International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-07-015**

A new practical method for estimation of input-output tables

Shinichiro Fujimori (fshinichi@t23.mbox.media.kyoto-u.ac.jp)

Approved by

Marek Makowski
Leader, Integrated Modeling Environment Project

(May, 2007)

Foreword

This report describes the research results achieved by the author during the Young Scientists Summer Program (YSSP) 2006. Since the goals for this research were very ambitious, more computational experiments had been run after the author returned to Japan. Therefore, this version of the report contains also the results achieved after the 2006 YSSP.

The research reported in this paper is an element of the longer-term research by the author, which in turn is a part of a large-scale activity. This content of the work determined rather strict requirements for the reliability and efficiency of the estimation described in the report. The problem of estimation of input-output tables is not new: there are several methods and many publications documenting diverse approaches. However, none of the existing methods were adequate for the problem to be researched by the author.
Therefore is was necessary to develop a new method, which is built on the cross-entropy approach.
In order to meet the requirements for the long-term research the author has:

(1) extended the specification of the classical formulation of estimation of input-output tables to adequately analyze the material flow problem;
(2) developed a new method for effective implementation of the modified cross-entropy method to the actual case study; in particular the author implemented and combined two procedures for an adequate preprocessing of data for non-linear solvers, namely removing non-substantial elements, and scaling the non-linear optimization problem;
(3) tested the approach on representative real-case sets of data, and proved the effectiveness and efficiency of the developed method.

The paper reports intermediate results. However, if such intermediate results will be interesting for researchers and practitioners working on estimation of medium-size input-output tables characterized by: (1) incomplete information for some sectors, and (2) additional information for other sectors. Moreover, the proposed method has accuracy not as inferior than the documented approach but is dramatically faster (in terms of computational time), thus it is especially recommended for studies in which estimations of many input-output tables are required.

Finally, I wish to stress that the author achieved the impressive results not only during a very short time but also at the very beginning of his research carrier (just few months after receiving his MSc. degree). This is a very good sign for his future research.

Marek Makowski
Leader, Integrated Modeling Environment Project

Abstract

This paper describes a new and effective method for the estimation of input-output tables. The method is based on the cross-entropy method (Kullback, 1959). The cross-entropy method has been applied for the estimation of input-output tables (see e.g., Robinson et al., 2001; Golan et al., 1994). However, considering this, the known approaches were not effective to the problem. Therefore, we had to solve some methodological problems which in turn led to substantial improvements of the cross-entropy method for actual applications.

In the future, we are going to estimate a few decades' input-output tables. Thus, we applied the developed method to estimate input-output tables of Japan for twenty year. The results reported in this paper show that the proposed method not only provides correct results but is much more efficient, and therefore can be effectively used for estimating material flows for many countries and for the period of several decades.

Acknowledgments

I would like to thank Marek Makowski for his supervision and for granting me a lot of discussion time and appropriate suggestions. I am grateful to Tatiana Ermolieva and Yuri Ermoliev who taught me the mathematical method. In addition, I would like to thank Yuzuru Matsuoka of Kyoto University, Japan for many discussions and guidance in this research.

I would like to express my thanks to Aviott John and Natalia Ovchinnikova from the IIASA library for their valuable help in finding required literature. I would like to thank Deirdre Zeller and Kathryn Kenyon. Also, many thanks to Barbara Hauser, Tanja Huber, Joanne Linnerooth-Bayer, Matthias Jonas and Serge Medow who organized, coordinated and supported the 2006 Young Summer Scientists Program at IIASA.

About the Author

Shinichiro Fujimori was a participant of the 2006 Young Scientists Summer Program (YSSP) in the Integrated Modeling Environment project at IIASA. He received an engineering masters in Urban and Environmental Engineering from Kyoto University in 2006. Currently he is a Ph.D. candidate in the same department. He is researching global Material Flow Accounting and Analysis.

Contents

A new practical method for estimation of input-output tables

Shinichiro Fujimori*

1. Introduction

1.1 Background

Material Flow Analysis (MFA) builds on earlier concepts of material and energy balancing, as presented by Ayres (1978; Ayres and Kneese 1968). MFA is the analysis of the throughput of process chains extraction or harvest, manufacturing, consumption, recycling and disposal of materials. It is based on accounts in physical term quantifying the inputs and outputs of those processes. The first material flow accounts on the national level have been presented at the beginning of the 1990s for Austria (Steurer, 1992) and Japan (Environment Agency Japan, 1992). Since then, MFA was a rapidly growing field of scientific interest and major efforts have been undertaken to harmonize the different methodological approaches developed by different research teams (e.g. Matthews *et al.*, 2000). Previous works revealed that human society have a great deal of “ecological rucksacks” which are indirect flows that do not become part of a product but which are concomitant to its production and how indicators of sustainable development have been changed.

A Physical Input-Output Table (PIOT) provides one of the most comprehensive descriptions of anthropogenic resource flows. A PIOT describes the material and energy flows between the socio-economic system and the environment (thus providing the same information as economy-wide material flow accounts) and in addition the flows between the different sectors within an economic system. Furthermore the net-accumulation of materials in the economic system is accounted for (EUROSTAT 2001).

The concept of PIOT is based on the principles laid out in the “System of Integrated Environmental and Economic Accounts (SEEA)“ of the United Nations (1993, 2001). Together with MFA and energy accounts it forms the methodological core of physical flow accounting systems within the SEEA framework. However, there are only a few country’s application of PIOT (e.g. Stahmer *et al.*, 1997).

So far there are no comprehensive material flow accountings which cover the entire world including trade flows. Such data would substantially support policy making in numerous issues pertinent to material utilizations. Therefore, we developed a calculation method to estimate those flows and implemented them for 2001(Fujimori and Matsuoka, 2007).

---

* Kyoto University, Graduate school of Engineering, Department of Urban and Environmental Engineering, JAPAN, fshinichi@t23.mbox.media.kyoto-u.ac.jp

This method requires world countries input-output tables, trading matrix and material information. An input-output table depicts the transactions associated with the production processes of an economic system. Currently many countries (especially developed counties) have reliable input-output tables for a few decades. However, Global Trade Analysis Project (GTAP; Hertel, 2001, 2005), is one of the few input-output table databases covering the entire world. GTAP is available only for 1997 and 2001. In order to analyze problems related to material flows it is necessary to estimate annual material flow for a longer period for 20 or 30 years, which in turn requires estimation of annual monetary input-output tables for these years for which they are not available.

Some methods updating or developing input-output tables have been developed. RAS method (Bacharach, 1970) and Cross-entropy method (Robinson *et al.*, 2001; Golan *et al.*, 1994) are the major examples. Stone method (Stone *et al.*, 1942; Byron, 1978) is another example, which is proposed for the data reconciliation method. Despite a good background for estimating or reconciling methods, there have been some methodological problems that needed to be solved for an actual application (we will indicate these problems in Section 2.3).

1.2 The scope of the reported research

This paper deals with two problems. One is to provide a new and effective method for the estimation of input-output tables. The other is to assess the accuracy of the estimation and to suggest the method can be applied for a large set of countries, and for a long period (e.g, more than 20 years) for each country.

The method is based on the cross-entropy method (Kullback, 1959). As mentioned above, the cross-entropy method has been applied for the estimation of input-output tables (see e.g., Robinson et al., 2001; Golan et al., 1994). However, considering this the known approaches were not effective to our problem and they only applied the one year. Therefore, we had to solve some methodological problems which in turn led to substantial improvements of the cross-entropy method for actual applications. We applied the developed method to estimate input-output tables of Japan for 1985, 1980, 1975 and 1970, and proved that the method is effective for our purpose.

1.3 Structure of this paper

Section 2 introduces input-output tables and a definition of estimating input-output table problems. Then, we introduce the classical methods to estimate input-output tables. Section 3 indicates our approach to estimate input-output tables. The practical application and the results of the application will be discussed in Section 4. Finally, we will suggest the conclusion in Section 5.

2. Estimating input-output table

2.1 Input-output table

In this Section, we explain the concept of an input-output table. An input-output table represents the transactions associated with the production processes of an economic system. The input requirement of each production process is recorded, as well as, the output of products.

