Assessment Study for the RAS Method based on China's Input-Output Tables

Xu Jian
He Jing
Zhao Yanyun

Management School of Graduate University

Chinese Academy of Sciences
100080, Beijing(China)

E-mail: xujian@gucas.ac.cn

Abstract

In China, RAS method is used to construct the extended input-output table by National Bureau of Statistics and is also usually used to update or balance input-output table by scholars because of its good maneuverability. Consequently, the test for the accuracy of RAS method has attracted the attention from many researchers. Though a great deal of achievements have been obtained on this issue, there still exists some serious problem. Firstly, the standard test procedures have a decline of overestimating accuracy. The paper proposes two designs to do a comparative analysis of real accuracy and upper-limit accuracy and addresses the extent of overestimation. Secondly, in literatures, the comparison between modified RAS and standard RAS were usually made. A basic conclusion is that modified RAS is more reliable than standard RAS. The paper will research where the improvement of accuracy is from. It concerns the DF-error relationship in nature. In order to draw reliable conclusions, we design 13 groups of pre-identified coefficients that have different number. The most important two conclusions are: 1) the reduction of DF, namely the increase of pre-identified coefficients, is able to decrease overall error reliably but to a large extent the decrease of overall error should be attributed to zero-error of pre-identified coefficients; 2) there exists threshold effect of DF-error relationship. Thus for modified RAS method, the accuracy of those unknown coefficients (excluding pre-identified coefficients) will have not improvement if pre-identified coefficients can not reach necessary number, which at least should account for 50% of the number of all coefficients.

Key words: RAS; accuracy; pre-identified coefficients

Assessment Study for the RAS method based on China's Input-Output Tables

Xu Jian He Jing Zhao Yanyun

1.Introduction

Input-output analysis plays an important role in researching economic structure. However, because constructing a complete table requires heavy input in term of both money and time, input-output tables are usually compiled for every few year. In China, National Bureau of Statistics compiles full-survey tables every 5 years and these tables are published in 2-3 years after the compiling-table year. So, in order to improve the timeliness of tables, IO researchers have focused on the non-survey or semi-survey techniques for updating and constructing IO tables.

Among the diversified non-survey techniques, RAS method, proposed by Stone in 1960s, has had the broadest application because of its good maneuverability. Consequently, the test for the accuracy of RAS method has attracted the attention from many researchers.

Availability of two or more IO tables for a majority of nations have promoted and allowed for empirical testing for RAS and other methods and a great deal of achievements have been obtained. But, McMenamin and Haring（1974）pointed out that all of the empirical evaluations of RAS method use actual vectors of target year as control vectors, but in practice, these vectors also need be estimated, so the accuracy measures presented represent an upper limit to the accuracy attainable. Their perspective holds true up to today. The only supplement we can do is that the upper limit is impossible to reach because the error of control variables will be not zero forever.

The same problem exists in the test of the accuracy of modified RAS. Most people have an opinion that modified RAS is more reliable than standard RAS. Moreover, we want to know where the improvement of accuracy is from. Lynch(1979) uses IO tables for the UK to compare modified RAS with standard RAS and draw a conclusion that modified RAS method improves the overall accuracy but has no effect on those non-identified elements. It tells us a neglected fact that the improvement of accuracy are partly attributed to test method that use actual value as the value of pre-identified elements. When we do so, the error of all pre-identified elements will be zero. As a sequence, it is certain that the overall accuracy get improvement.

But in practice those pre-identified elements have the estimation error, so the evaluation result will be bias.

The paper tries to answer the following questions: 1) what is the extent to which the accuracy of RAS are overestimated; 2) whether could modified RAS improve the estimation of those non-identified elements.

This paper is organized as follows. Following this introduction, we introduce the concept and measure indicator used in this paper. In Section 3, the first question is researched. In Section 4, we project two sets of pre-identified coefficients to analyze DF-error relationship. Following Section is conclusion.

2.Concept and Measure Indicator

In non-survey input-output literatures, there are some different definitions for accuracy. The most basic two definitions are partitive accuracy and holistic accuracy. The former focuses on the cell-by-cell accuracy, the latter focuses on the capability that updating matrix represent really economic structure. The detail discussion of this problem can be found in Jensen's (1980) article. The concept of accuracy will decide how to measure it. In this paper, the accuracy is confined to partitive accuracy. Therefor, error is equal to the distance between target matrix and estimated matrix.

