Comparative Analysis of Mathematical Methods for Updating Input-Output Tables

Joachim Lamel
Josef Richter
Werner Teufelsbauer
Karl Zelle

Follow this and additional works at: https://researchrepository.wvu.edu/rri_iotheorymethods

Recommended Citation

This Article is brought to you for free and open access by the RRI Input-Output Archive at The Research Repository @ WVU. It has been accepted for inclusion in Theory and Methods by an authorized administrator of The Research Repository @ WVU. For more information, please contact beau.smith@mail.wvu.edu.
Session 3: Studies of national applications

COMPARATIVE ANALYSIS OF
MATHEMATICAL METHODS FOR
UPDATING INPUT-OUTPUT TABLES

by

Joachim Lamel, Josef Richter,
Werner Teufelsbauer and Karl Zelle
Federal Economic Chamber
and Computing Centre Graz
Vienna and Graz
Austria

Received on 20. 3. 1974
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>I  SCOPE AND METHOD OF THE INVESTIGATION</td>
<td>4</td>
</tr>
<tr>
<td>II STATISTICAL INDICATORS FOR THE PERFORMANCE OF THE UPDATING ALGORITHMS</td>
<td>7</td>
</tr>
<tr>
<td>III THE INFLUENCE OF PRECOS ON RAS PERFORMANCE</td>
<td>8</td>
</tr>
<tr>
<td>III.1 Main Results</td>
<td>8</td>
</tr>
<tr>
<td>III.2 Big PRECOs versus many PRECOs</td>
<td>13</td>
</tr>
<tr>
<td>IV THE INFLUENCE OF PRECOS ON ALMON PERFORMANCE</td>
<td>16</td>
</tr>
<tr>
<td>IV.1 Main Results</td>
<td>17</td>
</tr>
<tr>
<td>IV.2 ALMON versus RAS</td>
<td>18</td>
</tr>
<tr>
<td>V  THE DYMOD PROCEDURE FOR UPDATING IO TABLES</td>
<td>23</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>31</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION

In most countries empirical input-output (IO) tables can only be published several years after their year of reference because of lags in the basic statistics and the tremendous work necessary to put these statistics together into an IO table. This delay constitutes a serious problem for any practical application. In order to overcome this problem several algorithms have been developed to get IO tables for a more recent year by combining as much relevant information as possible (at least the margins of the IO table) for this recent year with the complete and consistent information content in the latest empirical IO table.

There has been a lot of theoretical discussions about the various methods of updating during the last years. But only few tests of efficiency with respect to the aim of minimizing errors have been carried out (6, 12). Due to theoretical considerations and empirical tests Stone's RAS method - based on the idea of bi-proportionality - and its various modifications have found the widest recognition among all updating methods under discussion (4, 5, 8, 9, 16, 17).

1) The project was financed by the "Fonds zur Förderung der Wissenschaftlichen Forschung", Vienna; computations were carried out by the "Rechenzentrum Graz", (UNIVAC 494)

2) Mrs. RITEULA typed the manuscript
Although some empirical tests (12, 13) gave evidence that the model of biproportionality works quite well in describing the changing structure of IO tables there are always some flows which do not follow this model. Paelinck and Waelbroek (12), Schintke, Stäglin and Wessels (14, 15) and others have already shown that the elimination of such flows improves the results of RAS considerably. But so far - to the knowledge of the authors - no systematic research has been done on the performance of the different updating methods with respect to the exogenous consideration of intermediate flows for the updating year. The lack of such investigations might partly be due to the fact that there are few countries with time series of fully comparable IO tables for which tests can be carried out.

This paper concentrates on the empirical test of only two major aspects of the introduction of intermediate flows:
- on the updating performance of RAS, when an increasing number of genuine information is considered which is only available at increasing costs,
- on the performance of a slightly modified version of Almon's algorithm (1, 2, 3) which allows to use information on the different reliability of additional information (considering increasing costs of additional information).

The paper also proposes a new kind of updating procedure (DYMOD = DYnamic MODel). This procedure starts with the estimation of a matrix containing both exogenous information and flows computed by means of a modification of the model of biproportionality. The second step uses a weighted least squares approach to satisfy the margins. In addition this procedure allows to introduce given "subset constraints" or knowledge on sub-matrices into the estimation process. Furthermore it is possible to modify, if necessary, a given estimation of the updated IO table by adding further plausible "subset constraints".