Table1 shows the general structure of an input-output table. The columns depict the input requirements of a production unit. Let *I* and *J* be the sets of indices of the row and the column. Some inputs are purchased from other production units: these are the "intermediate inputs", represented by matrix **Z1**[†]. The sets of column/row indices of **Z1** are denoted by J1 and I1, respectively; thus *I*1 and *J*1 are the subsets of *I* and *J* composed of production units. *I*1 and *J*1 is the same classification. Inputs that are not produced by other production units are called "primary inputs" (matrix **Z2**). For example, the primary inputs are labor and capital depreciation. The sets of column/row indices of **Z2** are denoted by I2 and J1. *I*2 is the subset of *I* composed of primary inputs. The rows indicate the destination of the outputs of the production units. The products which are not purchased by production units are supplied to "final demand" (matrix **Z3**). This includes categories such as private consumption, government consumption, capital formation, imports and exports. The sets of column/row indices of **Z3** are denoted by I2 and J2. *J*2 is the subset of *J* composed of final demand. Final demand doesn't have primary inputs, thus the lower-right corner of the table is composed of elements equal to **0**. **a1** denotes the total output from production units *i*1 and **a2** is the total of primary inputs *i*2. Vector **b1** denotes the total input of production units J1 (equation (2.1)) and **b2** is the total input of final demand J2 (equation (2.1)).

Table 1 Structure of an input-output table

	Production units (J1)	Final demand (J2)	Total output
Production units (I1)	**Z1**	**Z3**	**a1**
Primary inputs (I2)	**Z2**	**0**	**a2**
Total inputs	**b1**	**b2**	

[†] Throughout the paper a bold upper-case letter denotes a corresponding matrix with elements denoted by the corresponding lower-case letter. The subscripts *i* and *j* denote row and column indices, correspondingly.

The Input and output of each production units is balanced; this is represented by (2.3).

$$\sum_{j\in J1} z1_{i,j} + \sum_{j\in J2} z3_{i,j} = \mathbf{a1} \qquad (2.1)$$

$$\sum_{i\in I2} z1_{i,j} + \sum_{i\in I2} z2_{i,j} = \mathbf{b1} \qquad (2.2)$$

$$\mathbf{a1} = \mathbf{b1} \qquad (2.3)$$

The classical definition of the input-output table is composed of only **Z1** matrix. In this paper we extended the classical approach by adding matrices **Z2** and **Z3**.

2.2 Specification of the input-output table estimation problem

We define how the problem of estimation of input-output table. For solving the estimation problem we introduce four matrices: **X**, **Q**, **Y** and **P**.

The matrix **X** is a given the base year input-output table, and it is conventionally called the prior matrix because it is used for estimation of input-output tables for the years for which data is not available. *I* and *J* are sets of indices of rows and columns, and therefore the numbers of their elements are equal to m and n, respectively.

The matrix **Q** is composed of normalized elements of **X**, i.e.

$$q_{i,j} = \frac{x_{i,j}}{\sum_{i\in I} x_{i,j}}, \quad i \in I,\ j \in J \qquad (2.4)$$

The matrix **Y** is composed of estimated values of the input-output table, and it is defined for each year for which the estimation is to be done. The corresponding matrix **P** is composed of normalized elements of **Y**, i.e.

$$p_{i,j} = \frac{y_{i,j}}{b_j}, \quad i \in I,\ j \in J \qquad (2.5)$$

In other words the problem is to find estimate the elements of the matrix **Y** using the prior matrix **X**. However, this is done by solving an auxiliary problem, i.e., to estimate elements of the normalized matrix **P** from the given normalized matrix **Q**.

Table 2 Structure of an input-output table

	Production units (J1)	Final demand (J2)	Total output
Production units (I1)	**Z1**	**Z3**	**a1**
Primary inputs (I2)	**Z2**	**0**	**a2**
Total inputs	**b1**	**b2**	

Let us denote a sum of row elements by **a** and **b**. We assume that the matrix **X**, and vectors **a** and **b** are known (see Tables 3 and 4 for illustration). The problem is to estimate values of the matrix $y_{i,j}$ that fulfill the following conditions:

$$a_i = \sum_{j \in J} y_{i,j} \,, \quad i \in I \tag{2.6}$$

$$b_j = \sum_{i \in I} y_{i,j} \,, \quad j \in J \tag{2.7}$$

Table 3 Known values in prior matrix (basic year)

	Production units	Final demand	Total output
Production	**X**		
Primary inputs			
Total inputs			

Table 4 Known values in estimated matrices

	Production units	Final demand	Total output
Production			**a**
Primary inputs			
Total inputs	**b**		

2.3 Relevant methods

Further on, we concrete on describing the methods dealing with the auxiliary problem, i.e., estimation of elements of the normalized matrix **P** from the given normalized matrix **Q**. There are two major methods for such estimations; RAS method and cross-entropy method. The RAS method is proposed by Bacharach

(1970). Though cross-entropy method is originally proposed by Kullback (1959), its application to estimating input-output table is indicated in Golan *et al.* (1994).

2.3.1 RAS method

The RAS method is a basic method to estimate an input-output table when we have information on the row and column sum, but do not have information on the input-output table. The RAS problem can be presented as finding the vectors **r** and **s** such that,

$$\mathbf{P} = \mathbf{r} \cdot \mathbf{Q} \cdot \mathbf{s} \tag{2.8}$$

subject to equation (2.4), (2.5) (2.6) and (2.7). Where **r** and **s** are diagnosed matrices composed of factors adjusting elements of each respective row and column.

This method amounts to a successive biproportional adjustment of the rows and columns of the base matrix **Q**, until convergence is reached.

2.3.2 *Cross-entropy method*

2.3.2.1 *Basic formulation*

Cross-entropy method was opposed by in Kullback (1959) for downscaling. This method can be applied to the problem which has prior information. Cross-entropy method uses the following function

$$\sum_{j\in J}\sum_{i\in I} p_{i,j} \ln \frac{p_{i,j}}{q_{i,j}} \qquad \left(q_{i,j} > 0\right) \tag{2.9}$$

This function is called Kullback-Leibler distance between distributions of **Q** and **P**; the function

$$-\sum_{j\in J}\sum_{i\in I} p_{i,j} \ln \frac{p_{i,j}}{q_{i,j}} \qquad \left(q_{i,j} > 0\right) \tag{2.10}$$

is called the cross-entropy. Under the principle of minimum discriminability the distance is minimized. Consequently, the problem of estimation of **P** can be formulated as:

$$\min \sum_{j\in J}\sum_{i\in I} p_{i,j} \ln \frac{p_{i,j}}{q_{i,j}} \qquad \left(q_{i,j} > 0\right), \tag{2.11}$$

subject to (2.4), (2.5), (2.6) and (2.7).

2.3.2.1 *Additional constraints*

In addition to constraints (2.6) and (2.7) which reflect the given balances for rows and columns of the input-output table, often other information about the input-output table is available. For example, we can get the sector's value add data. This information can be summarized as additional linear adding-up constraints on various elements (indicated in Robinson *et al.*(2001)). Let us assume that there are *k* such constraints, which given by:

$$\sum_{j\in J}\sum_{i\in I} g^k_{i,j}\cdot y_{i,j} = d^k \qquad k\in K \tag{2.12}$$

where d^k is the additional value and $g^k_{i,j}$ has ones for cells in the aggregated data and zeros otherwise. K is the set of indices of the additional constraints. For example, let us assume we could get the additional information d^{k1} described in equation (2.13). Then, we set $g^{k1}_{i,j}$ be as equation (2.14).

$$\sum_{j\in J^{k1}}\sum_{i\in I^{k1}} y_{i,j} = d^{k1} \qquad I^{k1}\subseteq I\,,\quad J^{k1}\subseteq J \tag{2.13}$$

$$\begin{cases} g^{k1}_{i,j} = 1 & i\in I^{k1},\ j\in J^{k1} \\ g^{k1}_{i,j} = 0 & i\in I - I^{k1},\ j\in J - J^{k1} \end{cases} \tag{2.14}$$

3. Methodological problems and solutions

The cross-entropy method has been applied to estimating input-output tables, (e.g. Golan *et al.* (1994), Robinson *et al.* (2001)). However, there have been some methodological problems in applying it to the problem defined in Section 2.3. Therefore, we reformulate the cross-entropy method as shown in 3.1. In order to deal with large matrices and to shorten the calculation time, we developed two specialized methods for removing non-substantial values, and for scaling matrix respectively. These methods are presented respectively in Section 3.2 and 3.3, respectively.

