assumption and optimization methods such as RAS to fill in missing data. However, a large amount of inferred data leads to errors in heterogeneous input-output models. Existing research has not considered the issue of model accuracy, and it is not known how errors in the original model will affect the research results. Based on a systematic study of the conventional construction method of heterogeneous input-output models, this paper identifies the main sources of errors in the model. Monte Carlo simulation is conducted based on the distribution law of elements in the input-output table, and the simulated table is adjusted using TRAS to have the same structure as the original heterogeneous input-output table. On this basis, the propagation law of errors in the heterogeneous input-output model is analyzed in depth, and potential biases in research results are measured. Suggestions are proposed to improve the accuracy of heterogeneous input-output models.

The structure of the remaining parts of this paper is as follows: Part 2 systematically reviews the conventional construction method for heterogeneous input-output models, identifies potential sources of errors. Part 3 explains the difficulties in studying the accuracy of heterogeneous input-output models and introduces methods based on Monte Carlo simulation and TRAS to study model accuracy. Part 4 demonstrates the selection and processing of research materials, and introduces the core calculation formulas of input-output analysis. Part 5 measures the errors of the Leontief inverse matrix, output multipliers, and export value-added, and explores the propagation law of errors in the actual application process of the model. Part 6 summarizes the main conclusions of this paper and proposes suggestions to improve the accuracy of heterogeneous input-output models.

2. The conventional construction method of heterogeneous input-output model

The essence of heterogeneous input-output model is to add additional dimensions to the basic input-output model that distinguishes between industrial or product sectors. The core of this model is to split the intermediate flow matrix, final demand matrix, value-added matrix, and output vector of the basic input-output table based on the division criteria of this additional dimension.

The conventional construction method of heterogeneous input-output model can be roughly divided into three steps. Firstly, based on customs statistics, statistical yearbooks and other databases to obtain or use gravity models and other methods to estimate the total quantity data such as the final demand matrix, value-added matrix and output vector after splitting. Secondly, obtain the initial values of the elements in the intermediate flow matrix. Due to the lack of sufficient detailed data for splitting the intermediate flow matrix, it is necessary to rely on proportionality assumption in the compilation process. Based on the total quantity matrices such as output, value-added, final demand, import and export, the proportion of different types of enterprises' output in sector i (marked as ε_i^{c1}) and the proportion of intermediate inputs used in sector j (marked as δ_j^{c2}) are calculated. Then, the intermediate flow matrix element z_{ij} in the original input-output table is multiplied by these two ratios to obtain the initial value of the intermediate flow matrix element in the segmented heterogeneous input-output table. For example, the consumption of $c1$ type enterprises in sector i by $c2$ type enterprises in sector j:

$$z_{ij}^{c1c2} = \varepsilon_i^{c1} \delta_j^{c2} z_{ij}$$

Finally, according to methods such as RAS and quadratic optimization, the initial values of the intermediate flow matrix are coordinated to satisfy the balance structure of input-output model, forming a complete heterogeneous input-output model.

From the above conventional construction steps of heterogeneous input-output model, it can be seen that the aggregate matrices such as output, value-added, final demand, import and export in the model are obtained or estimated according to economic statistical data from various sources, and thus have a relatively high level of accuracy. The estimation of intermediate flow matrix relies on proportionality assumption and optimization methods such as RAS, which is the main source of error in heterogeneous input-output models. Proportionality assumption implies that different types of enterprises within the same sector use intermediate inputs from different types of enterprises in other sectors based only on the proportion of their output, without any preference for intermediate goods from different types of enterprises. This is a very strong assumption that does not align with reality. Tang (2016) found that foreign-owned enterprises tend to use imported intermediate inputs rather than domestically produced ones, and enterprises with different ownership exhibit significant preference differences in the use of intermediate inputs. In addition, optimization methods such as RAS only adjust the intermediate flow matrix to meet the basic balance relationship of input-output model, and the influence of this method on the accuracy of the input-output model needs further study. Furthermore, as the core of input-output analysis, the Leontief inverse matrix relies on the output vector and the intermediate flow matrix in its calculation. The value-added and final demand matrices are only used as satellite matrices multiplied by the Leontief inverse matrix when studying specific issues. Their impact on the overall accuracy of the input-output model is lower than that of the Leontief inverse matrix. Therefore, the key to studying the accuracy of heterogeneous input-output models is to systematically analyze the accuracy issues caused by the assumption of proportionality and balance optimization methods used in the construction of the intermediate flow matrix, and to analyze the impact of errors in intermediate flow matrix elements on model accuracy.

3. Methods for measuring the accuracy of heterogeneous input-output model

3.1. The main method for measuring the accuracy of input-output model

Since the inception of input-output analysis, research on its accuracy has received widespread attention from scholars. This is because information on measuring uncertainty helps decision-makers understand the assumptions and limitations behind the data, enabling them to make sound decisions in a well-informed manner (Lenzen, 2012). In addition, research on model accuracy helps to improve the methods of compiling input-output tables, achieving the goal of improving model accuracy under constraints of funding and personnel.

The use of non-survey and partial survey methods to update and compile input-output tables has facilitated the development of research on measuring model accuracy. To measure the accuracy of I-O tables prepared using the RAS method and other methods, scholars have undertaken a comparative analysis of non-survey or partial survey I-O tables with those prepared using survey methods. In this context, they have constructed indicators to measure matrix differences which directly measure the comprehensive error of input-output matrix (Leontief, 1966; Theil, 1971; Sawyer and Miller, 1983; Miller and Blair, 1982, 1983; Lahr, 2001). However, these indicators have different applicable scenarios, and the calculation results of different indicators may conflict with each other regarding the accuracy of the same model. Improper use of these indicators may lead to erroneous conclusions. Researchers such as Knudsen and Fotheringham (1986), Lahr (1998), Gallego and Lenzen (2009), and Wiebe and Lenzen (2016) have systematically compared various indicators for measuring differences in matrices. They found that weighted standardized total percentage error (WSTPE) and standardized weighted absolute difference (SWAD) are widely used due to their simplicity and easy of understanding. Meanwhile, metrics such as mean absolute difference, root mean squared error, correlation coefficient and its inverse, and R-squared are symmetrical and suitable for matrices with many zero elements after standardization.

Multi-regional input-output models (MRIO) have extremely high requirements for interregional trade data and rely on many assumptions during their construction. There are no highly credible MRIO tables prepared using survey methods to serve as a benchmark for accuracy research. In order to address the lack of a reference basis, scholars compare the accuracy of MRIO models through mutual comparison with world input-output tables such as Eora, WIOD, EXIOBASE, and GTAP. Geschke et al. (2014) created new GMRIO tables by combining raw data, constraints, and integration methods from different GMRIO tables and compared them with the original tables. They found that the quality of the basic data plays a crucial role in determining the quality of the GMRIO table. The integration method used in Eora model can produce higher quality GMRIO tables based on the raw data from other GMRIO systems.

In addition to directly measuring the error of the input-output table matrix elements, another important approach to study the accuracy of the model is to analyze the process of error propagation from the intermediate flow matrix to the Leontief inverse matrix and multiplier, in order to measure the bias in the final calculation results of the model. The majority of empirical studies suggest that so long as the errors in the intermediate flow matrix are kept within an acceptable range, the significant presence of elements which overestimate or underestimate the actual flow of goods and services between sectors, along with the model's intricate interrelationships, lead to a net neutralization of errors in the final results (Jensen, 1980; Peters, 2007; Yamakawa and Peters, 2009; Lenzen and Peters, 2010). Therefore, from initial intermediate flow matrix to Leontief inverse matrix and then to multipliers and export value-added, the errors gradually decrease. These research conclusions are consistent with the concept of overall accuracy proposed by Lenzen et al. (2013), in which the accuracy of individual elements in input-output analysis is insignificant. As long as the overall analysis results produced by the model can reflect the real economic situation and meet the decision-makers' purposes, the model's accuracy can be accepted.

In the study of the accuracy of heterogeneous input-output models, the lack of a survey-based heterogeneous table as a reference benchmark, coupled with only one table for each dimension of heterogeneity, renders it impossible to employ the approach of assessing GMRIO model accuracy by cross-referencing multiple input-output tables depicting the same dimension of heterogeneity. As such, the key to studying the accuracy of heterogeneous input-output models lies in the construction of a reference basis.

3.2. Generating intermediate flow matrix through Monte Carlo simulation

Due to the specific row-column balance structure and the fact that the elements of the intermediate flow matrix in input-output models are not deterministic parameters but rather stochastic variables that follow a stable distribution pattern (Torres-González and Yang, 2019), scholars have used Monte Carlo simulation to generate a large number of input-output matrices and satellite accounts that satisfy the above characteristics. These matrices and accounts cover all possible forms that the actual input-output models may present, and serve as mutual reference benchmarks for studying the accuracy of input-output models.

Lenzen et al. (2010) investigated the error of carbon emissions caused by consumption through 5000 Monte Carlo simulations. They found that the overall consumption-based carbon emissions were accurate, but there were significant errors in sectoral consumption-based carbon emissions. Mroan and Wood (2014) conducted a sensitivity analysis of the carbon footprint estimated by different MRIO models using Monte Carlo simulations. They found that the accuracy of national or regional carbon footprint estimates is positively correlated with the size of the economy, and there is temporal continuity in the estimates. Dietzenbacher (2006) used Monte Carlo simulation to study the distribution of output multipliers when the elements of the intermediate flow matrix are randomly distributed. It was found that the output multipliers exhibited statistically significant positive bias, but the impact was negligible. Lenzen (2011) used Monte Carlo simulation method to demonstrate that the use of small-scale survey data can significantly improve the accuracy of MRIO tables.

Based on previous research, this paper employs Monte Carlo simulation to generate the intermediate flows matrix, capturing the possible errors in the heterogeneous input-output model. Then, the Leontief inverse matrix, output multipliers, and export value-added are calculated for each simulation model to measure their differences and study the impact of errors on the model accuracy as well as the propagation mechanism from the intermediate flows matrix to the final results.