A variant of DYMOD is implemented in FORTRAN IV at the "Rechenzentrum GRAZ" (UNIVAC 494). For this method empirical tests have not been done yet.
The author's preoccupation with the selection and treatment of exogenous information reflects the pragmatic nature of the aim of their investigation. As soon as margins (intermediate row and column totals) are disposable for a target year a lot of exogenous information on intermediate flows will also be available and further information can be obtained at increasing costs. The aim of this paper is to present some useful advice on how to select such information and how to choose an appropriate algorithm within given cost limits.
I. SCOPE AND METHOD OF THE INVESTIGATION

The basic method used in all tests is to update an IO table for a base year to a target year and to compare the updated table with the respective genuine table of the target year. Margins and PRECOs\(^1\) are taken from the genuine table of the target year. Nine different ways of selecting PRECOs have been considered:

- selections by random process. These selections reflect the situation as it frequently is before systematic research is done to get additional information. As soon as margins are available there is always a lot of statistics which provide information on intermediate flows irrespective of their type, value or position. Four different random selections were considered. Flows \(x_{ij} \leq 30\) were not allowed to enter the selections. (Selection number 1, 2, 3, 4).

- selection according to the size of the flow, starting with the biggest flow, diagonal elements included. This selection and the following ones consider the fact that in many cases, statistics on large flows are published earlier and are of relatively high statistical quality. (selection number 5).

- selection according to the size of the flow, starting with the biggest flow, diagonal elements not included. (Selection number 6).

- selection according to the size, starting with the largest flow in each column and row. (Selection number 7).

- selection according to the largest indirect coefficients \((E - A)^{-1} - E - A\) of the base year. Large coefficients indicate a high degree of structural interdependence and the size of the propagation effect of an error in the respective flow. (Selection number 8).

- selection according to the largest indirect requirements \(((E - A)^{-1} - E - A)Y\) of the base year. Large indirect requirements give a picture of the degree of interdependence

---

1) According to Almon's terminology PRECO (PREConceived Opinion) stands for all exogenous information on intermediate flows of the target year.
of an industry with respect to given Final Demand. They are used as indicators for the size of the propagation effect of an error in a specific flow taking the base year structure as a proxy for the target year's structure. (Selection number 9).

According to the two algorithms used, two different types of PRECOs were introduced:
- exact information (taken out of the genuine table of the target year) on intermediate flows for RAS
- exact information (taken out of the genuine table for the target year) on intermediate flows biased by random error terms for Almon's algorithm.

The data was taken from a time series of fully comparable IO tables for Norway (Dimension of first quadrant:141 x 141). ¹)

Further investigations are intended to be done with data for other countries. The Norwegian tables were aggregated (51 x 51) in order to obtain a classification similar to the Austrian and to the EEC classification.

The IO tables used include both domestic and imported goods. Although a purely domestic version also was available there were good reasons to leave it out of consideration. Since the product-mix between imported and domestic goods is not technologically determined but rather a result of prices, capacity limits and fashion, the share of imports - with the exception of non-competitive imports - is not constant over time. Taking a domestic table for the updating job implies the necessity of estimating these unmeasurables.

Although there is also a time series of Norwegian IO tables in constant prices, only tables in current prices were used. First this choice in favour of current prices was influenced by the fact, that in most countries neither margins nor PRECOs are available at constant prices. Secondly, there is some empirical evidence that forecasts based on current price

¹) The authors are deeply indebted to Mr. Jacob Bjerve and Mr. Per Sevaldson, Central Statistical Bureau, Oslo, who furnished the data.
coefficients yield better results than forecasts based on constant price coefficients (20).

For the present study 1964 was chosen as base year and 1968 as target year.

In order to avoid excessive use of computer time all iterative procedures were stopped when the absolute deviations from the margins were smaller than 5.0 for all sectors. A series of preceding tests had shown that this degree of accuracy provides almost the same results as one of 0.5 (rounding) but cuts down the amount of computer time.
II. STATISTICAL INDICATORS FOR THE PERFORMANCE OF THE UPDATING ALGORITHMS

The application of appropriate indicators to the results of an updating procedure appeared to be one of the key problems of the investigation. Among the different indicators computed to compare the updated matrix with the genuine matrix (14, 18, 21) one very simple indicator and one which tries to find a common scale for different deviations of an estimate from the genuine value depending on the size of the flow were chosen for the presentation of the results:

\[ \text{ABSA}(F) = \sum |d_{ij}| \]  
where \( d_{ij} \) is the deviation of the estimate from the genuine value \( (x_{ij} - \hat{x}_{ij}) \)

\[ \text{ABSGM}(F) = \sum k_{ij}, k_{ij} = \frac{|d_{ij}|}{10x_{ij}^{0.1}} \]

Summation is done for all flows for which \( x_{ij} > 0 \) in the base year as well as in the target year.