3.1 Reformulating the cross-entropy method

3.1.1 Dealing with negative values

We are aiming estimate input-output table. Input-output tables have negative values (e.g. import value). The cross-entropy function rejects negative values because of the logarithm. Therefore, we deal with negative values by using $\mu_{i,j}$. $\mu_{i,j}$ has one for positive value in prior matrix **X** and has negative one for negative value in prior matrix **X** (equation (3.1)). Equation (2.6), (2.7) and (2.12) are reformulated as;

$$\begin{cases} \mu_{i,j} = 1 & x_{i,j} \geq 0 \\ \mu_{i,j} = -1 & x_{i,j} < 0 \end{cases} \tag{3.1}$$

$$q_{i,j} = \frac{x_{i,j}\cdot\mu_{i,j}}{\sum_{j\in J} x_{i,j}\cdot\mu_{i,j}}\ ,\quad i\in I,\ j\in J\ . \tag{3.2}$$

$$\sum_{j\in J}\mu_{i,j}\cdot y_{i,j} = a_i,\quad i\in I \tag{3.3}$$

$$\sum_{i \in I} \mu_{i,j} \cdot y_{i,j} = b_j, \quad j \in J \tag{3.4}$$

$$\sum_{j \in J} \sum_{i \in I} g_{i,j}^k \cdot \mu_{i,j} \cdot y_{i,j} = d^k \quad k \in K \tag{3.5}$$

3.1.2 Missing values

As we indicated in Section 1.1, we are going to estimate input-output tables covering the world. Though some international statistics estimate missing values (e.g. International Energy Agency (2004)), many international statistics have missing values (e.g. United Nations Industrial Development Organization (2006)). Consequently, we cannot possibly get the data for all of column and row sums (**a** and **b**), especially it is difficult to get service sector's total output. Therefore, we reformulate the estimation method for estimating even if there are some missing values in column and row sums. We have to change the definition of $p_{i,j}$ as equation (3.6)

$$p_{i,j} = \frac{y_{i,j}}{\sum_{j \in J} y_{i,j}} \ , \quad i \in I, \ j \in J \ . \tag{3.6}$$

Even if some elements of **b** are missing, $p_{i,j}$ can be defined.

3.1.3 Statistical errors

We deal with the noise in economical statistics like shown in Robinson *et al.* (2001). Firstly, assume that parameters **a**, **b** and **d** have statistical errors $\mathbf{e^1}$, $\mathbf{e^2}$ and $\mathbf{e^3}$. Thus we reformulate equation (3.3), (3.4) and (3.5) as;

$$\sum_{j \in J} \mu_{i,j} \cdot y_{i,j} + e_i^1 = a_i, \quad i \in I \tag{3.7}$$

$$\sum_{i \in I} \mu_{i,j} \cdot y_{i,j} + e_j^2 = b_j, \quad j \in J \tag{3.8}$$

$$\sum_{j \in J} \sum_{i \in I} g_{i,j}^k \cdot \mu_{i,j} \cdot y_{i,j} + e_k^3 = d^k \quad k \in K \tag{3.9}$$

Where $\mathbf{e^1}$, $\mathbf{e^2}$ and $\mathbf{e^3}$ are statistical errors. Following Golan *et al.* (1996), the errors are defined as weighted averages of known constants;

$$e_i^1 = \sum_{h=1}^{h} w_{i,h}^1 v_{i,h}^1 \ , \quad i \in I \ , \tag{3.10}$$

$$e_j^2 = \sum_{h=1}^{h} w_{j,h}^2 v_{j,h}^2 \ , \quad j \in J \ , \tag{3.11}$$

$$e_k^3 = \sum_{h=1}^{h} w_{k,h}^3 v_{k,h}^3 , \qquad k \in K . \tag{3.12}$$

Where $v_{i,h}^1$, $v_{j,h}^2$ and $v_{k,h}^3$ are constants and defined a prior for the error distribution. **h** is the set of weights, **w**. $w_{i,h}^1$, $w_{j,h}^2$ and $w_{k,h}^3$ are variables. Reformulate the objective function as following,

$$\min \quad \sum_{j \in J} \sum_{i \in I} p_{i,j} \ln \frac{p_{i,j}}{q_{i,j}} + \sum_{h=1}^{3} \left(\sum_{i \in I} w_{i,h}^1 \ln \frac{w_{i,h}^1}{u_{i,h}^1} + \sum_{j \in J} w_{j,h}^2 \ln \frac{w_{j,h}^2}{u_{j,h}^2} + \sum_{k \in K} w_{k,h}^3 \ln \frac{w_{k,h}^3}{u_{k,h}^3} \right) \quad ,(3.13)$$

Where $u_{ijk,h}^o$ is prior weight of $w_{ijk,h}^o$. Each sectors input and output is balanced as following.

$$\sum_{j \in J} y_{i,j} = \sum_{j \in J} y_{j,i} , \qquad i \in I1 \tag{3.14}$$

*I*1 is the subset of ***I*** composed of production units as defined in Section 2.1.

3.2 Removing non-substantial values

Good modeling practice requires that only substantial elements are included in the model specification, i.e. constraints and objective function. Substantial in this context means having large enough values. Including non-substantial elements may make effective scaling (see the following Section) impossible, which in turn is likely to cause numerical problems. Therefore relatively small values should be removed from the matrix **Q**.

To identify the non-substantial elements of the matrix **Q** we should evaluate their values with respect to values of the other elements in the same row and the same column. For this we consider not only $q_{i,j}$ but also $t_{i,j}$. $t_{i,j}$ is defined as,

$$t_{i,j} = \frac{x_{i,j}}{\sum_{j \in J} x_{i,j}} \qquad i \in I, \ j \in J . \tag{3.15}$$

The values which are non substantial in both $q_{i,j}$ and $t_{i,j}$ should be removed. This criterion is as following;

$$t_{i,i} < \varepsilon \quad and \quad q_{i,i} < \varepsilon \qquad i \in I, \ j \in J . \tag{3.16}$$

Where ε is the criterion for removing non-substantial values. In practical application (Section 4) we assume ε equal to 0.0001

3.3 Scaling method

3.3.1 Reformulate the original problem into scaled problem

As I specified in the Section 3.1, this method is a non-linear one. We used GAMS and CONOPT as the solver in the calculation procedure. As opposed to linear problems routinely scaled by solvers, non-linear problems have to be well scaled by users (GAMS Development Corporation, 2005). Values of matrix **Q** elements differ by several of magnitude orders, i.e. the ratio of the maximal to the minimal value of non-zero elements can be larger than 10^6 while non-linear solvers require that such a ratio should be smaller than 10^4.

In order to properly define our optimization problem for the non-linear solver we had to adapt and implement scaling of the matrix **Q**. A scaling consists of finding the values of two vectors **rs** and **cs**, elements of which are the scaling coefficients of the matrix **Q**, for rows and columns, respectively. Elements of the scaled matrix **Q'** are defined by the equation (3.17).

$$q'_{i,j} = q_{i,j} \cdot \frac{1}{rs_i} \cdot \frac{1}{cs_j} \quad , \quad i \in I,\ j \in J \tag{3.17}$$

where rs_i and cs_j are scaling coefficient ($rs_i > 0,\ cs_j > 0$).