Monte Carlo simulation requires a clear understanding of the distribution of intermediate flow matrix elements. Wilting (2012) suggests that intermediate flow matrix elements are approximately normally distributed, while Lenzen et al. (2010) argue that intermediate flow matrix elements follow a log-normal distribution, with larger values being relatively accurate. Based on these studies, to ensure the robustness of our results, we constructed two simulation scenarios. Scenario one assumes that intermediate flow matrix elements follow a normal distribution with the values in the heterogeneous input-output table as the mean, while scenario two assumes a log-normal distribution for intermediate flow matrix elements.

In previous studies, researchers such as Rypdal and Winiwarter (2001), Wilting (2012), and Moran et al. (2014) assumed that the relative standard deviation of intermediate matrix elements is 0.1. In our study, we adopt this assumption in scenario one. In scenario two, we refer to the research results of Lenzen et al. (2010) and assume the relative standard deviation of intermediate matrix elements as:

$$\sigma z_{ij}^0/\left|z_{ij}^0\right| = 0.393\left|z_{ij}^0\right|^{-0.302}$$

The reliability of research results can be significantly influenced by the number of Monte Carlo simulations performed. Dietzenbacher (2006) suggests that increasing the number of simulations beyond 1000 times will not substantially alter the research findings. As the intermediate flow matrix simulated in this study is relatively large, to ensure reliable research, the empirical research section of this paper includes over 1000 simulations.

3.3. Adjusting the structure of intermediate flow matrix with TRAS

Due to the fact that the total intermediate demand and intermediate input of each sector in a heterogeneous input-output model can be calculated or estimated from economic statistics data with high accuracy, and the intermediate flow matrix elements in a heterogeneous input-output model are obtained by splitting the intermediate flow matrix elements in the basic model. Therefore, the generated intermediate flow matrix in simulation should meet the following conditions: (1) the row sum of intermediate flow matrix elements equals the determined intermediate input of the sectors; (2) the column sum of intermediate flow matrix elements equals the determined intermediate demand of the sectors; (3) the intermediate flow matrix elements in the basic model equal to the sum of multiple intermediate flow matrix elements in the split heterogeneous model; and (4) there are no negative intermediate flow matrix elements in a heterogeneous input-output model. This should be taken into consideration in the empirical analysis of this study.

However, due to the randomness of Monte Carlo simulation, the intermediate input-output matrix generated by simulation cannot satisfy the above requirements. In this paper, TRAS method is used for adjustment. This paper takes the total intermediate input and total intermediate demand of each sector as row and column constraints, and takes the intermediate flow matrix elements of the original input-output model before splitting as constraints for the corresponding intermediate flow matrix element blocks after splitting. The detailed process of the TRAS method refers to Gilchrist and St Louis (1999). The above adjustments not only realize the randomness of the intermediate flow matrix generated by simulation, maintain the structural characteristics of the heterogeneous input-output model, but also highly restore the compilation process of the heterogeneous input-output model, effectively improving the reliability of the research results.

4. Selection and processing of research materials

4.1. Selection of research materials

This study utilizes the 2016 Input-Output Framework for Foreign and Domestic Firms (ICIO-DF) released by OECD as research materials. The ICIO-DF distinguishes 34 sectors of 60 countries or regions (36 OECD countries, 23 non-OECD countries, and 1 other region in the world) into domestic or foreign firms, constructing a large heterogeneous input-output table that contains a 4080×4080 matrix of intermediate flows. This dataset has been widely used by scholars in empirical research.

The construction of ICIO-DF table generally follows the conventional construction method of heterogeneous input-output model. The detailed process can be found in Cadestin's (2018) report, which can be divided into three main steps as follows.

The first step involves constructing matrices that distinguish between countries or regions, sectors, and ownership of output, value-added, and trade (exports and imports). The initial values of the matrices are obtained through the Activities of Multinational Enterprises (AMNEs), the OECD Trade by Enterprise Characteristics (TEC) database and data from National Statistics Offices, missing values are estimated through FDI data or statistical methods such as gravity models. The initial values are then adjusted to match the ICIO table provided by the World Input-Output Database (WIOD). The second step involves splitting the intermediate flow matrix and final demand matrix of the original ICIO table based on the assumption of proportionality, as shown in Figure 1. The method for splitting the intermediate flow matrix can be summarized as follows: based on the proportions of output produced by domestic or foreign-owned firms in sector i (marked as ε_i^c) and the proportions of intermediate input used by sector j (marked as δ_j^c), the original intermediate flow matrix element z_{ij} is split to obtain the initial values of the intermediate flow matrix elements z_{ij}^{cc} in the ICIO-DF table. For example, the initial value of the direct consumption of domestic-owned firms in sector i by foreign-owned firms in sector j is:

$$z_{ij}^{df} = \varepsilon_i^d \delta_j^f z_{ij}$$

The third step utilizes a quadratic optimization method to achieve a balanced heterogeneous input-output table that distinguishes between domestic and foreign investment.

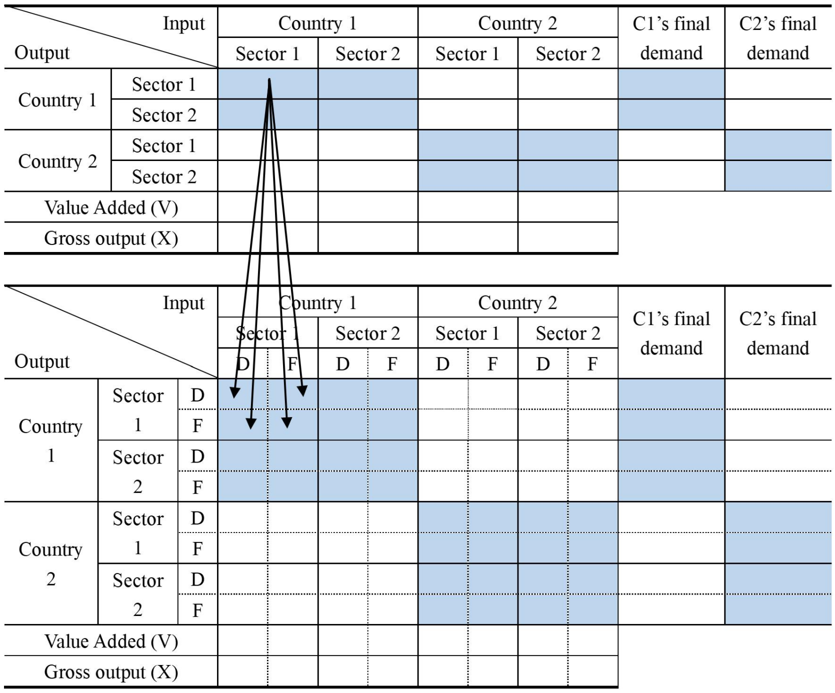

Figure 1: Breaking down the ICIO table into ICIO-DF tables

Due to the enormous size of ICIO-DF, Monte Carlo simulations and calculations of the Leontief inverse matrix, output multipliers, and export value-added exceeded the computational capacity of personal computers. To save computation time, this paper converted the ICIO-DF table into a non-competitive table for China, with the following processing steps: (1) adding up the intermediate inputs from different countries or regions as the import intermediate input matrix for division by sector and ownership; (2) adding up the final demand from different countries or regions as the import final demand matrix for division by sector and ownership; (3) adding up China's supply of intermediate inputs and final demand for other countries or regions as the export vector. (4) Since the intermediate input, value-added, final demand, and output for private households with employed persons in ICIO-DF table are all zero, this paper removed this sector from the differentiated domestic and foreign investment non-competitive input-output table for China.

The adjustments mentioned above only involve the treatment of the import intermediate input matrix and final demand matrix for China, while the domestic intermediate input matrix, output vector, and value-added vector remain unchanged and retain the values of the ICIO-DF. Therefore, the adjusted non-competitive table of China with domestic and foreign investment classification can still be used to test the impact on accuracy of the model caused by assumption of proportionality and quadratic optimization method.

4.2. Calculate the Leontief inverse matrix, output multiplier and export added-value

As the intermediate flow matrix Z^{m0} in the Chinese non-competitive input-output table with domestic and foreign investment classification is obtained by aggregating the intermediate inputs from different countries or regions to China in the ICIO-DF table, it possesses a level of uncertainty comparable to that of the domestic intermediate flow matrix Z^{d0}. Therefore, in the Monte Carlo simulation performed in this study, the intermediate flow matrices for both domestic and imported goods were processed according to the same rules.

Z^{d0} and Z^{m0} represent the domestic and imported intermediate flow matrices, respectively, in the non-competitive input-output table of China with domestic and foreign investment classification, while z_{ij}^{d0} and z_{ij}^{m0} represent the elements in these matrices. During the Monte Carlo simulation, we assume that the elements of the initial domestic and imported intermediate flow matrices follow a normal distribution with mean values as z_{ij}^{d0} and z_{ij}^{m0}, respectively. This assumption means that elements of initial intermediate flow matrices in each simulation is obtained by adding a disturbance term (e_{ij}^{d} and e_{ij}^{m}) to the base value (z_{ij}^{d0} and z_{ij}^{m0}). The disturbance terms follow normal distributions: $e_{ij}^{d}\sim\mathrm{N}(0,\sigma z_{ij}^{d0})$ and $e_{ij}^{m}\sim\mathrm{N}(0,\sigma z_{ij}^{m0})$. Therefore, in each simulation, the initial domestic intermediate flow matrix is marked as $Z^{di}=Z^{d0}+E^{d}$, and the initial imported intermediate flow matrix is marked as $Z^{mi}=Z^{m0}+E^{m}$, where E^{d} and E^{m} are the disturbance matrices of domestic and imported intermediate flow matrices, respectively.