There are many indictors which have been broadly applied to evaluate partitive accuracy in literatures, such as STPE(Standardized Total Percentage Error), SMAD(Standardized Mean Absolute Difference), DSI(Dissimilarity Index), MIC(Mean Information Content), RMSE(Root Mean Square Error) and MAPE(Mean Absolute Percentage Error). In this paper, STPE will be used as a major measure indictor of error because of its stability, see Szyrmer(1989). STPE can be defined as follows:

$$STPE = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}^{1} - a_{ij}^{0}\right| \Big/ \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^{1}$$

a_{ij}^{1} and a_{ij}^{0} represent element of target matrix and estimated matrix *respectively*

For a matrix whose marginal totals are fixed, degree of freedom (DF) expresses the number of free coefficients. DF is equal to (n-1)×(k-1)-h, n, k and h represent the number of row, the number of column and the number of fixed cells respectively. It should be noted that a column or row whose elements are all zero should not be included and a fixed cell lying in zero column or row should not be included yet.

We could have a more in-depth understanding on what is real accuracy of modified RAS and whether modified RAS could improve the estimation of those non-identified elements by analyze Degree of Freedom-Updating Error relationship.

3.A comparative analysis of real accuracy and upper-limit accuracy

Experimental design

Firstly, We use 17-sector direct input coefficient matrix for 1997 as the target matrix and use actual gross output, intermediate input and intermediate output from 1997 IO table as marginal constraints. Applying RAS algorithm to 1992 direct input coefficient matrix, we can obtain updating matrixes at the base of which STPE can be computed. Obviously, the STPE obtained from the above design is a measurement of upper-limit accuracy or lower-limit error. In this paper, its notation is $STPE_U$.

Secondly, we firstly estimate gross output, intermediate input and intermediate output for 1997 by all available non-IO data in 1997 as if 1997 IO table were not published. Then RAS is applied to update 1992 table to 1997 with the estimated results as control. When we obtain updating matrixes, another STPE can be computed. The above design simulates the real process of running RAS, so the STPE is a measurement of real accuracy or real error we have to face. In this paper, its notation is $STPE_R$.

The estimation process of gross output, intermediate input and intermediate output for 1997 is as follows:

(1) Obtaining the estimation of total output of agriculture sector and the secondary industry sectors

Though there is no IO table, we still get total output of agriculture sector and the secondary industry sectors for 1997 from China Statistical Yearbook 1998 published in 1998. But in China, There are two important differences between the data from the China Statistic Yearbook and the one we want: ① The yearbook's data is based on industry while the total output in the IO table is based on commodity. ② The yearbook's data was collected from firms whose sale was more than 5,000,000 RMB a year while the total output of IO table includes all the firms, so we can not use the yearbook's data directly as the controlling value.

Detailed estimation procedure is:

① Collecting the data of the total output for 1992 and 1997 from the China Statistical Yearbooks.

② Comparing the total output for 1992 from the China Statistical Yearbook with and the total output from IO table for 1992 , we can get a set of adjustment coefficients.

③ The total output for 1997 from the China Statistical Yearbooks multiplied by the above adjustment coefficients, we can get the estimation of total output of agriculture sector and the secondary industry sectors satisfying the demand of IO table.

(2) Obtaining the estimation of value added of agriculture sector and the secondary industry sectors

Using the same procedure as total output, value added consistent with the scope of IO table could be obtained.

(3) Obtaining the estimation of both total output and value added of tertiary industry sectors

Different with agriculture sector and the secondary industry sectors, only value added can be obtained from China Statistical Yearbook for tertiary industry. So estimation procedure has a little different:

① Collecting the data of the value added for 1992 and 1997 from the China Statistical Yearbooks.

② Comparing the value added for 1992 from the China Statistical Yearbook with and the value added from IO table for 1992 , we can get a set of adjustment coefficients.

③ The value added for 1997 from the China Statistical Yearbooks multiplied by the above adjustment coefficients, we can get the final estimation of value added of tertiary industry sectors.

④ Calculating value added coefficients for 1992 based on 1992 IO data.

⑤ The results of step 3 divided by value added coefficients for 1992, we can get a estimation of total output for 1997.