By applying the scaling, the objective function (3.13) of the original non-linear optimization problem (equation (3.2), (3.6), (3.7), (3.8), (3.9), (3.10), (3.11), (3.12), (3.13), (3.14)) is replaced by following:

$$\min \quad \sum_{j \in J} \sum_{i \in I} p'_{i,j} \ln \frac{p'_{i,j}}{q'_{i,j}} + \sum_{h=1}^{3} \left(\sum_{i \in I} w_{i,h}^{1} \ln \frac{w_{i,h}^{1}}{u_{i,h}^{1}} + \sum_{j \in J} w_{j,h}^{2} \ln \frac{w_{j,h}^{2}}{u_{j,h}^{2}} + \sum_{k \in K} w_{k,h}^{3} \ln \frac{w_{k,h}^{3}}{u_{k,h}^{3}} \right) \tag{3.18}$$

The solution of the above problem (denoted by **P'**) has to be rescaled to provide a solution of the original problem, i.e. the matrix **P**. This is done by:

$$p_{i,j} = p'_{i,j}\, rs_i \cdot cs_j \quad , \quad i \in I,\ j \in J \tag{3.19}$$

3.3.2 Algorithm

We describe the algorithm to obtain scaling coefficient rs_i and cs_j by the following iterative procedure. We refer to Makowski M. and J. Sosnowski(1981) [22] for this algorithm.

Step 1:

Let *s* be the index of scaling iteration, initiated by 0.

Set initial values to $q_{i,j}^{0} = q_{i,j}$ and $rs_i^0 = 1,\ cs_j^0 = 1$.

Step 2:

Calculate the maximum value α_i^s and minimum non-zero value β_i^s in each row *i* as equation (3.20) and (3.21). Calculate the square root of maximum and minimum value; τ_i (3.22).

$$\alpha_i^{s+1} = \max_j q_{i,j}^s \qquad i \in I \tag{3.20}$$

$$\beta_i^{s+1} = \min_{j \in Js} q_{i,j}^s \qquad Js = \left\{ j : q_{i,j}^s \neq 0 \right\}, \quad i \in I \tag{3.21}$$

$$\tau_i = \sqrt{\alpha_i^{s+1} \cdot \beta_i^{s+1}} \qquad i \in I\ . \tag{3.22}$$

Step 3:

Update rs_i^s to

$$rs_i^{s+1} = \tau_i \cdot rs_i^s \qquad i \in I \tag{3.23}$$

Step 2 and step 3 are the procedure for calculating row scaling coefficient *rs*$_i$. Apply the same procedure to columns. Calculate the maximum value γ_j^s and minimum non-zero value δ_j^s in each column *j* as equation (3.24) and (3.25). Calculate the square root of maximum and minimum value; σ_j (3.26).

Step 4:

$$\gamma_j^{s+1} = \max_i \frac{q_{i,j}^s}{\tau_i} \qquad j \in J \tag{3.24}$$

$$\delta_j^{s+1} = \min_{i \in I} \frac{q_{i,j}^s}{\tau_i} \qquad Is = \left\{ i : q_{i,j}^s \neq 0 \right\}, \quad j \in J \tag{3.25}$$

$$\sigma_j = \sqrt{\gamma_i^{s+1} \cdot \delta_i^{s+1}} \qquad j \in J\ . \tag{3.26}$$

Step 5:

Update cr_j^s to

$$cr_j^{s+1} = \sigma_j \cdot cr_j^s \qquad j \in J \tag{3.27}$$

Step 6:

Update $q_{i,j}^s$ to

$$q_{i,j}^{s+1} = \frac{q_{i,j}^s}{\tau_i \cdot \sigma_j} \qquad i \in I,\ j \in J \tag{3.28}$$

and iterate Steps 2 – 6, until satisfying:

$$\frac{\lambda^{s+1}}{\lambda^s} \geq 1 - \omega \tag{3.29}$$

Where λ^s is defined by following:

$$\lambda^s = \frac{\min_{i,j \in I,J} q_{i,j}^s}{\max_{i,j \in I,J} q_{i,j}^s} \quad (3.30)$$

We assume ω equals to 10^{-15}.

Consequently, we apply the rs_i^s and cs_j^s as the scaling coefficients for *i*-th row and *j*-th column of matrix **Q**, respectively. After the estimation of matrix **P** is ready, the same coefficients are used for "rescaling" the estimated matrix **P**.

4. Practical application

In this Section we present the practical application of the modified (as described in Section 3) cross-entropy method to estimation of input-output tables. We applied this method to estimate input-output tables of Japan for years 1985, 1980, 1975 and 1970.

4.1 Data set

4.1.1 Data

OECD publishes input-output tables of some countries (OECD, 1995)[23]. This data covers the input-output tables from 1970 to 1990 but the coverage year depends on the countries. In this article we applied Japanese input-output table. OECD also published latest version of input-output tables (OECD, 2002) [24]. However, the sector classification of the latest version is different from that of old one. Thus, we did not use the latest one.

In this application, we applied Japanese input-output table. In the OECD input-output table, there are four input-output of Japan and the coverage year is 1990, 1985, 1980, 1975 and 1970. This input-output table uses OECD sectoral classification. We re-arrange the sectoral classification described in Section 4.1.2 for the application.

4.1.2 Row and column classification

As we showed in Table 1, input-output table has "Production units" and "Primary inputs" in rows, and "Production units", "Final demand" in columns. Table A1, Table A2 and Table A3 (in the Appendix) show this research classification of production units, primary inputs and final demand. We aggregate OECD sectoral classification. Primary inputs and final demands are also classified as in Table A5 and Table A6. We will show how the OECD sectoral classification, primary inputs and final consumptions are mapped to each our commodity category in the Appendix.

4.1.3 Applied data

Table 5 summarizes the application data. We used 1990's aggregated input-output table as the prior matrix Q. Row and column sums of input-output table for each year(1985, 1980, 1975 and 1970) are applied to a_i and b_j. We set d^k for total of all sectors output and commodity trade (54; export and 55; import and the codes of commodities are 1 to 25) in each year. $v^l_{o,h}$ is set as $v^l_{o,1} = -v^l_{o,3}$ and $v^l_{o,2} = 0$, and $v^l_{o,1}$ as 5% of each constraint(a_i, b_j, d^k). And we set all $u^l_{o,h}$ as $1/3$.

Table 5 Application data

Parameters	Application data	year
$q_{i,j}$	input-output table 1990	1990
a_i	sum of rows of input-output table (other than sector from 1 to 31)	each year
b_j	sum of columns of input-output table	each year
d^{k1}	import value of sectors from 1 to 25	each year
d^{k2}	export value of sectors from 1 to 25	each year
d^{k3}	total output of sectors of 1 to 31	each year
$v^l_{o,1}\left(=-v^l_{o,3}\right)$	5% of each constraint(a_i, b_j, d^k)	each year
$u^l_{o,h}$	1/3	each year

4.2 Results

4.2.1 Assessment of methodological improvement

We performed four types of comparative calculations denoted by A, B, C, D:

A. using original matrix ($q_{i,j}$)

B. using scaled prior matrix (without removing non-substantial values)

C. removing non-substantial values but not scaling the matrix

D. removing non-substantial values and scaling the matrix

To assess the effectiveness of the above defined combinations of methods (defined by letters A, B, C, and D) we use, for evaluating their accuracy, the following four commonly used indicators; in their definition $q^y_{i,j}$ denotes reported values in year y and $p_{i,j}$ denotes estimated value. $q^y_{i,j}$ is calculated by equation (2.4).

Theil's U (Theil, 1971)

$$U = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{i,j} - q^y_{i,j}\right)^2}{\sum_{i=1}^{m}\sum_{j=1}^{n} {q^y_{i,j}}^2}} \qquad (4.1)$$

Standardized weighted absolute (Lahr, 2001)

$$SWAD = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}^{y} \left| p_{i,j} - q_{i,j}^{y} \right|}{\sum_{i=1}^{m}\sum_{j=1}^{n} {q_{i,j}^{y}}^{2}} \qquad (4.2)$$

Fit C (Roy *et al.*, 2001)

$$C = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}^{y} \log q_{i,j}^{y} - \sum_{i=1}^{m}\sum_{j=1}^{n} p_{i,j} \log p_{i,j}}{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}^{y} \log q_{i,j}^{y}} \qquad (4.3)$$

Standardized total percentage error

$$STPE = 100 * \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \left| q_{i,j}^{y} - p_{i,j} \right|}{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}^{y}} \qquad (4.4)$$

We show the information about prior matrix Q for each calculation in Table 6 (minimum value (except zero), maximum value, variance, mean and the solution time). Matrix size is 32 by 36 (it includes zero elements). The ratio between minimum and maximize values of the method B and D are smaller than method A and C. It is due to the scaling effect. And comparing that of method B and D, method D is smaller than method B.