To maintain the structural characteristics of the heterogeneous input-output model, we performed TRAS adjustments on initial domestic and imported intermediate flow matrices, denoted as Z^{di} and Z^{di}, respectively, resulting in adjusted matrices Z^{d} and Z^{m}.

When calculating the total input vector of a sector, in order to satisfy the column balance relationship of the input-output model, we sum the domestic intermediate input matrix, import intermediate input matrix, and value-added vector column-wise, that is, $\mathrm{x}=uZ^{d}+uZ^{m}+\mathrm{v}$. Here, u is a row vector with all elements equal to 1, and v is the value-added row vector.

Based on this, the direct consumption coefficient matrix of domestic products can be calculated as

$$A^{d}=Z^{d}\hat{x}^{-1}$$

where the elements a_{ij}^{d} represents the direct consumption of domestic products of sector i caused by the unit output of sector j. Similarly, the direct consumption coefficient matrix of imported products can be obtained as

$$A^{m}=Z^{m}\hat{x}^{-1}$$

According to the basic theory of input-output analysis, the Leontief inverse matrix is calculated

$$\mathrm{B}=(\mathrm{I}-A^{d})^{-1}$$

where I is the unit matrix with the same dimension as A^{d}. The Leontief inverse matrix, also known as the total requirement coefficient matrix, has elements b_{ij} that represents the final demand for the output of sector i in order to consume the unit value of sector j. The output multiplier vector

$$\lambda=\mathrm{uB}$$

can be obtained by adding up the columns of the Leontief inverse matrix, where each element λ_j represents the total output of all sectors directly or indirectly induced by the unit final demand of sector j.

Finally, the heterogeneous input-output model obtained from each simulation is used to estimate the value added generated by exports. Specifically, the domestic value added driven by exports is calculated as

$$y^d = A_v(\mathrm{I} - A^d)^{-1}e$$

and the foreign value added driven by exports is

$$y^m = uA^m(\mathrm{I} - A^d)^{-1}e$$

where A_v is the row vector of value-added and e is the export column vector.

5. Conclusion

5.1. Accuracy of Leontief inverse matrix elements

In order to investigate the effect of the number of Monte Carlo simulations on the research results, and to determine the number of simulations needed to obtain stable results, this study conducted 10 to 10,000 simulations of the non-competitive input-output table of China adapted from the ICIO-DF table in two scenarios. The simulation results show that whether the elements of the intermediate flow matrix follow a normal distribution with a standard deviation of 0.1 times the mean or a lognormal distribution, when the number of simulations exceeds 500, the distribution of the standard deviation and coefficient of variation of the Leontief inverse matrix elements obtained by simulation tends to be stable. This is consistent with the research findings of Dietzenbacher (2006). To ensure the reliability of the research conclusions, we use the results of 10,000 Monte Carlo simulations as the basis for investigating the impact of assumption of proportionality and optimization methods on the accuracy of heterogeneous input-output models.

We conducted 10,000 Monte Carlo simulations under the scenario one where the elements of the intermediate flow matrix followed a normal distribution with a standard deviation of 0.1 times the mean. The maximum standard deviation of the elements in the Leontief inverse matrix is 0.04275, the median is 0.00008, and 75% of the elements have a standard deviation of less than 0.0004. Thus, the vast majority of the elements only have minimal uncertainty.

According to the heat map, the standard deviations of the elements near the diagonal of the Leontief inverse matrix before the 38th column are relatively large, indicating a significant absolute error in the total requirement coefficients of mining and manufacturing enterprises for different ownership attribute enterprises within their respective industries. In addition, the standard deviations of the 21st row's 21st to 40th columns in the Leontief inverse matrix are also relatively large. These elements correspond to the total requirement coefficients of the manufacturing and construction sectors for domestic enterprises in the basic metal sector. Since basic metals are an important input in the production processes of the manufacturing and construction sectors, the corresponding total requirement coefficients are relatively large and can generate large standard deviations.

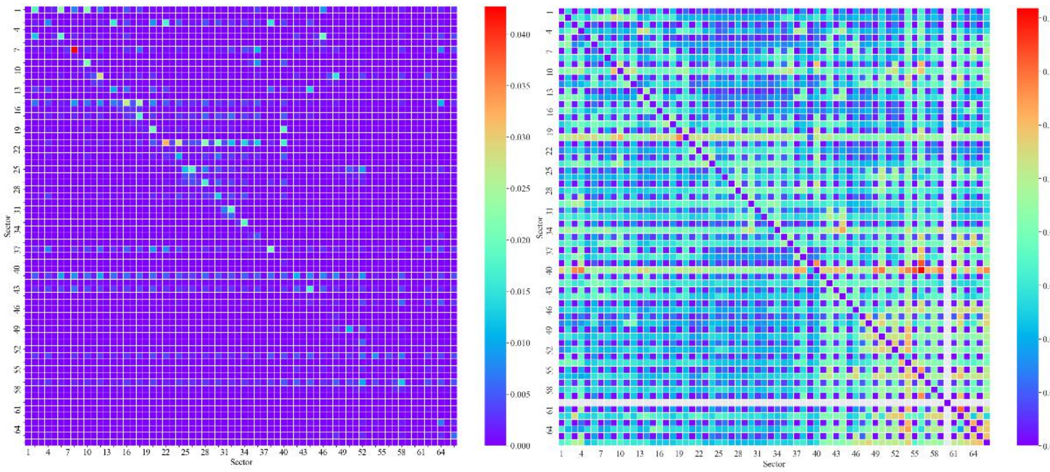

Figure 2: Heatmap of standard deviation and coefficient of variation of the elements in the Leontief inverse matrix in scenario one

Through 10,000 Monte Carlo simulations in scenario one, the maximum coefficient of variation of the elements in the Leontief inverse matrix is 0.16378, with a median of 0.0482, and 95% of the elements have a coefficient of variation less than 0.1. Thus, only a small number of elements in the Leontief inverse matrix have slightly larger coefficient of variation than their corresponding elements in the intermediate flow matrix, indicating that the Leontief inverse matrix is more precise than the intermediate flow matrix. The uncertainty of the intermediate flow matrix is significantly reduced when calculating the Leontief inverse matrix.

The coefficient of variation (CV) heatmap of the Leontief inverse matrix for scenario 1 shows that the coefficient of variation of the main diagonal elements is relatively small, implying that the relative error of total requirements coefficients for firms with the same ownership in the same sector is significantly lower than average. This is mainly because these coefficients have larger numerical values. Moreover, the elements in the lower right corner of the Leontief inverse matrix have a generally higher coefficient of variation, indicating that the total requirements coefficients between service sectors are more sensitive to the initial uncertainty of the input-output matrix even if the initial elements are normally distributed with a relative standard deviation of 0.1. Missing coefficient of variation values in the 60th row and column are due to the absence of foreign-funded enterprises in public management, national defense, and compulsory social security sectors, resulting in corresponding intermediate flow matrix and Leontief inverse matrix values of zero.

In Scenario 2, assuming that the elements of the intermediate flow matrix follow a lognormal distribution with a standard deviation as

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

Through 10,000 Monte Carlo simulations, the maximum standard deviation of the elements in the Leontief inverse matrix is 0.01061, the median is 0.00005, and more than 75% of the element standard deviations are less than 0.0002. The maximum value of the coefficient of variation of the Leontief inverse matrix elements is 0.62301, with an average of 0.050649 and a median of 0.02788, and more than 75% of the element coefficients of variation are less than 0.06, with very few elements having a larger coefficient of variation. It can be seen that the distribution of the Leontief inverse matrix elements in Scenario 2 is more concentrated than in Scenario 1, indicating a higher level of relative accuracy.

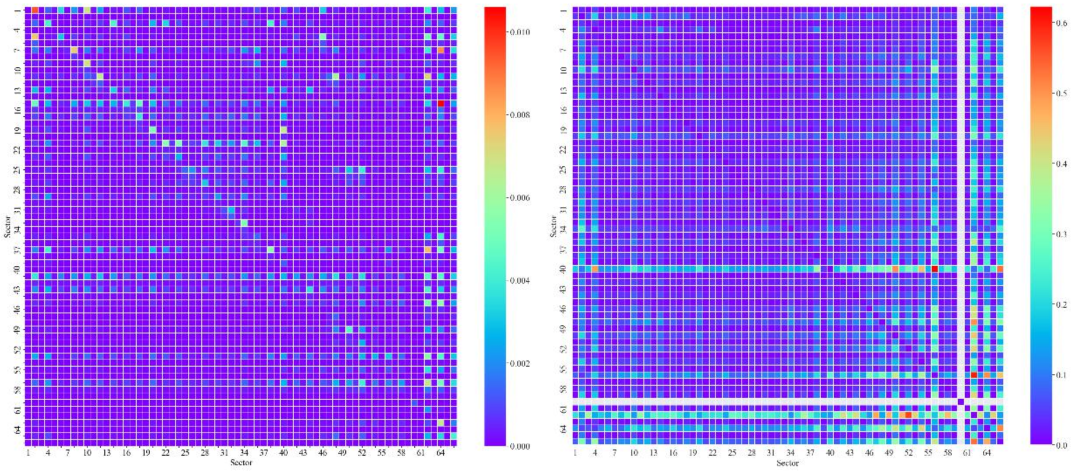

Figure 3: Heatmap of standard deviation and coefficient of variation of the elements in the Leontief inverse matrix in scenario two

Figure 3 reveals that the total requirements coefficient of foreign-funded enterprises in the agriculture, forestry and fishing sector to domestic enterprises has the maximum standard deviation among the elements in the Leontief inverse matrix. Moreover, the standard deviation of the elements near the main diagonal and the elements in columns 21-40 of row 21 in the Leontief inverse matrix are relatively high, which is similar to the characteristics observed in Scenario 1.