The construction of \( \text{ABSGM}(F) \) implies that a deviation of e.g. ± 400 for a flow of 10,000 is as tolerable as a deviation of ± 63 for a flow of size 100 and as tolerable as a deviation of ± 160 for a flow of size 1,000. \( \text{ABSGM}(F) \) reflects the statisticians preference for small percentage errors in big flows and also the fact that small flows are of minor statistical quality even in genuine IO tables.
III. THE INFLUENCE OF PRECOS ON RAS PERFORMANCE

The results of RAS can easily be improved by the introduction of PRECOS. The value of such known elements is deducted from the margins and only remaining flows are treated according to the model of biproportionality. Usually the hypothesis of constant input coefficients is used to obtain starting values for the iteration. For the computations reported on a weighted average \(^1\) of constant intermediate input- and constant intermediate output-coefficients was used. Empirical tests indicated that the results of the RAS procedure are not significantly influenced by the starting values. The selected method, which corresponds to the MODOP procedure (16), ascertained quick convergence in most cases.

III.1 MAIN RESULTS

The presentation of main results of the nine different methods of selecting PRECOS concentrates on the two indicators ABSA(F) and ABSGM(F). Other indicators produce almost the same results. They are omitted here for sake of clarity.

As it may be seen from the two graphs (1 and 2) both for ABSA(F) and ABSGM(F) two clearly distinct bundles of error performance lines according to the different selection methods can be distinguished.

The upper bundle in each graph contains three of the four lines corresponding to random process selections (No. 1, 2, 3) which therefore can be said to be of lower quality than the others. The processes of selecting PRECOS systematically are superior to the random processes, especially methods

\[ x_{ij}^+ = \frac{g_i(1) h_j(1)}{g_i(0) h_j(0)} x_{ij}, \text{ where } g_i \text{ and } h_j \text{ are the row and column margins respectively} \]
Graph 1

RAS: Error performance of different PRECO selections

AWSA(P)

PRECO Selection:
1. Random
2. Random
3. Random
4. Random
5. Size of flows, biggest first
6. Size of flows, biggest first
7. Biggest flow of each row and col.
8. Indirect coefficients
9. Indirect requirements
Graph 2  
RAC: Error performance of different PRECO selections

PRECO Selection:
1. Random
2. Random
3. Random
4. Random
5. Size of flows, biggest first
6. Size of flows, biggest first
7. Biggest flow of each row and col.
8. Indirect coefficients
9. Indirect requirements
5 and 6 show a very good performance.

Table 1 shows the values of ABSA(F) and ABSGM(F) for the highest values of NQ (percentage rate of the number of PRECOs to the total number of flows) and VY (corresponding percentage rate of the value of PRECOs to the total value of flows) for the different methods of selecting PRECOs.

<table>
<thead>
<tr>
<th>PRECO selection</th>
<th>NQ</th>
<th>VY</th>
<th>ABSA(F)</th>
<th>ABSGM(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Random</td>
<td>18,00</td>
<td>68,92</td>
<td>6.118</td>
<td>124,2</td>
</tr>
<tr>
<td>2. Random</td>
<td>18,00</td>
<td>85,51</td>
<td>4.312</td>
<td>106,0</td>
</tr>
<tr>
<td>3. Random</td>
<td>17,98</td>
<td>81,70</td>
<td>7.821</td>
<td>118,4</td>
</tr>
<tr>
<td>4. Random</td>
<td>10,07</td>
<td>90,72</td>
<td>8.295</td>
<td>188,2</td>
</tr>
<tr>
<td>5. Size of flows, biggest first, incl. main diag.</td>
<td>2.27</td>
<td>69,86</td>
<td>25.807</td>
<td>362,6</td>
</tr>
<tr>
<td>6. Size of flows, biggest first, excl. main diag.</td>
<td>7.96</td>
<td>70,07</td>
<td>12.860</td>
<td>243,7</td>
</tr>
<tr>
<td>7. Biggest flow of each row and column</td>
<td>3.69</td>
<td>69,92</td>
<td>24.709</td>
<td>349,2</td>
</tr>
<tr>
<td>8. Indirect coefficients</td>
<td>9.61</td>
<td>64,55</td>
<td>17.591</td>
<td>261,9</td>
</tr>
<tr>
<td>9. Indirect requirements</td>
<td>4.23</td>
<td>69,90</td>
<td>22.607</td>
<td>319,7</td>
</tr>
</tbody>
</table>