Table 6 Information about the prior matrix **Q**

	Calculation			
	A	B	C	D
Minimum value	4.80E-07	2.84E-03	8.18E-05	4.12E-02
Maximum value	7.76E-01	3.52E+02	7.78E-01	2.43E+01
Ratio of Max and Min	1.62E+06	1.24E+05	9.51E+03	5.88E+02
Number of non-zero elements	1072	1072	862	862
Variance	0.0108	814.9	0.0154	23.20
Mean	0.0336	7.191	0.0418	1.892

Table 7 shows whether the problem was solved or not, four indicators and solution time of each method for 1985, 1980, 1975 and 1970. In Table 7, the blank means the solver could not find the solution.

Three observations from the results presented in Table 7 justify the strengths of the developed method.

Firstly, we can see that the calculation method C and A could not solve the problem for some years. It means that the scaling method, which is used by methods B and D, helps solving the problem.

Secondly, let us compare the four indicators. Comparing the values of time-series, we can see the difference between the estimated values and expected values become worse in most indicators, as the year is apart from the basic year. When it comes to comparing among the calculation methods, the values of indicators are almost same. The biggest different can be seen in 1970, and that the value of method B is higher than that of method C and D.

Thirdly, let us compare the solution time. Comparing method A and B, C and D, the solution time of the method A and C are better than that of method B and D. it indicates that scaling method increases the solution time. However, the method A and C has unsolved problems. Thus instead of increasing the solution time, the method B and D find the solutions.

Comparing method A and C, B and D, the solution time of the method C and D are better than that of method A and B. It indicates that removing non-substantial values improves the solution time.

Consequently, combination of the two methods; removing non-substantial values and scaling is highly effective. It results in not only shortening the solving time but also provides the correct solution.

Table 7 Four indicators and solution time for each method

	1985			
	A	B	C	D
Theil's U	0.131235	0.129590	0.131244	0.131244
SWAD	0.259253	0.253573	0.259026	0.259026
Fit C	0.012748	0.011795	0.009024	0.009024
STPE	15.25759	14.90616	15.32633	15.32633
Solution time(seconds)	408	859	112	123
	1980			
	A	B	C	D
Theil's U	0.191514	0.191514		0.191535
SWAD	0.334198	0.334198		0.333614
Fit C	0.013909	0.013909		0.010011
STPE	23.37556	23.37556		23.43713
Solution time(seconds)	380	495		227
	1975			
	A	B	C	D
Theil's U	0.204268	0.204268		0.204240
SWAD	0.348587	0.348587		0.348000
Fit C	0.030531	0.030531		0.026483
STPE	24.86755	24.86756		24.93811
Solution time(seconds)	483	1364		156
	1970			
	A	B	C	D
Theil's U		0.226565	0.201707	0.201707
SWAD		0.390249	0.364518	0.364518
Fit C		0.032668	0.024182	0.024182
STPE		29.48064	26.62262	26.62262
Solution time(seconds)		801	282	203

4.2.2 Assessment of the results by using ratio

In Section 4.2.1, we assess the differences among the methods. In this Section, we show how the estimation fits with the reported values. Figure 1, 2, 3 and 4 are the histogram of absolute ratio of the estimated value $p_{i,j}$ to the reported value $q^y_{i,j}$ for each year. Table 8 shows the information about Figure 1, 2, 3 and 4. Figures 1, 2, 3 and 4 describe the method D results of 1985 1980, 1975 and 1970. X-axis is the range of ratio and the range is shown in Table 8. The bars are the number of the values in each category and the line means cumulative percentage. We assume that the ratio is 1 if the estimated values of $p_{i,j}$ and reported values are $q^y_{i,j}$ equal to zero. If the bars are located on the center of the graphs, the estimation corresponds to the expected values. The year older, the correspondence become worse in Figure 1, 2, 3 and 4. For instance, the frequency in the range from 0.5 to 2 (from (6) to (15)) is 87.8%, 78.0%, 76.4% and 72.9% in 1985 1980, 1975 and 1970. And the summation of the frequency of less than 0.1 and more than 10 is 5.1%, 7.5%, 9.4% and 10.2% in 1985 1980, 1975 and 1970.

Table 8 Frequency of absolute ratios between $p_{i,j}$ and $q^y_{i,j}$ for each year y

				1985		1980		1975		1970	
	Range			Frequency	%	Frequency	%	Frequency	%	Frequency	%
(1)	0.0	-	0.1	26	2.3%	34	3.0%	44	3.8%	66	5.7%
(2)	0.1	-	0.2	3	0.3%	14	1.2%	9	0.8%	10	0.9%
(3)	0.2	-	0.3	4	0.3%	11	1.0%	23	2.0%	21	1.8%
(4)	0.3	-	0.4	10	0.9%	25	2.2%	17	1.5%	29	2.5%
(5)	0.4	-	0.5	21	1.8%	32	2.8%	32	2.8%	35	3.0%
(6)	0.5	-	0.6	29	2.5%	42	3.6%	47	4.1%	47	4.1%
(7)	0.6	-	0.7	36	3.1%	47	4.1%	40	3.5%	55	4.8%
(8)	0.7	-	0.8	60	5.2%	58	5.0%	50	4.3%	48	4.2%
(9)	0.8	-	0.9	82	7.1%	62	5.4%	63	5.5%	56	4.9%
(10)	0.9	-	1.0	132	11.5%	91	7.9%	96	8.3%	83	7.2%
(11)	1.0	-	1.1	402	34.9%	339	29.4%	329	28.6%	315	27.3%
(12)	1.1	-	1.3	97	8.4%	75	6.5%	63	5.5%	66	5.7%
(13)	1.3	-	1.4	76	6.6%	64	5.6%	55	4.8%	55	4.8%
(14)	1.4	-	1.7	64	5.6%	74	6.4%	71	6.2%	59	5.1%
(15)	1.7	-	2.0	34	3.0%	46	4.0%	66	5.7%	56	4.9%
(16)	2.0	-	2.5	22	1.9%	43	3.7%	35	3.0%	42	3.6%
(17)	2.5	-	3.3	12	1.0%	21	1.8%	17	1.5%	25	2.2%
(18)	3.3	-	5.0	6	0.5%	12	1.0%	19	1.6%	15	1.3%
(19)	5.0	-	10.0	3	0.3%	10	0.9%	12	1.0%	17	1.5%
(20)	10.0	-		33	2.9%	52	4.5%	64	5.6%	52	4.5%

In this section we use the ratio of the estimated value to the reported value in order to assess the estimation. The ratio deals with all $p_{i,j}$ and $q^y_{i,j}$ as same way. It means that even if $p_{i,j}$ or $q^y_{i,j}$ are very small, the errors can be significant. Thus, in the next section we are going to assess the difference of the estimated value $p_{i,j}$ and the reported value $q^y_{i,j}$.