The coefficients of variation for several rows and columns in the Leontief inverse matrix are relatively large, including the 40th, 56th, 62nd, 64th, and 66th rows, as well as the 56th, 62nd, 64th, and 66th columns. The main reason for this phenomenon is the relatively low participation of foreign invested enterprises in sectors such as construction, real estate, education, human health and social work, arts, entertainment, recreation and other service sectors. The corresponding elements in the intermediate flow matrix are small, and in Scenario 2 where the variability of matrix elements is negatively correlated with their absolute value, the variability of these elements in the Leontief inverse matrix is relatively large.

In order to study the propagation law of uncertainty in Scenario 2, this paper calculates the ratio of the coefficient of variation of the corresponding elements in the Leontief inverse matrix and the intermediate flow matrix, and draws a heat map. As shown in Figure 4, the coefficient of variation of the Leontief inverse matrix elements in Scenario 2 is smaller than that of the intermediate flow matrix elements. This phenomenon is similar to Scenario 1, indicating that the Leontief inverse matrix is less uncertain than the intermediate flow matrix. In addition, the ratio of the coefficient of variation is smaller on the main diagonal, indicating that the uncertainty of the total requirements coefficients of enterprises with the same ownership attributes in the same sector is significantly reduced when calculating the Leontief inverse matrix. Moreover, the elements in the lower right corner of the figure are relatively large. This phenomenon is also reflected in Scenario 1, indicating that the uncertainty of the service sectors is reduced to a lesser extent compared to the mining and manufacturing sectors when calculating the Leontief inverse matrix.

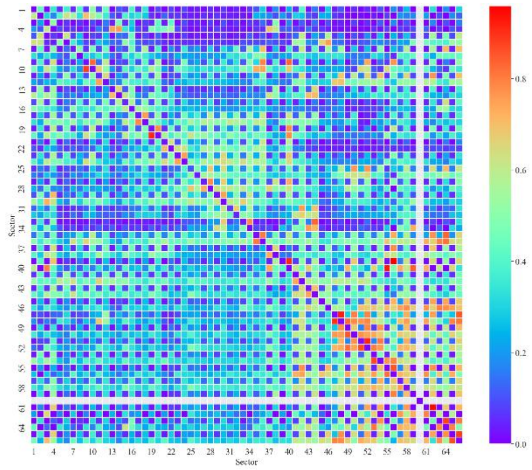

Figure 4: Ratio of coefficient of variation between elements in the Leontief inverse matrix and the intermediate flow matrix under Scenario 2

In summary, regardless of whether the elements of the intermediate flow matrix are random variables that follow a normal distribution or a lognormal distribution, the uncertainty of the intermediate flow matrix will be significantly reduced in the Leontief inverse matrix. When constructing heterogeneous input-output models such as ICIO-DF, the uncertainty caused by assumption of proportionality and optimization method on the intermediate flow matrix will be significantly alleviated when calculating the Leontief inverse matrix, and these conventional techniques for compiling heterogeneous input-output tables will not have a significant impact on the accuracy of the model.

5.2. The accuracy of the output multipliers

Output multipliers is a row vector obtained by summing the columns of Leontief inverse matrix

$$\lambda = \mathrm{uB}$$

where the element λ_j represents the total amount of output that is generated in all sectors by an increase of one unit of final demand in sector j. Multipliers, such as the output multipliers, are of great importance in input-output analysis, and it is necessary to investigate the impact of using assumption of proportionality and optimization methods on the accuracy of output multipliers.

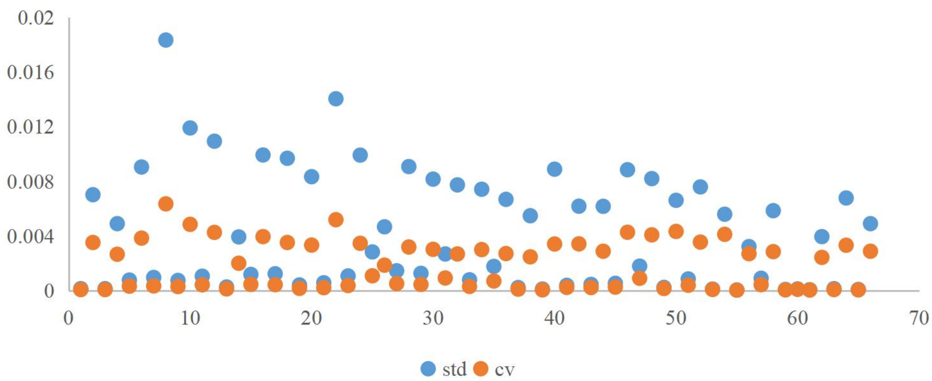

Figure 5: Standard deviation and coefficient of variation of output multipliers for each sector in Scenario 1

Figure 5 shows the standard deviation and coefficient of variation of output multipliers for each sector obtained from 10,000 Monte Carlo simulations in Scenario 1. It can be seen that the standard deviation of output multipliers for each sector is within the range of 0-0.02, and the coefficient of variation of output multipliers is within the range of 0-0.008, indicating that the assumptions in Scenario 1 do not cause significant fluctuations in the estimated values of sectoral output multipliers and that output multipliers are highly accurate. Among all domestic and foreign invested sectors, the foreign invested sectors of textiles, wearing apparel, leather and related products has the largest standard deviation and coefficient of variation of output multipliers, at 0.01834 and 0.00634, respectively. The standard deviation and coefficient of variation of output multipliers for each sector are positively correlated, but the distribution of coefficient of variation of output multipliers is more concentrated. The standard deviation of output multipliers for manufacturing sectors is generally larger than that for service sectors, but the difference in coefficient of variation is not significant.

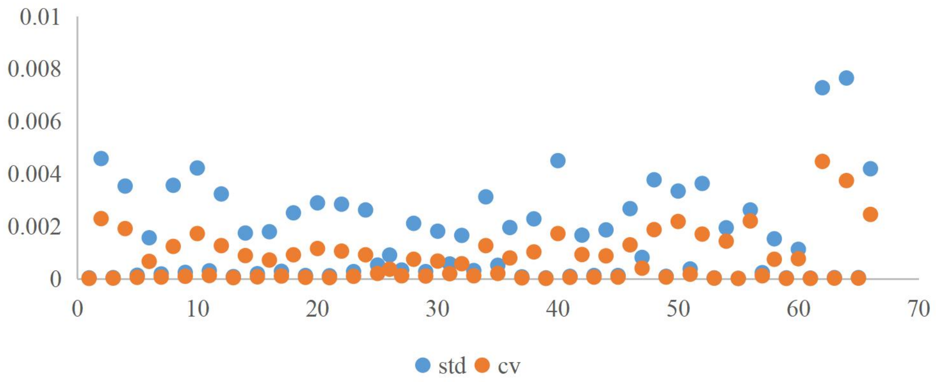

Figure 6: Standard deviation and coefficient of variation of output multipliers for each sector in Scenario 2

Figure 6 shows the standard deviation and coefficient of variation of output multipliers for each sector in scenario 2. Compared with scenario 1, the standard deviation and coefficient of variation of output multipliers are smaller, with the standard deviation distribution ranging from 0 to 0.01 and the coefficient of variation distribution ranging from 0 to 0.005. However, unlike scenario 1, the standard deviations and coefficients of variation of output multipliers are larger for the foreign invested sectors in education, human health and social work, arts, entertainment, recreation, and other services than for the manufacturing sectors. This is because foreign invested enterprises are less involved in these industries, and the corresponding values of the intermediate flow matrix elements are smaller. Under the assumption of lognormal distribution, the uncertainty of the intermediate flow matrix elements is greater.

Combining Figures 5 and 6, it can be observed that regardless of whether the intermediate flow matrix elements follow a normal or lognormal distribution, the relative uncertainty measured by the coefficient of variation decreases gradually from the intermediate flow matrix elements to the elements of the Leontief inverse matrix and then to the output multipliers. The errors in intermediate flow matrix are neutralized during the relevant calculations of the input-output model.

5.3. Accuracy of export value-added

The heterogeneous input-output model has been widely applied in measuring export value-added, addressing the bias caused by the inability of traditional input-output models to distinguish trade patterns and enterprise types. However, there has been no discussion regarding the accuracy of the research results. This paper uses the non-competitive input-output table of

China with domestic and foreign investment classification, adapted from ICIO-DF, to measure the domestic and foreign value-added in China's exports, and discusses the error range of China's export value-added in two different scenarios.

Due to the significant differences in absolute values of export value-added among different sectors, this paper normalizes the export value-added of each sector by dividing the value obtained from each simulation by the mean value obtained from 10,000 simulations. This allows for analysis of the relative error range of export value-added for each sector. Each violin plot shows the red dot representing the export value-added calculated from the original data of the non-competitive input-output table of China with domestic and foreign investment classification. The three horizontal lines represent the maximum, minimum, and median values of the simulation results, while the black bars represent the upper and lower quartiles of the simulation results.

Based on the findings from Figure 7 and Figure 8, we can observe that in scenario one, where the intermediate flow matrix elements are random variables following a normal distribution with a standard deviation of 0.1 times the mean, the simulation results of domestic and foreign export value-added for all sectors exhibit a normal distribution. Additionally, it is noteworthy that our original data calculated export value-added is at the center of the simulation results. The upper and lower quartiles of the normalized simulation results for all sectors are within the range of 0.94-1.06. Except for the foreign-invested sector in construction, the extremes of the normalized simulation results are within the range of 0.8-1.2. Among all 66 sectors, 62 sectors have normalized simulation results with extremes within the range of 0.9-1.1.