(4) Obtaining the estimation of total intermediate input

The difference between total output and value added is just total intermediate input.

(5) Calculating final output

All final demands has the same calculation procedure, so we only take household consumption expenditure as example to illustrate the procedure.

① Collecting the data of household consumption expenditure for 1992 and 1997 from the China Statistical Yearbooks.

② Comparing the household consumption expenditure for 1992 from the China Statistical Yearbook with and that from IO table for 1992 , we can get a adjustment coefficients.

③ The household consumption expenditure for 1997 from the China Statistical Yearbooks multiplied by the above adjustment coefficients, we can get the final estimation as control.

④ Multiplying the allocation coefficient of household consumption expenditure for 1992 with corresponding total output for 1997, we could get the household consumption expenditure column for 1997.

⑤ By adjustments to assure that the sum of household consumption expenditure column is equal tothe results of step 3, we can get the final estimation of household consumption expenditure.

(6) Obtaining the estimation of total intermediate output

The difference between total output and final output is just total intermediate output.

Results Analysis

$STPE_U$ is 0.2393; $STPE_R$ is 0.2619; STPE of no updating is 0.2816.

As far as the value is concerned, there is no significant difference between $STPE_U$ and $STPE_R$. But, if we notice that STPE is only 0.2819 without any update done, we have to say that $STPE_U$ overestimates the benefit of RAS method to a relatively great extent.

Farther analysis draws two conclusions:

1) $STPE_R$ has the increasing trend with the increase of the estimation error of control variables. It means that directly using input coefficients for base year as the estimation of those for target year will be better choice than applying RAS if the estimation of control value are not able to be done properly.

2) The error of control value is distributed uniform in the whole direct input coefficients matrix. It means that there is no strong correlation relationship between the error of control value by a certain sector and the errors of direct input coefficients by this sector.

4. Degree of Freedom and Updating Error

Experimental design

In order to obtain enough representative samples, we design two sets of pre-identified coefficients at the base of 1987 constant price 18-sector direct input coefficient matrix for China,. Firstly, according to size of direct input coefficient, we chose seven groups of coefficients as pre-identified coefficients. The range of every group is listed in first column of table 1. Secondly, another six groups of pre-identified coefficients are randomly selected; the number of pre-identified coefficients is 25,50,100,150,200,250 respectively.

We use 1992 direct input coefficient matrix for China as target matrix and use actual gross output, total intermediate input and total intermediate output of 1992 as marginal constraint. Modified RAS algorithm is applied to 13 base matrixes, in which pre-identified coefficients are set equal to zero in iteration process and are replaced with the corresponding actual coefficients for 1992 after iteration process ends, and estimated matrix can be obtained. Comparing estimated matrix with target matrix, we can compute two kinds of STPE. STPE1 refers to mean error of all coefficients. STPE2 only refers to mean error of unknown coefficients. The latter can be defined as follows:

$$\mathrm{STPE2} = \sum\sum\left|a_{ij}^{1} - a_{ij}^{0}\right| \Big/ \sum\sum a_{ij}^{1}, \quad a_{ij} \text{ excluding pre-identified coefficients.}$$

In order to more precisely evaluate the effect of DF on updating error, we ulteriorly computed STPE3 which refers to mean error of those corresponding unknown coefficients when standard RAS algorithm is applied, namely no pre-identified coefficient.

Results Analysis

Table 1. STPE using base matrixes whose pre-identified coefficients are determined by the size of coefficients

size of direct input coeff.	degree of freedom	number of pre-identified	STPE1	STPE2	STPE3
>0.1	264	25	0.1534	0.2610	0.2577
>0.05	227	62	0.1075	0.2974	0.2858
>0.02	165	124	0.0570	0.3558	0.3858
>0.01	115	174	0.0208	0.3090	0.4308
>0.005	84	210	0.0105	0.3104	0.4565
>0.002	39	258	0.0024	0.3192	0.5410
>0.001	16	285	0.0011	0.3192	0.6357

Table 2. STPE using base matrixes whose pre-identified coefficients are randomly selected

degree of freedom	number of pre-identified	STPE1	STPE2	STPE3
289	0	0.2052	0.2052	0.2052
264	25	0.1906	0.2073	0.2072
239	50	0.1776	0.2018	0.2018
189	100	0.1510	0.2163	0.2140
139	150	0.1318	0.2192	0.2226
89	200	0.0793	0.1839	0.1660
40	250	0.0385	0.2377	0.3417

From table 1 and table 2, we can find STPE1 obviously tends to decrease and STPE2 still keep stability with the reduction of DF. This result enable us to draw a conclusion that the reduction of DF is able to decrease overall error reliably but to a large extent the decrease of overall error should be attributed to zero-error of pre-identified coefficients.