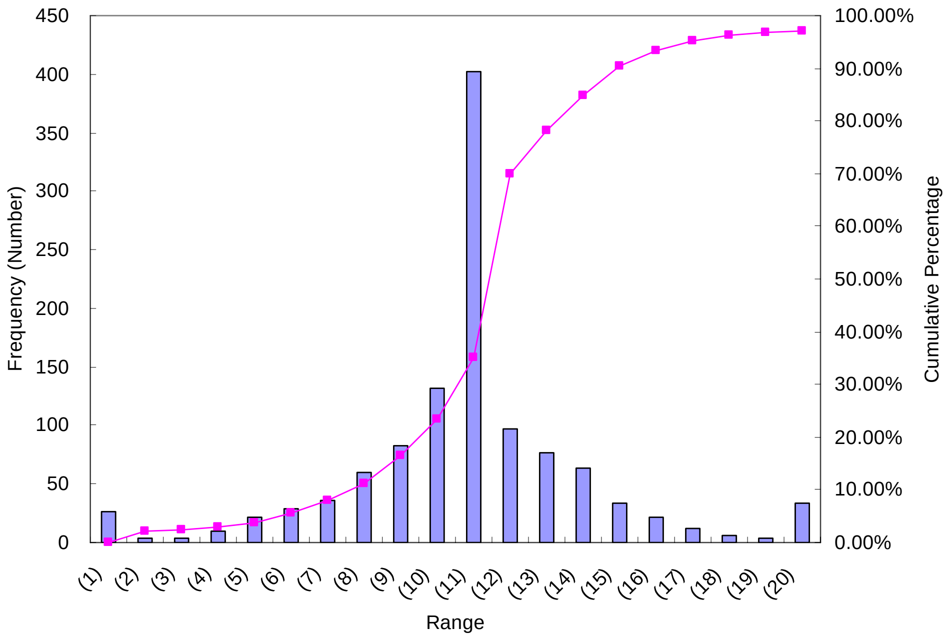

Figure 1 Histogram of absolute ratio between $p_{i,j}$ and $q_{i,j}^{y}$ (1985)

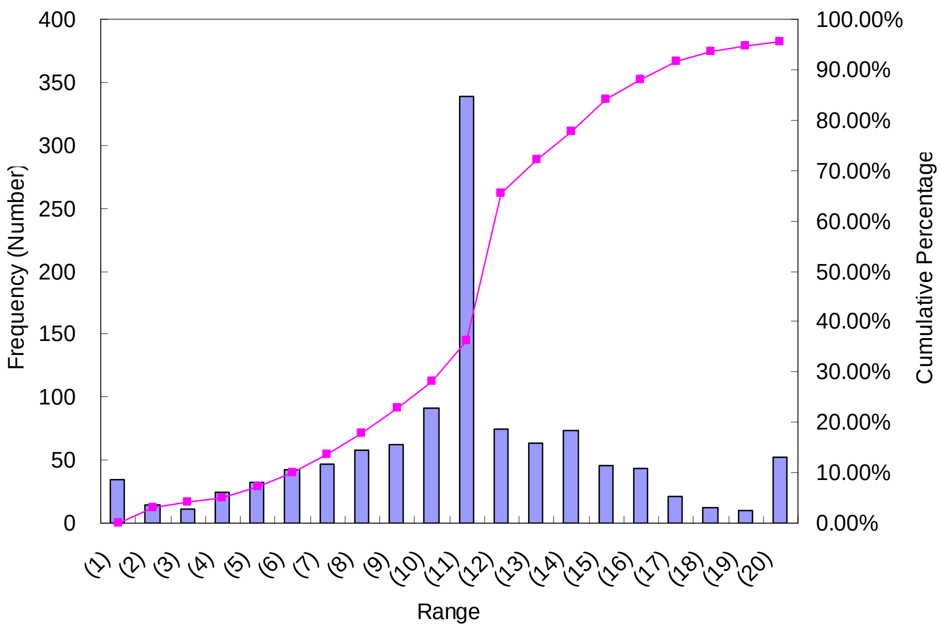

Figure 2 Histogram of absolute ratio between $p_{i,j}$ and $q_{i,j}^{y}$ (1980)

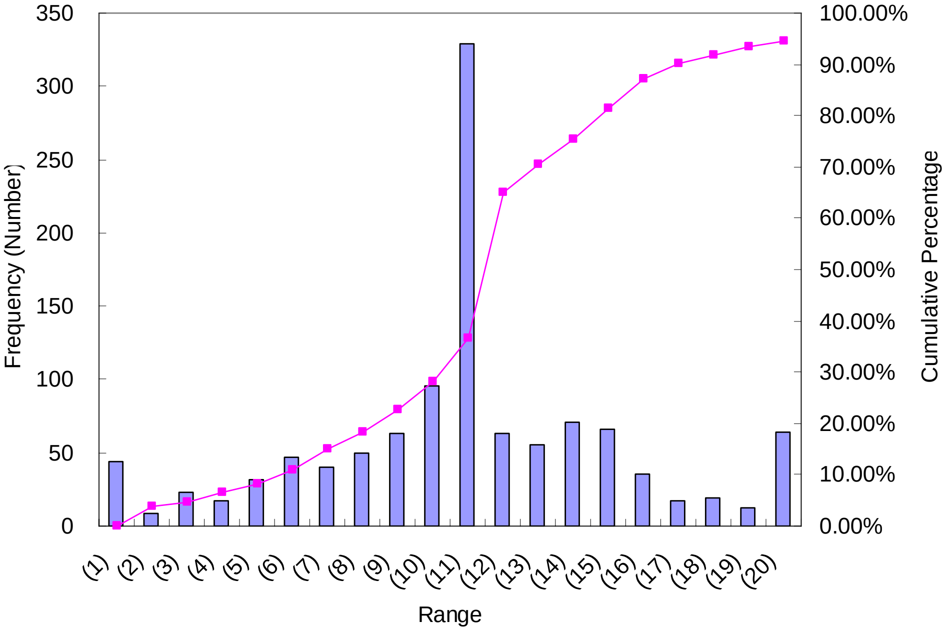

Figure 3 Histogram of absolute ratio between $p_{i,j}$ and $q^{y}_{i,j}$ (1975)

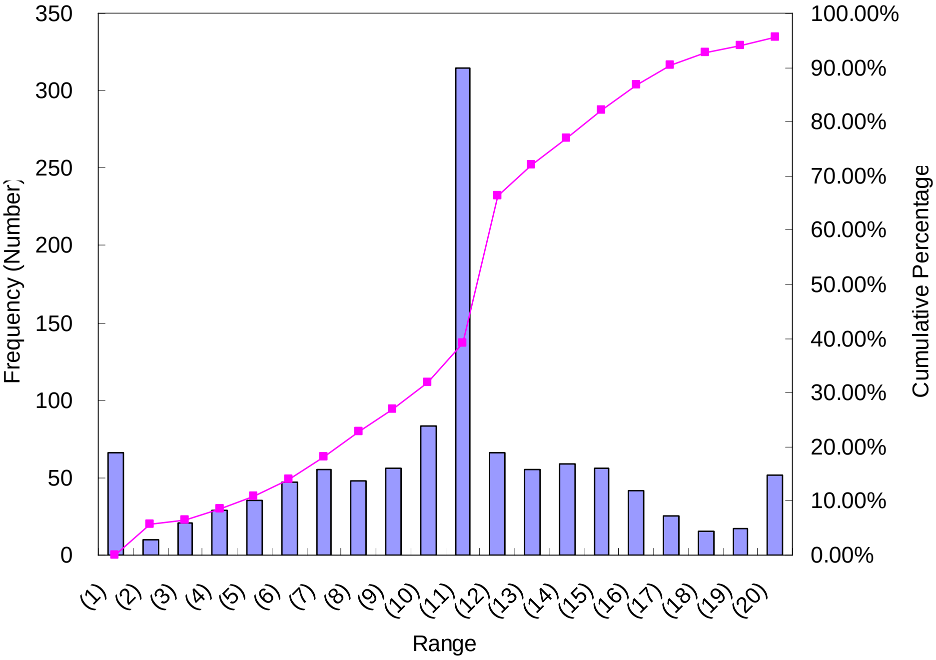

Figure 4 Histogram of absolute ratio between $p_{i,j}$ and $q^{y}_{i,j}$ (1970)

4.2.3 Assessment of the results by using difference

In this section we assess the results by using the difference of the estimated value $p_{i,j}$ and the reported value $q^{y}_{i,j}$. Table 9 shows the frequency of the difference of each year in each ranges. Range 0 – $10^{-7.0}$ means both of the estimated value $p_{i,j}$ and the reported value $q^y_{i,j}$ are zero. The frequency of the differences in the range from $10^{-2.5}$ to $10^{-0.9}$ ((11) to (14)) are dominant in every year. Moreover, The frequency of the differences less than $10^{-0.8}$ = 0.158(range from (1) to (15)) is 99.8%, 99.6%, 99.3% and 99.2% in 1985 1980, 1975 and 1970. Comparing among the years, 1985 and others shows different tendency. In 1985 the frequency in the range (10) and (11) are higher than that of others; 83 and 202, and the frequency in the range (13) and (15) are lower.