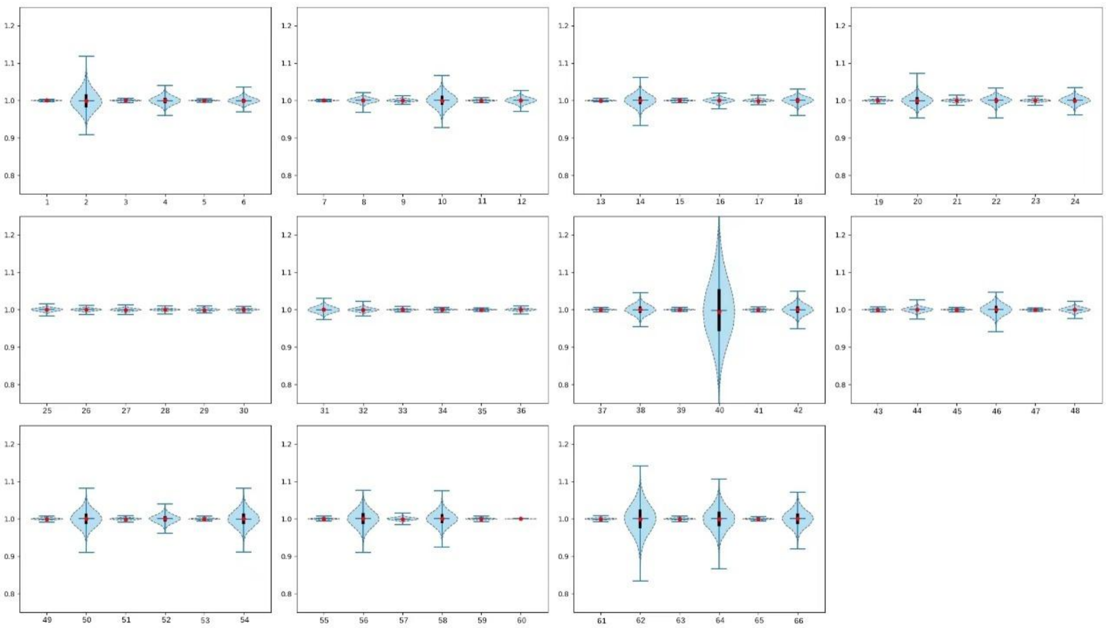

Figure 7: **The violin plots of domestic export value-added for each sector under Scenario 1**

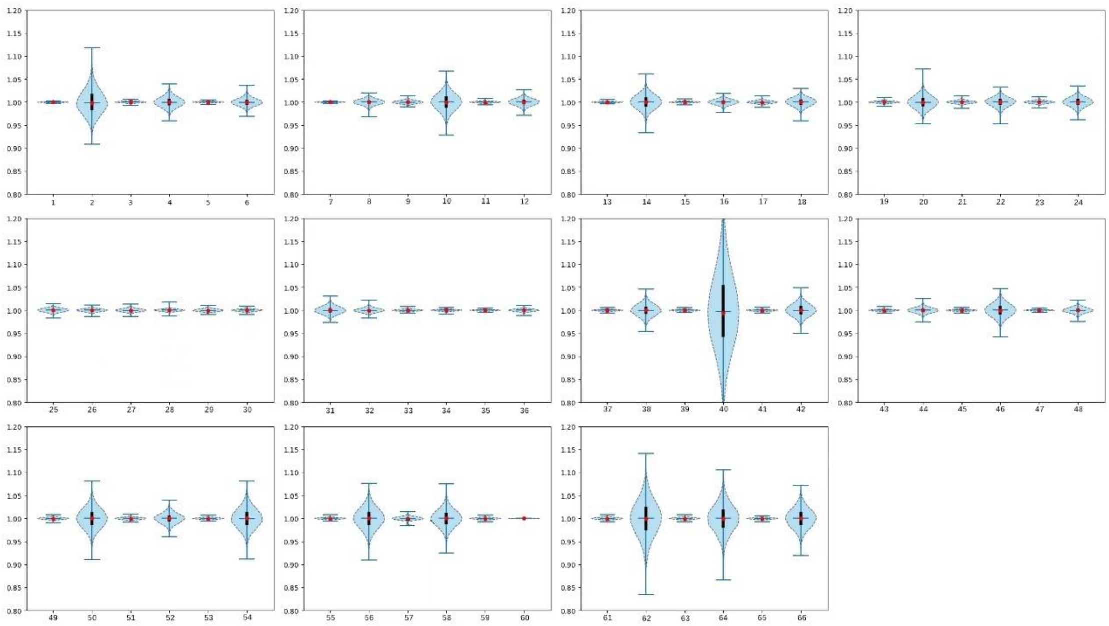

Figure 8: The violin plots of foreign export value-added for each sector under Scenario 1

In Scenario 2, the distribution range of the normalized simulation results of most sectors' export value-added is smaller than that in Scenario 1, except for the foreign invested sector in human health and social work, which are an exception and have extreme values that exceed the range of 0.8-1.2. This is because the corresponding elements in the intermediate flow matrix for this sector is extremely small, and the relative standard deviation of the elements is magnified under the assumption of Scenario 2, resulting in a more scattered simulation result for the export value-added. However, the domestic and foreign value-added included in the sector's export are both less than 1 million US dollars, so the absolute error of the sector's export value-added can still be ignored. Overall, the export value-added of sectors in Scenario 2 is more accurate than that in Scenario 1.

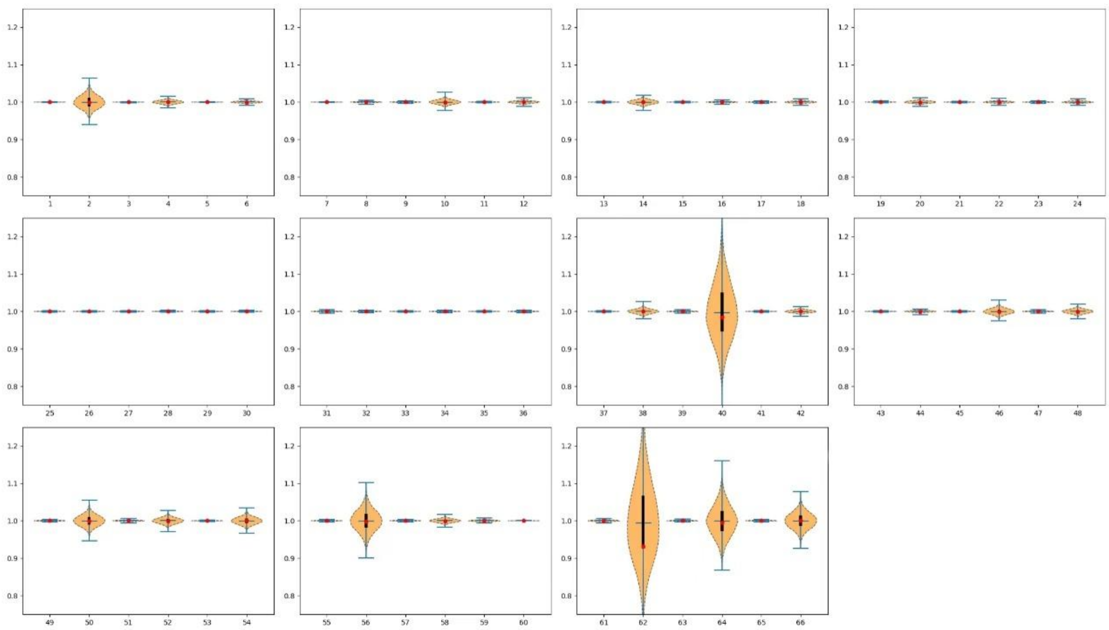

Figure 9: The violin plots of domestic export value-added for each sector under Scenario 2

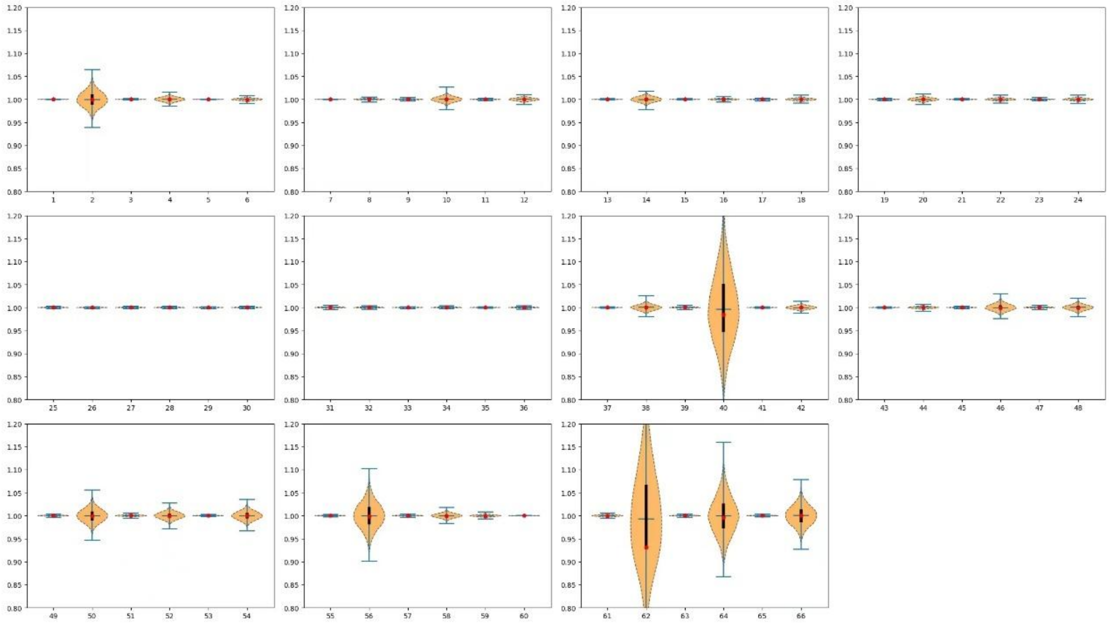

Figure 10: The violin plots of foreign export value-added for each sector under Scenario 2

The above findings suggest that sectors' export value-added has relatively low uncertainty, regardless of the distribution of intermediate flow matrix elements. Specifically, the relative error range of sectors' export value-added is smaller than that of intermediate flow matrix elements and Leontief inverse matrix elements. The export value-added calculated by the non-competitive input-output model of China with domestic and foreign investment classification has a high degree of accuracy. Furthermore, assumption of proportionality and optimization method widely employed in compiling heterogeneous input-output tables do not significantly affect overall results such as multipliers and export value-added. The above conclusion aligns with research findings on the decreasing uncertainty of SRIO and MRIO tables from the intermediate flow matrix to the Leontief inverse matrix and multipliers. Therefore, this verifies the universality of the concept of overall accuracy proposed by Lenzen et al. (2013) in various input-output models.