More important than the results concerning the different error performance of the selected methods seems the fact that no decreasing returns to scale were found within the range of PRECOs considered. At any stage an increase of NQ means more or less the same reduction of error. Table 2 shows the correlation coefficients between NQ (VY) and ABSA(F) as proof for the mentioned linearity.
<table>
<thead>
<tr>
<th>PRECO selection</th>
<th>NQ</th>
<th>VY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Random</td>
<td>- 0.996</td>
<td>- 0.984</td>
</tr>
<tr>
<td>2. Random</td>
<td>- 0.997</td>
<td>- 0.987</td>
</tr>
<tr>
<td>3. Random</td>
<td>- 0.941</td>
<td>- 0.988</td>
</tr>
<tr>
<td>4. Random</td>
<td>- 0.882</td>
<td>- 0.990</td>
</tr>
<tr>
<td>5. Size of flows, biggest first, incl. main diag.</td>
<td>- 0.987</td>
<td>- 0.964</td>
</tr>
<tr>
<td>6. Size of flows, biggest first, excl. main diag.</td>
<td>- 0.958</td>
<td>- 0.963</td>
</tr>
<tr>
<td>7. Biggest flow of each row and column</td>
<td>- 0.991</td>
<td>- 0.981</td>
</tr>
<tr>
<td>8. Indirect coefficients</td>
<td>- 0.982</td>
<td>- 0.991</td>
</tr>
<tr>
<td>9. Indirect requirements</td>
<td>- 0.987</td>
<td>- 0.989</td>
</tr>
</tbody>
</table>
III.2 BIG PRECOS VERSUS MANY PRECOS

The first part of the investigation showed that the selection of PRECOs according to their size (starting with the biggest flows) provided better results than the selection of PRECOs by random process.

In practice the statistician who wants to update an IO table will rarely be free to choose PRECOs according to criteria such as size of flows, indirect coefficients, indirect requirements etc. His selection will often be determined by the availability of relevant statistical information, labor capacity and time. In the case of limited labor capacity the question arises whether he should concentrate on big flows or on many flows, if a selection of PRECOs according to one of the above mentioned criteria is hindered by the lack of statistical information.

If the statistician is free to choose, the key question is, whether 10% more information with respect to the number of PRECOs ascertains better results than 10% additional information with respect to the value of PRECOs. Of course this question cannot be answered isolated because any increase in the number of PRECOs is always associated with an increase in the total value of PRECOs and vice versa. In addition it is only meaningful to put this question if there are significant differences in the labor requirements per estimated flow. Otherwise the selection of PRECOs should start - as far as data is available - according to the criteria mentioned above.

However, it should be noted that in the case many PRECOs versus big PRECOs there is no "perfect" substitution between these two criteria, since any increase in the number is at least marginally associated with an increase in the value and vice versa. Graph 3 shows the extreme boundaries\(^1\) of the alternative and the

---

1) The upper boundary corresponds to a selection of PRECOs according to the size of flows starting with the biggest flow, the lower boundary to a selection starting with the smallest flow.
Theoretical case that a movement towards more PRECOs (a movement along the x axis) improves the results of updating as much as an identical movement towards a higher value of PRECOs (a movement along the y axis). It follows from the fact of "imperfect" substitution between many and big flows that the movement never will be parallel to the axes, but is determined by the slope of the extreme boundaries. In Graph 3 NQ again denotes the percentage rate of PRECOs to the total number of flows (zero flows included), VY the value of selected PRECOs in % of the total value of all intermediate transactions. The isoquants should be seen as the loci of all points of identical updating performances. C is a big versus many cost indifference curve illustrating the big-many combinations which can be obtained with a certain amount of labor input. As it may be seen easily from Graph 3 in this case the statistician should concentrate on more PRECOs.

Graph 3  Big PRECOs versus many PRECOs - a theoretical case
In order to obtain empirical estimates for the slopes of the isoquants the results reported above were used. Since these tests with the Norwegian IO tables were not done in such a detail that NQ/VY combinations with identical updating performance could be picked out, in many cases estimates have been obtained by linear interpolation. This, of course, only provides proxies but the procedure seems to be backed by the almost linear relationship between updating performance and NQ and VY respectively.

Fitting a curve of the specification \( VY = A \cdot NQ^{-\beta} \) to the observations for ABSGM(F) = 420, 400, 380, 360, 340 and 320 gave statistically meaningful estimates, which are shown in graph 4. The extreme boundaries are those of the Norwegian case. The big versus many cost indifference curve \( C \) was fitted intuitively according to the experience from updating an Austrian IO table for 1964 to 1970. Graph 4 shows-at least for the Norwegian case-that the question big versus many has to be answered in favor of many.