This assessment indicates two things. Firstly, the estimation of each year is not so different in 1980, 1975 and 1970. Secondly, the estimation of 1980, 1975 and 1970 are worse than that of 1985 but the difference of estimation and expected values are not so significant.

Table 9 Frequency absolute difference between $p_{i,j}$ and $q^y_{i,j}$

	Range			1985		1980		1975		1970	
				Frequency	%	Frequency	%	Frequency	%	Frequency	%
(1)	0	-	$10^{-7.0}$	265	23.0%	257	22.3%	249	21.6%	235	20.4%
(2)	$10^{-7.0}$	-	$10^{-6.5}$	0	0.0%	0	0.0%	0	0.0%	0	0.0%
(3)	$10^{-6.5}$	-	$10^{-6.0}$	0	0.0%	0	0.0%	0	0.0%	1	0.1%
(4)	$10^{-6.0}$	-	$10^{-5.5}$	0	0.0%	0	0.0%	0	0.0%	0	0.0%
(5)	$10^{-5.5}$	-	$10^{-5.0}$	2	0.2%	0	0.0%	0	0.0%	0	0.0%
(6)	$10^{-5.0}$	-	$10^{-4.5}$	1	0.1%	2	0.2%	0	0.0%	0	0.0%
(7)	$10^{-4.5}$	-	$10^{-4.0}$	9	0.8%	0	0.0%	1	0.1%	1	0.1%
(8)	$10^{-4.0}$	-	$10^{-3.5}$	9	0.8%	7	0.6%	4	0.3%	7	0.6%
(9)	$10^{-3.5}$	-	$10^{-3.0}$	32	2.8%	22	1.9%	25	2.2%	18	1.6%
(10)	$10^{-3.0}$	-	$10^{-2.5}$	83	7.2%	57	4.9%	57	4.9%	50	4.3%
(11)	$10^{-2.5}$	-	$10^{-2.0}$	202	17.5%	137	11.9%	132	11.5%	141	12.2%
(12)	$10^{-2.0}$	-	$10^{-1.5}$	213	18.5%	235	20.4%	237	20.6%	236	20.5%
(13)	$10^{-1.5}$	-	$10^{-1.0}$	171	14.8%	223	19.4%	219	19.0%	220	19.1%
(14)	$10^{-1.0}$	-	$10^{-0.9}$	131	11.4%	140	12.2%	156	13.5%	160	13.9%
(15)	$10^{-0.9}$	-	$10^{-0.8}$	32	2.8%	67	5.8%	64	5.6%	74	6.4%
(16)	$10^{-0.8}$	-	$10^{-0.7}$	0	0.0%	1	0.1%	6	0.5%	7	0.6%
(17)	$10^{-0.7}$	-	$10^{-0.6}$	1	0.1%	3	0.3%	0	0.0%	1	0.1%
(18)	$10^{-0.6}$	-	$10^{-0.5}$	0	0.0%	0	0.0%	0	0.0%	1	0.1%
(19)	$10^{-0.5}$	-	$10^{-0.4}$	1	0.1%	1	0.1%	2	0.2%	0	0.0%
(20)	$10^{-0.4}$	-	$10^{-0.3}$	0	0.0%	0	0.0%	0	0.0%	0	0.0%

4.2.4 What makes the errors

In the former section, we could see the accuracy of the results is good. In this section, we discuss what makes the errors. As mentioned above, the solution for 1970 is worse than that for 1985. In order to describe the reason, we show the histogram of absolute ratio of the estimated value $p_{i,j}$ for 1970 and the prior matrix $q_{i,j}$ (for 1990) in Figure 5. The histogram indicates the estimated values are quite close to the prior matrix. It can be seemed clearly comparing the

Figure 4 and Figure 5. It means the estimation strongly depends on the prior matrix. The cross-entropy function is multiplied the logarithm of the ratio $\ln \frac{p_{i,j}}{q_{i,j}}$ by weighted by $p_{i,j}$. Therefore, the lager $p_{i,j}$ are relatively estimated the value close to $q_{i,j}$. In other words, if $q_{i,j}$ is relatively small value, the accuracy of estimated value $p_{i,j}$ would be worse.

But the real input-output tables Q changes with the time especially for longer periods. Therefore there can be errors such as described in the former sections.

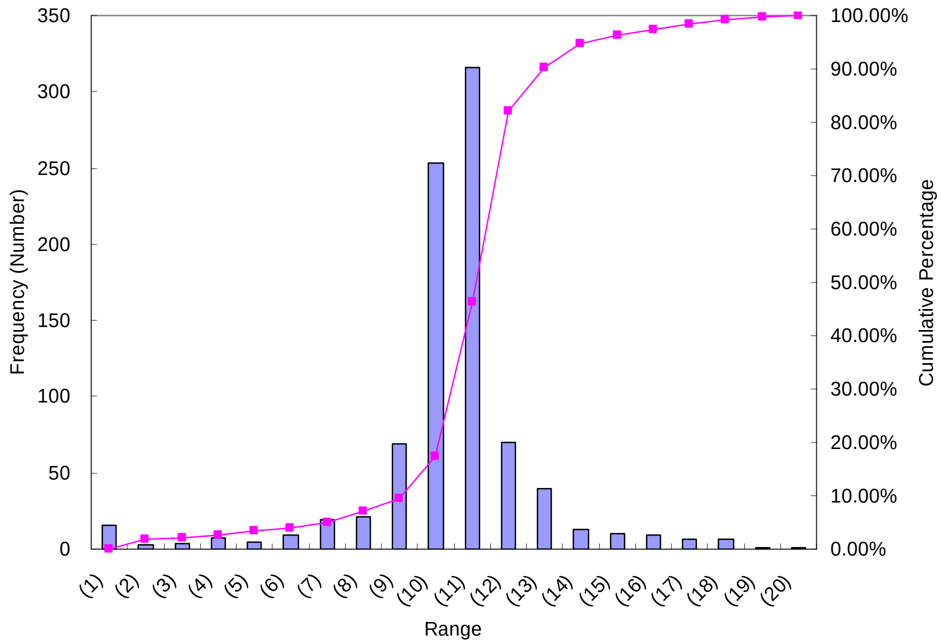

Figure 5 Histogram of absolute ratio of estimated $p_{i,j}$ of 1970 to the prior matrix $q_{i,j}$ of 1990

5. Concluding remarks and further steps

This paper presents the new and effective method for estimation of input-output tables developed to meet the requirements of estimation of material flows. Although several methods for estimation of input-output tables have been suggested, none of them could be effectively used for our problem. In particular we described the methodological problems related to using the cross-entropy method to estimation input-output tables having properties that cause numerical problems. Effective solutions to these problems were developed and implemented. The two main novel elements of the new approach are:

(1) methods for identification of non-substantive elements that should be removed from the prior matrix
(2) scaling of the non-linear optimization problem

We implemented this method in Japan for 1985, 1980, 1975 and 1970 using the 1990 input-output table for a prior matrix, and other aggregated information including sums of row and column for each year. Due to the described approach we could get the improvement in the results.

- Without the method which we proposed, the solver could not find the solution in some years.
- The estimation accuracy is guaranteed even if the estimation is more than 20 years. Most of the absolute differences between the estimated values and reported values are less than $10^{-0.8}$ (99.8%, 99.6%, 99.3% and 99.2% in 1985 1980, 1975 and 1970).
- The solving time is substantially improved.

Though we could get better solutions, there are still outliers in the solutions. In the practical application, we used only trade and total output for other additional information. However, other information, for instance energy statistics and value added, can make the accuracy of the estimation better. Therefore, we plan to use such kind of additional information. Finally we are planning to apply the method to estimate the global input-output tables.