5.4. The impact of standard deviation variation in intermediate flow matrix elements on accuracy

To investigate the accuracy of the non-competitive input-output model of China with domestic and foreign investment classification, this paper adopts two scenarios where the intermediate flow matrix elements follow a normal distribution and a log-normal distribution respectively. However, under each distribution type, the standard deviation of the elements only satisfies one fixed functional form. In order to enhance the robustness of the research results, we change the mechanism for forming the standard deviation of the intermediate flow matrix elements and explore the impact of different functional forms of the standard deviation on the accuracy of the Leontief inverse matrix, output multipliers, and export value-added.

Based on the previous analysis, it can be concluded that the simulation results become stable after 1000 Monte Carlo simulations. In scenario one, where the intermediate flow matrix elements are assumed to follow a normal distribution, this study conducted 1000 Monte Carlo simulations with a relative standard deviation ranging from 0.1 to 1.0 in increments of 0.01 to measure the distribution of the Leontief inverse matrix elements and output multipliers.

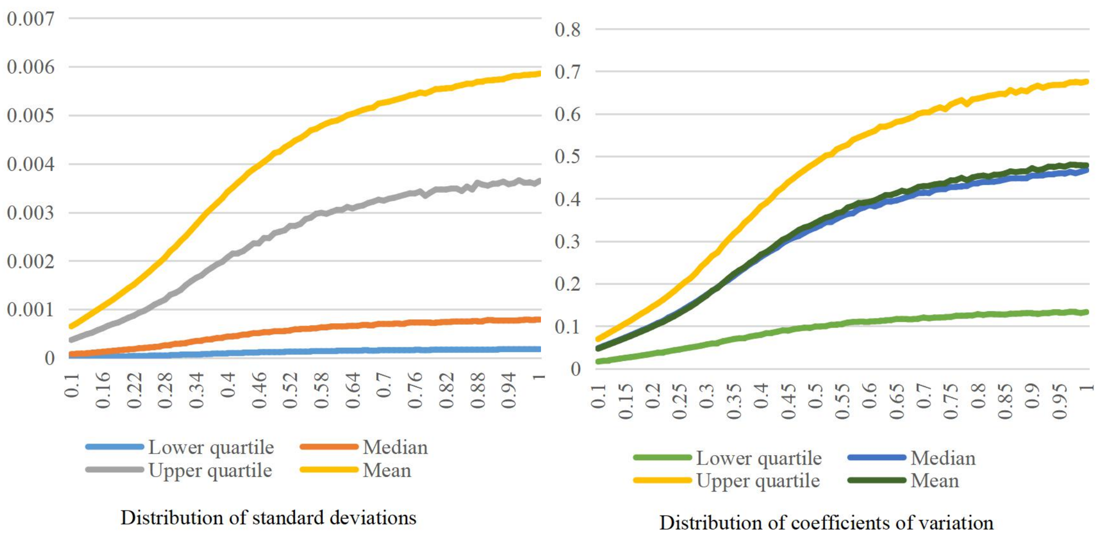

Figure 11: Distribution of Leontief inverse matrix elements under different relative standard deviations of intermediate flow matrix elements in Scenario 1

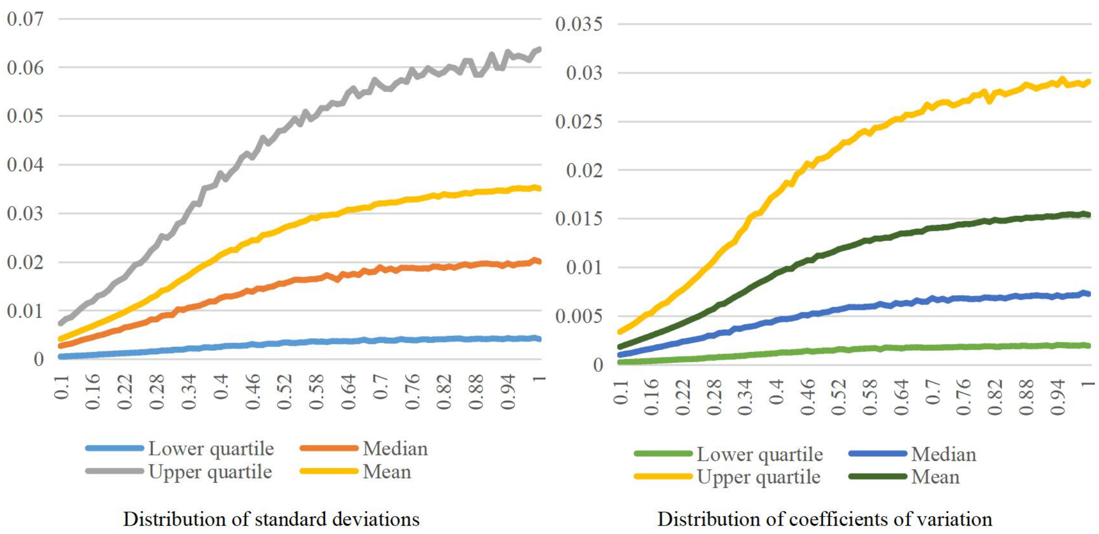

Figure 12: Distribution of output multipliers under different relative standard deviations of intermediate flow matrix elements in scenario 1

From Figures 11 and 12, it can be observed that the standard deviation and coefficient of variation of the Leontief inverse matrix elements and output multipliers slowly increase as the relative standard deviation of the intermediate flow matrix elements increases. The mean and upper quartile of the standard deviation of the Leontief inverse matrix elements do not exceed 0.006, and the mean and upper quartile of the coefficient of variation do not exceed 0.7. The mean and upper quartile of the standard deviation of the output multiplier are less than 0.07, and the mean and upper quartile of the coefficient of variation are less than 0.03. These phenomena indicate that when the distribution of intermediate flow matrix elements in the non-competitive input-output model of China with domestic and foreign investment classification is more concentrated than the normal distribution with a relative standard deviation of 1, the accuracy of the Leontief inverse matrix elements obtained from the table is higher than that of the intermediate flow matrix elements, and the accuracy significantly improves from the Leontief inverse matrix to the output multipliers.

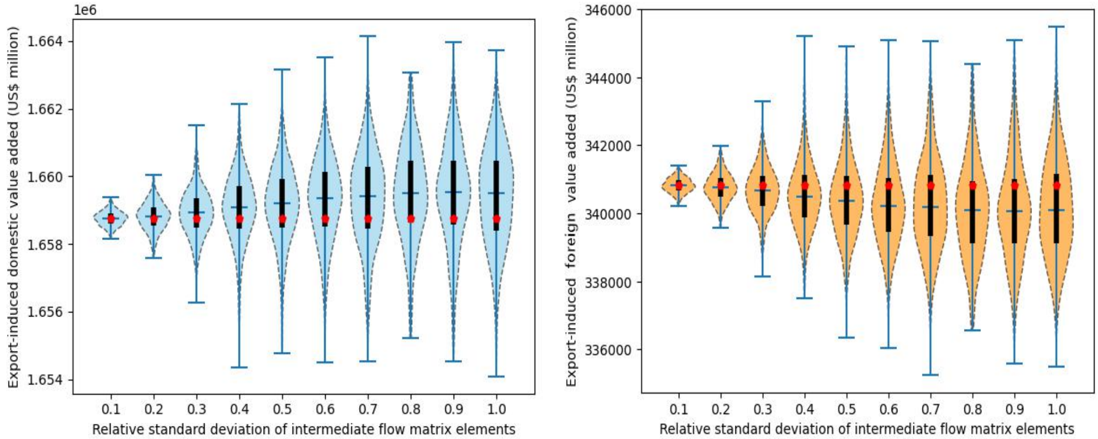

Figure 13: Distribution of export value-added under different relative standard deviation of intermediate flow matrix elements in Scenario 1

From Figure 13, it can be seen that with the increase of the relative standard deviation of intermediate flow matrix elements, the distribution range of export value-added simulation results shows a trend of rapid expansion followed by stabilization. Moreover, the distribution range of domestic value-added driven by exports does not exceed 1% of the mean, while the distribution range of foreign value-added driven by exports does not exceed 3% of the mean. This indicates that even when the uncertainty of intermediate flow matrix elements reaches the standard deviation to 1 times of the mean, the accuracy of export value-added is still very high. In addition, with the increase of the relative standard deviation of intermediate flow matrix elements, the domestic value-added driven by exports calculated from the original data of the non-competitive input-output model of China with domestic and foreign investment classification, gradually approaches the simulation result from below, while the foreign value-added driven by exports gradually approaches the simulation result from above. This phenomenon confirms the research results of Dietzenbacher (2006), which found that when the intermediate flow matrix elements are random variables that follow a certain distribution rule, the Leontief inverse matrix elements have a statistically significant but negligible positive bias, meaning that elements with larger values are more likely to be overestimated. Due to the fact that the elements in the Leontief inverse matrix for domestic enterprises are generally larger than those for foreign enterprises in the same industry, and domestic enterprises have a higher domestic value-added rate while foreign enterprises have a higher foreign value-added rate, it results in a phenomenon of upward bias for domestic value-added and downward bias for foreign value-added when calculating the export value-added. However, this bias has little effect on the calculation results of export value-added and can be ignored in empirical research.