The function of RAS method is to obtain the estimation of unknown coefficients at very low cost, moreover, the error of pre-identified coefficients depends on the survey and is independent to RAS method in practice. Thus evaluating or comparing the accuracy of RAS method should regard the unknown coefficients as object. In this sense, STPE2 is more proper measure indicator than STPE1, and STPE3 is also a proper counterpart. The effect of DF on error can be evaluated by comparing STPE2 with STPE3.

For first set of base matrix, STPE2 is obviously smaller than STPE3 when DF is reduced to 115. Therefor, only when known coefficients reach enough number, at least exceed 124, could the reduction of DF improve the accuracy of those unknown coefficients obviously. Otherwise, there is no correlation between DF and the error of unknown coefficients. Then, there exists threshold effect of DF-error relationship. For second set of base matrix, there also exists similar relationship, the major difference lies in the position of threshold. Table 4 shows: Until DF is reduced to 40, STPE2 is obviously smaller than

STPE3.

5. Conclusions

In a great number non-survey and semi-survey methods, RAS method has the broadest applications. Accordingly, empirical evaluation of RAS method also receives broad attentions. However, the standard test procedures have a decline of overestimating accuracy. As far as the standard RAS is concerned, its error is not only from the error of biproportional algorithms but also from both the error of control value. However, test procedure set the latter as zero. As far as the modified RAS is concerned, its error is not only from the error of algorithms but also from both the error of pre-identified elements. However, test procedure also set the latter as zero.

The paper gives estimations of the extent that accuracy is overestimated and draws some important conclusions. It could help users of RAS to have a more accurate recognition on the reliability of this famous method.

References

[1] Ali, Reza Jalili, Comparison of Two Methods of Identifying Input-Output Coefficients for Exogenous Estimation, Economic Systems Research, Vol. 12, No. 1, pp113-129, 2000

[2] Gilchrist, Donald A. and Larry V. St Louis, Completing Input-Output Tables using Partial Information, with an Application to Canadian Data, Economic Systems Research, Vol. 11, No. 2,pp185-194, 1999

[3] Jensen, R. C., The Concept of Accuracy in Regional Input-Output Models, International Regional Science Review, Vol. 5, No. 2, pp139-154, 1980

[4] Lynch, R. G. , An Assessment of the RAS Method for Updating Input-Output Tables, In Readings in Input-Output Analysis: Theory and Applications, edited by Ira Sohn, New York: Oxford University Press, pp271-284, 1979

[5] Malizia, Emil and Daniel L. Bond, Empirical Tests of the RAS Method of Interindustry Coefficient Adjustment, Journal of Regional Science, Vol. 14, No. 3, pp355-365,1974

[6] McMenamin, David G. and Joseph E. Haring, An Appraisal of Nonsurvey Techniques for Estimating Regional Input-Output Models, Journal of Regional Science, Vol. 14, No. 2, pp191-205,1974

[7] Robinson, Sherman, Andrea Cattaneo and Moataz El-said, Updating and Estimating a Social Accounting Matrix using Cross Entropy Methods, Economic Systems Research, Vol. 13, No. 1,pp47-63, 2001

[8] Sonis, Michael and Geoffrey J. D. Hewings, Coefficient Change in Input-Output Models: Theory and Applications, Economic Systems Research, Vol. 4, No. 2, pp143-157, 1992

[9] Szyrmer Janusz, Trade-Off between Error and Information in the RAS Procedure, In Frontiers of Input-Output Analysis, edited by Ronald E. Miller, Karen R. Polenske and Adam Z. Rose, New York: Oxford University Press, pp258-277, 1989

[10] Toh, Mun-Heng, The RAS Approach in Updating Input-Output Matrices : An Instrumental Variable Interpretation and Analysis of Structural Change, Economic Systems Research, Vol. 10, No. 1,pp63-79, 1998