References

[1] Ayres, R. (1978): Resources, Environment and Economics, New York.
[2] Ayres, R.U. and Kneese, A.V. (1968): Environmental Pollution, Federal Programs for the Development of Human Resources, Washington DC.
[3] Steurer, A. (1992): Stoffstrombilanz Oesterreich, 1988, Schriftenreihe Soziale Oekologie, No. Band 26. IFF/Abteilung Soziale Oekologie, Wien.
[4] Environment Agency Japan (1992): Quality of the environment in Japan 1992, Tokyo.
[5] Matthews, E., Bringezu, S., Fischer-Kowalski, M., Huetller, W., Kleijn, R., Moriguchi, Y., Ottke, C., Rodenburg, E., Rogich, D., Schandl, H., Schuetz, H., van der Voet, E., and Weisz, H. (2000): The weight of nations, Material outflows from industrial economies World Resources Institute, Washington.
[6] EUROSTAT (2001): Economy-wide material flow accounts and derived indicators, A methodological guide Office for Official Publications of the European Union, Luxembourg.
[7] United Nations (1993): Handbook of National Accounting: Integrated Environmental and Economic Accounting, Studies in Methods United Nations, New York.
[8] United Nations (2001): System of Environmental and Economic Accounting, SEEA 2000 Revision United Nations, New York.
[9] Stahmer, C., Kuhn, M., and Braun, N. (1997): Physische Input-Ouput-Tabellen 1990 Statistisches Bundesamt, Wiesbaden.
[10] Fujimori, S., and Matsuoka, Y. (2007): Development of estimating method of global carbon, nitrogen, and phosphorus flows caused by human activity, *Ecological Economics*, 62(3-4), 399-418.
[11] Hertel, T. W. (2005): Global Trade, Assistance, and Production The GTAP 6 Data Base, Purdue University.
[12] Hertel, T. W., (2001): Global Trade, Assistance, and Production The GTAP 5 Data Base, Purdue University.
[13] Bacharach, M. (1970): Biproportional matrices and input-output change, Cambridge, Cambridge University.
[14] Robinson, S., Cattaneo, A. and El-Said, M. (2001): Updating and estimating a social accounting matrix using cross entropy methods, *Economic Systems Research*, 13(1), 47-64.
[15] Golan, A., Judge, G. and Robinson, S. (1994): Recovering Information from Incomplete or Partial Multisectoral Economic Data, *Review of Economics and Statistics*, 76, 541-549.
[16] Stone, R., Champernowne, D. G. and Meade, J. E. (1942): The precision of national income estimates, *The Review of Economics and studies*, 9(2), 111-125.
[17] Byron, R. P. (1978): The estimation of large social account matrices, *Journal of the Royal statistical Society. Series A (General)*, 141(3), 359-367.
[18] Kullback, J.(1959): Information theory and statistics, John Wily and Sons, New York.

[19] International Energy Agency (2004): Energy Balance OECD, International Energy Agency.
[20] United Nations Industrial Development Organization (2006): UNIDO INDSTAT4 2006 Industrial Statistics Database, United Nations Industrial Development Organization.
[21] GAMS Development Corporation (2005): GAMS solver manuals, Washington, DC.
[22] Makowski M. and J. Sosnowski (1981): Implementation of an Algorithm for Scaling Matrices and Other Programs Useful in Linear Programming, Working Paper WP-81-37, International Institute for Applied Systems Analysis, Laxenburg, Austria.
[23] OECD (1995): The OECD Input-output Database, OECD, Paris.
[24] OECD (2002): OECD Input-Output tables edition 2002, OECD, Paris.
[25] Theil, H. (1971): Applied economic forecasting, North-Holland, Amsterdam.
[26] Lahr M. L. (2001): A strategy for producing hybrid regional input-output tables, input-output analysis: Frontiers and extensions, Palgrave, New York, 211-242.
[27] Roy, J. Batten, D. and Lesse, P. (1982): Minimizing information loss in simple aggregation, *Environment and Planning A*, 14, 973-980.

Appendix

In this Appendix, we show the sectoral and the OECD industry sectoral classifications; primary inputs and final consumptions are mapped to each our commodity category. This research classification is in Table A1, Table A2 and Table A3. The industry sector mapping is shown in Table A4, primary inputs mapping is in Table A5 and final demand is in Table A6.

Table A1 Production units classification

Code	Production units
1	Agriculture, forestry & fishing
2	Mining & quarrying
3	Food, beverages & tobacco
4	Textiles, apparel & leather
5	Wood products & furniture
6	Paper, paper products & printing
7	Industrial chemicals
8	Drugs & medicines
9	Petroleum & coal products
10	Rubber & plastic products
11	Non-metallic mineral products
12	Iron & steel
13	Non-ferrous metals
14	Metal products
15	Non-electrical machinery
16	Office & computing machinery
17	Electrical apparatus, nec
18	Radio, TV & communication equipment
19	Other transport equipment
20	Motor vehicles
21	Professional goods
22	Other manufacturing
23	Electricity, gas & water
24	Construction
25	Wholesale & retail trade
26	Restaurants & hotels
27	Transport & storage
28	Communication
29	Finance & insurance
30	Real estate & business services
31	Other services

Table A2 Primary inputs classification

Code	Primary inputs
41	Value add

Table A3 Final demand classification

Code	Final demand
51	Houshold
52	Government
53	Capital formation and stock change
54	Export
55	Import

Table A4 Concordance between OECD and our industrial sector classification

Code	This research classification	OECD code	OECD classification
1	Agriculture, forestry & fishing	1	Agriculture, forestry & fishing
2	Mining & quarrying	2	Mining & quarrying
3	Food, beverages & tobacco	3	Food, beverages & tobacco
4	Textiles, apparel & leather	4	Textiles, apparel & leather
5	Wood products & furniture	5	Wood products & furniture
6	Paper, paper products & printing	6	Paper, paper products & printing
7	Industrial chemicals	7	Industrial chemicals
8	Drugs & medicines	8	Drugs & medicines
9	Petroleum & coal products	9	Petroleum & coal products
10	Rubber & plastic products	10	Rubber & plastic products
11	Non-metallic mineral products	11	Non-metallic mineral products
12	Iron & steel	12	Iron & steel
13	Non-ferrous metals	13	Non-ferrous metals
14	Metal products	14	Metal products
15	Non-electrical machinery	15	Non-electrical machinery
16	Office & computing machinery	16	Office & computing machinery
17	Electrical apparatus, nec	17	Electrical apparatus, nec
18	Radio, TV & communication equi	18	Radio, TV & communication equipn
19	Other transport equipment	19	Shipbuilding & repairing
19	Other transport equipment	20	Other transport
20	Motor vehicles	21	Motor vehicles
19	Other transport equipment	22	Aircraft
21	Professional goods	23	Professional goods
22	Other manufacturing	24	Other manufacturing
23	Electricity, gas & water	25	Electricity, gas & water
24	Construction	26	Construction
25	Wholesale & retail trade	27	Wholesale & retail trade
26	Restaurants & hotels	28	Restaurants & hotels
27	Transport & storage	29	Transport & storage
28	Communication	30	Communication
29	Finance & insurance	31	Finance & insurance
30	Real estate & business services	32	Real estate & business services
31	Other services	33	Community, social & personal servi
31	Other services	34	Producers of government services
31	Other services	35	Other producers

Table A5 Concordance between OECD and our primary inputs classification

Primary inputs			
Code	This research classification	OECD code	OECD classification
41	Value added	38	Compensation of employees
	Operating surplus	39	Operating surplus
	Consumption of fixed capital	40	Consumption of fixed capital
	Indirect taxes	41	Indirect taxes
	(less) subsidies	42	(less) subsidies
	Value added	43	Value added

Table A6 Concordance between OECD and our final demand classification

Final Demand			
Code	This research classification	OECD code	OECD classification
51	Houshold	38	Private domestic consumption
52	Government	39	Government consumption
53	Capital formation and stock change	41	Changes in Stocks
54	Export	42	Exports of goods and services
55	Import	44	Imports of goods and services