In scenario two, the intermediate flow matrix elements are random variables that follow a log-normal distribution

$$\sigma z_{ij}^{0} = a\left|z_{ij}^{0}\right|^{b}$$

As the mechanism by which changes in a affect the accuracy of the model in this scenario is the same as that in scenario one, we maintain the setting of $a = 0.393$ and study only the effect of changes in the exponent b on accuracy. Additionally, given the fact that larger values in the intermediate flow matrix of input-output tables correspond to more accurate estimates, we set the range of variation for b to 0-1 and conduct 1,000 Monte Carlo simulations at intervals of 0.01 to measure the distribution of Leontief inverse matrix elements, output multipliers, and export value-added.

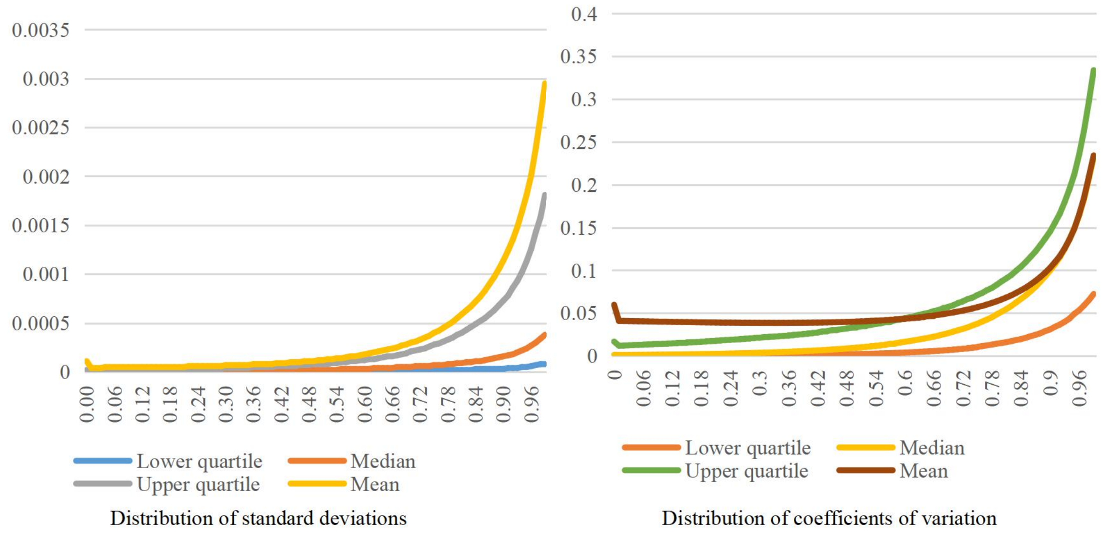

Figure 14: The distribution of Leontief inverse matrix elements under different values of b

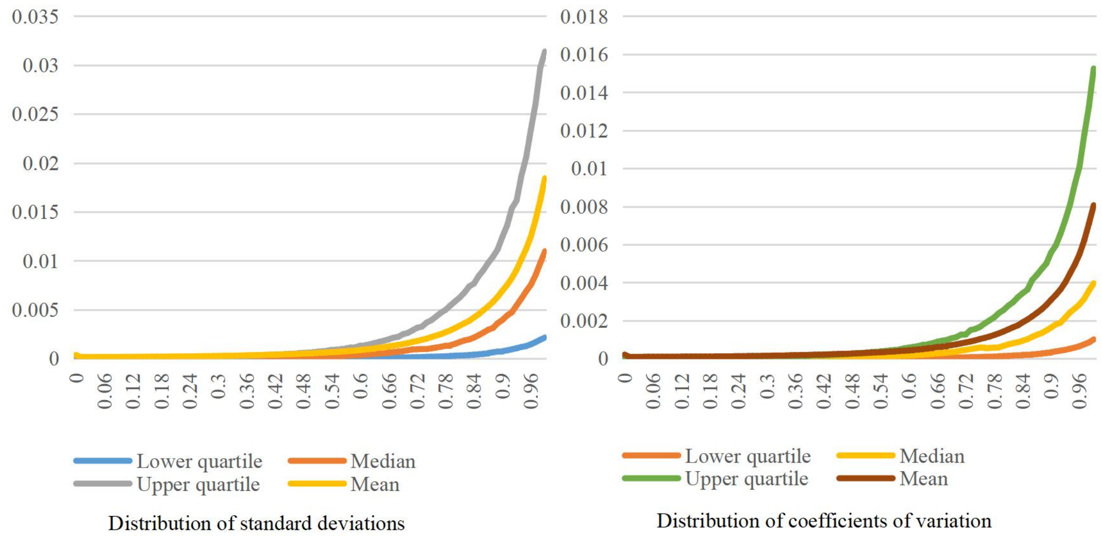

Figure 15: The distribution of output multipliers under different values of b

From Figure 14 and Figure 15, it can be seen that when b is less than 0.5, the standard deviation and coefficient of variation of the Leontief inverse matrix elements and output multipliers remain at a minimum level. When b is greater than 0.5, the standard deviation and coefficient of variation show an exponential increase with the increase of b. However, even when b approaches 1, the elements with larger values in the intermediate flow matrix are not estimated more accurately. The upper quartile and mean of the standard deviation of the Leontief inverse matrix elements are both less than 0.003, and the upper quartile and mean of the coefficient of variation are both less than 0.35. The upper quartile and mean of the standard deviation of the output multipliers are both less than 0.035, and the upper quartile and mean of the coefficient of variation are both less than 0.016. This indicates that when the exponent in the standard deviation function of the intermediate flow matrix elements varies within 1, the Leontief inverse matrix, especially the output multipliers, still maintain a low level of uncertainty. This uncertainty shows a decreasing trend from the intermediate flow matrix to the Leontief inverse matrix and then to the output multipliers.

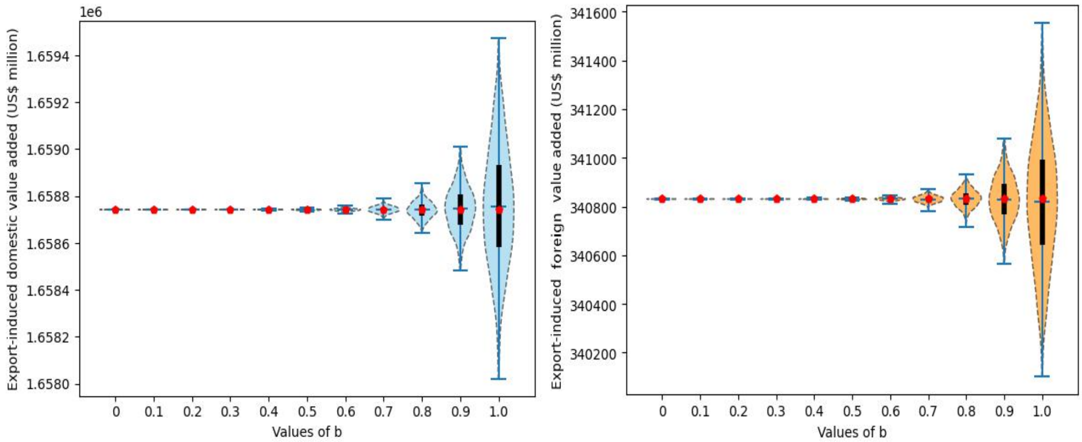

Figure 16: The distribution of export value-added under different values of b

Similar to the patterns of variation in the Leontief inverse matrix elements and output multipliers, the distribution of simulated results for export value-added is highly concentrated when b is less than 0.5. As b increases beyond 0.5, the range of simulated results for export value-added shows an exponential increase. However, even when $b = 1$, export value-added still exhibits a high degree of accuracy.

If we consider the normal distribution satisfied by the intermediate flow matrix elements in Scenario 1 as a special case of lognormal distribution, the determining function of the standard deviation of intermediate flow matrix elements can be expressed as

$$\sigma z_{ij}^{0} = a\left|z_{ij}^{0}\right|^{b}$$

where a and b are two variables that affect the accuracy of the heterogeneous input-output model.

Based on the analysis above, it is known that the uncertainty of Leontief inverse matrix, output multipliers, and export value added increases exponentially with the increase of the exponential part b. However, due to the fact that larger values in the intermediate flow matrix are estimated more accurately, the value of b is limited to the range of 0-1. Even when b is equal to 1, the model can still maintain a high level of accuracy.

The uncertainty of the heterogeneous input-output model also increases with the increase of parameter a, but the magnitude of the increase shows a converging trend. This is because the intermediate flow matrix must satisfy the specific structural form of the heterogeneous input-output model, which allows for the mutual neutralization of errors in input-output matrix elements. This mutual neutralization effect also ensures the overall accuracy of the model, and the accuracy of the intermediate flow matrix, Leontief inverse matrix, output multipliers, and export value-added gradually improves, with the accuracy of the calculation results approaching the total value being the highest.

There are sufficient reasons to believe that even if the intermediate flow matrix elements are random variables that follow a certain distribution, the data of the non-competitive input-output model of China, adapted from the ICIO-DF, have significant biases. However, the Leontief inverse matrix, output multipliers, and export value-added still have high accuracy and gradually improve in accuracy. Therefore, although the assumption of proportionality and optimization method widely used in compilation of heterogeneous input-output tables may cause biases in intermediate flow matrix, they will not seriously affect the accuracy of the model and empirical research based on the model.

6. Summary

A comprehensive understanding of the accuracy of the heterogeneous input-output model is a fundamental prerequisite for correctly evaluating model results and making decisions based on model conclusions. In recent years, the heterogeneous input-output model, as an emerging research topic, has been widely applied in empirical research and has yielded a large number of research results. However, to date, there has been no literature on the accuracy of the heterogeneous input-output model, which is a major deficiency in this field of research.

This paper delves into the conventional methods of constructing heterogeneous input-output models and finds that the assumption of proportionality and optimization method used in compiling the intermediate flow matrix are important sources of errors affecting the accuracy of heterogeneous input-output models. Based on this, we propose a method for measuring the accuracy of the Leontief inverse matrix, output multipliers, and export value-added, which is based on generating the intermediate flow matrix through Monte Carlo simulation and adjusting it to meet the specific structure of the heterogeneous input-output model by TRAS. With the use of this method, the current study investigates the accuracy of the non-competitive input-output table of China, which was adapted from the 2016 Input-Output Framework for Foreign and Domestic Firms (ICIO-DF). By simulating the intermediate flow matrix elements 10,000 times under normal distribution and lognormal distribution, and varying the mechanism of forming the standard deviation of intermediate flow matrix elements during the simulation, we find that regardless of the distribution form of the intermediate flow matrix elements, the uncertainty of the intermediate flow matrix, the Leontief inverse matrix, output multipliers, and export value-added all show a decreasing trend, and the model can maintain high overall accuracy.

The results of this study demonstrate that heterogeneous input-output models represented by ICIO-DF maintain good overall accuracy. As long as the total matrices such as output, value added, imports and exports are accurately estimated in the process of compiling heterogeneous tables, the assumption of proportionality and optimization method commonly used in the estimation of the intermediate flow matrix will not significantly affect the accuracy of the model.

References

[1] Duan, Y. and C. Yang (2018). Value added Decomposition on China's Regional Exports Based on the Heterogeneity of Trade Mode[J]. The Journal of World Economy 41(04): 75-98.

[2] Duan, Y., et al. (2013). Non-competitive input-output model capturing foreign-funded enterprise and trade pattern[J]. Systems Engineering-Theory & Practice 33(09): 2204-2211.

[3] Liu, Z., et al. (2010). Input-occupancy-output models of the non-competitive type and their application – an examination of the China-US trade surplus[J]. Social Sciences in China 31(01): 35-54.

[4] Zhang, J., et al. (2022). The employment effects and determinants in China's dual circulation considering firm size heterogeneity information[J]. Systems Engineering-Theory & Practice

42(12): 3151-3164.
[5] Zhang, J. (2021). China's Non-Competitive Input-Output Model Capturing Firm Size Heterogeneity and Its Application[D]. Beijing: Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 2021.
[6] Cadestin C, De Backer K, Desnoyers-James I, et al. Multinational enterprises and global value chains: the OECD analytical AMNE database[J]. 2018.
[7] Cadestin C, Backer K, Miroudot S, Moussiegt L, Rigo D, and Ye M. Multinational enterprises in domestic value chains. OECD Science Technology & Industry Policy Papers, 2019, No.63.
[8] Chen Q, Zhu K, Liu P, et al. Distinguishing China's processing trade in the world input-output table and quantifying its effects[J]. Economic Systems Research, 2019, 31(3): 361-381.
[9] Chen X, Cheng L, Fung K, and Lau, L. The estimation of domestic value added and employment induced by exports: An application to Chinese exports to the United States, Department of Economics, Stanford University, Working Paper, 2001.
[10] Dietzenbacher E. Multiplier estimates: to bias or not to bias? [J]. Journal of regional science, 2006, 46(4): 773-786.
[11] Dietzenbacher E, Pei J, Yang C. Trade, production fragmentation, and China's carbon dioxide emissions[J]. Journal of Environmental Economics and Management, 2012, 64(1): 88-101.
[12] Duan Y, and Jiang X. Haven or pollution halo? A re-evaluation on the role of multinational enterprises in global CO_2 emissions[J]. Energy Economics, 2021, 97: 105181.
[13] Jensen R C. The concept of accuracy in regional input-output models[J]. International Regional Science Review, 1980, 5(2): 139-154.
[14] Jensen R C, West G R. The effect of relative coefficient size on input—output multipliers[J]. Environment and Planning A, 1980, 12(6): 659-670.
[15] Jiang M, Liu L, Behrens P, et al. Improving subnational input–output analyses using regional trade data: a case-study and comparison[J]. Environmental science & technology, 2020, 54(19): 12732-12741.
[16] Jiang X, Zhu K, and Green C. China's energy saving potential from the perspective of energy efficiency advantages of foreign-invested enterprises[J]. Energy Economics, 2015a, 49: 104-112.
[17] Jiang X, Guan D, Zhang J, Zhu K, and Green C. Firm ownership, China's export related emissions, and the responsibility issue[J]. Energy Economics, 2015b, 51: 466-474.
[18] Koopman R, Wang Z, Wei S J. Estimating domestic content in exports when processing trade is pervasive[J]. Journal of development economics, 2012, 99(1): 178-189.
[19] Lenzen M. Aggregation versus disaggregation in input–output analysis of the environment[J]. Economic Systems Research, 2011, 23(1): 73-89.
[20] Lenzen M, Kanemoto K, Moran D, et al. Mapping the structure of the world economy[J]. Environmental science & technology, 2012, 46(15): 8374-8381.
[21] Lenzen M, Moran D, Kanemoto K, et al. Building Eora: a global multi-region input–output database at high country and sector resolution[J]. Economic Systems Research, 2013, 25(1): 20-49.
[22] Lenzen M, Peters G M. How city dwellers affect their resource hinterland[J]. Low Carbon Cities: Transforming Urban Systems, 2014: 131.
[23] 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.

[24] Ma H, Wang Z, Zhu K. Domestic content in China’s exports and its distribution by firm ownership[J]. Journal of Comparative Economics, 2015, 43(1): 3-18.

[25] Meng B, Liu Y, Andrew R, Zhou M, Hubacek K, Xue J, Peters G, and Gao Y. More than half of China’s CO_2 emissions are from micro, small and medium-sized enterprises[J]. Applied Energy, 2018, 230: 712-725.

[26] Meng B, and Ye M. Smile curves in global value chains: multinationals vs domestic firms; the US vs China[R]. Institute of Developing Economies, Japan External Trade Organization (JETRO), 2020.

[27] Moran D, Wood R. Convergence between the Eora, WIOD, EXIOBASE, and OpenEU's consumption-based carbon accounts[J]. Economic Systems Research, 2014, 26(3): 245-261.

[28] Peters G P, Andrew R, Lennox J. Constructing an environmentally-extended multi-regional input–output table using the GTAP database[J]. Economic Systems Research, 2011, 23(2): 131-152.

[29] Peters G P. Efficient algorithms for life cycle assessment, input-output analysis, and Monte-Carlo analysis[J]. The International Journal of Life Cycle Assessment, 2007, 12(6): 373-380.

[30] Rypdal K, Winiwarter W. Uncertainties in greenhouse gas emission inventories—evaluation, comparability and implications[J]. Environmental Science & Policy, 2001, 4(2-3): 107-116.

[31] Schulte S, Jakobs A, Pauliuk S. Relaxing the import proportionality assumption in multi-regional input–output modelling[J]. Journal of economic structures, 2021, 10(1): 1-21.

[32] Stadler K, Wood R, Bulavskaya T, et al. EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables[J]. Journal of Industrial Ecology, 2018, 22(3): 502-515.

[33] Tang H, Wang F, Wang Z. Extending the input-output table based on firm-level data[J]. Available at SSRN 2764787, 2016.

[34] Torres-González L D, Yang J. The persistent statistical structure of the US input–output coefficient matrices: 1963–2007[J]. Economic Systems Research, 2019, 31(4): 481-504.

[35] Weber C L. Uncertainties in constructing environmental multiregional input-output models[C]//International input-output meeting on managing the environment. 2008, 10.

[36] West G R. Approximating the moments and distributions of input-output multipliers[M]. University of Queensland, Department of Economics, 1982.

[37] West G R. A stochastic analysis of an input-output model[J]. Econometrica: Journal of the Econometric Society, 1986: 363-374.

[38] Wiedmann T. A review of recent multi-region input–output models used for consumption-based emission and resource accounting[J]. Ecological economics, 2009, 69(2): 211-222.

[39] Wilting H C. 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.

[40] Yamakawa A, Peters G P. Using time-series to measure uncertainty in environmental input–output analysis[J]. Economic Systems Research, 2009, 21(4): 337-362.

[41] Yang C, Dietzenbacher E, Pei J, Chen X, Zhu K, and Tang Z. Processing trade biases the measurement of vertical specialization in China[J]. Economic Systems Research, 2015, 27(1):

60-76.
[42] Zhang Z, Guan D, Wang R, Meng J, Zheng H, Zhu K, and Du H. Embodied carbon emissions in the supply chains of multinational enterprises[J]. Nature Climate Change, 2020b, 10(12): 1096-1101.

Appendix

Schedule 1: ICIO-DF Table Sector Classification

Code	Departments
A	Agriculture, forestry and fishing
B	Mining and extraction of energy producing products
C10T12	Food products, beverages and tobacco
C13T15	Textiles, wearing apparel, leather and related products
C16	Wood and products of wood and cork
C17T18	Paper products and printing
C19	Coke and refined petroleum products
C20T21	Chemicals and pharmaceutical products
C22	Rubber and plastic products
C23	Other non-metallic mineral products
C24	Basic metals
C25	Fabricated metal products
C26	Computer, electronic and optical products
C27	Electrical equipment
C28	Machinery and equipment, nec
C29	Motor vehicles, trailers and semi-trailers
C30	Other transport equipment
C31T33	Other manufacturing; repair and installation of machinery and equipment
DTE	Electricity, gas, water supply, sewerage, waste and remediation services
F	Construction
G	Wholesale and retail trade; repair of motor vehicles
H	Transportation and storage
I	Accommodation and food services
J58T60	Publishing, audiovisual and broadcasting activities
J61	Telecommunications
J62T63	IT and other information services
K	Financial and insurance activities
L	Real estate activities
MTN	Other business sector services
Or	Public admin. and defence; compulsory social security
P	Education
Q	Human health and social work
RTS	Arts, entertainment, recreation and other service activities
T	Private households with employed persons