The specification \( VY = A \cdot NQ^{-\beta} \) also provided direct estimates of VY with respect to NQ. These elasticities vary from -1.57 to -2.08 and give an illustration of the priority which should be given to the selection of as many PRECOs as possible. The statistical quality of the figures for the elasticities is limited by the fact that the estimates are based on nine observations only. Nevertheless they seem to provide at least an impression about the Norwegian case.

<table>
<thead>
<tr>
<th>ABSGM(F)</th>
<th>Elasticity</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>420</td>
<td>-1.72</td>
<td>4.5</td>
</tr>
<tr>
<td>400</td>
<td>-2.08</td>
<td>4.4</td>
</tr>
<tr>
<td>380</td>
<td>-2.01</td>
<td>5.1</td>
</tr>
<tr>
<td>360</td>
<td>-1.78</td>
<td>5.1</td>
</tr>
<tr>
<td>340</td>
<td>-1.57</td>
<td>3.9</td>
</tr>
<tr>
<td>320</td>
<td>-1.78</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Graph 4: Big PRECOs versus many PRECOs - the Norwegian case

BIG PRECOs VERSUS MANY PRECOs - THE NORWEGIAN CASE

\[ p_1 \quad \text{updating performance} \quad \text{ABSGM}(F)=420 \]
\[ p_2 \quad \text{updating performance} \quad \text{ABSGM}(F)=400 \]
\[ p_3 \quad \text{updating performance} \quad \text{ABSGM}(F)=380 \]
\[ p_4 \quad \text{updating performance} \quad \text{ABSGM}(F)=360 \]
\[ p_5 \quad \text{updating performance} \quad \text{ABSGM}(F)=340 \]
\[ p_6 \quad \text{updating performance} \quad \text{ABSGM}(F)=320 \]

Big - many cost indifference curve
IV  THE INFLUENCE OF PRECOS ON ALMON PERFORMANCE

PRECOS remain unchanged by PAS. This implies that the exact value of the respective flows must be known, otherwise PAS would generate errors in the unknown flows even if they were strictly biproportional. But knowledge about the true values of target year flows is very scarce. More frequently, there is some knowledge about the magnitude of a flow, but the knowledge is only reliable up to a certain degree. Using RAS, this would mean that a lot of information (all PRECOS with less than absolute reliability) about the target year had to be left out of consideration in order not to bias the strictly biproportional flows.

ALMON (1, 2, 3) provides a possibility of considering explicitly the different statistical confidence in different PRECOS by allowing the algorithm to change not only unknown flows but also PRECOS according to different degrees of confidence attached to them. For tests with the Norwegian data each intermediate flow was assigned to one of four reliability classes.

Reliability class 0: highest reliability, no change at all allowed
1: less reliability, small changes allowed
2: less reliability, more change allowed
3: lowest reliability, all flows estimated with base year coefficients.

Random errors were attributed to all flows to get figures corresponding to reality. The tests were carried out for the the same PRECO-selections already used in testing PAS. Flows of reliability 0, 1, 2 not included in the respective PRECO-selection were moved to reliability class 3.

The following slightly modified version of ALMON was implemented:

1) In this chapter ALMON stands for "Almon's updating method"
\[ y^{(1)}_i - T_i = U_i r_i + V_i r_i + W_i r_i \]

\[ y^{(1)}_i : i\text{-th row sum for the target year} \]
\[ T_i : \text{sum of flows of reliability class 0} \]
\[ U_i : \text{sum of flows of reliability class 1} \]
\[ V_i : \text{sum of flows of reliability class 2} \]
\[ W_i : \text{sum of flows of reliability class 3} \]

\[ \alpha_1 = \frac{U_i}{(U_i + V_i + W_i)} \quad \quad \alpha_2 = \frac{U_i + V_i}{(U_i + V_i + W_i)} \]

\[ 0 < \alpha_1 \leq \alpha_2 \leq 1 \]

This means that PRECOs are changed according to the degree of reliability attached to them. The higher the reliability of PRECOs the smaller will be the changes because \( \alpha \) decreases with increasing reliability. The equations are solved with Newton's method by iteration. Then a similar adaption process is done for the columns, etc. After 20 steps PAS is used to balance the remaining minor deviations from the margins.

IV.1 MAIN RESULTS

Table 4 gives the values of \( ABSA(F) \) and \( ABSSGM(F) \) for the highest values of \( NQ \) and \( VY \) for the different methods of selecting PRECOs.

<table>
<thead>
<tr>
<th>PRECO selection</th>
<th>NQ</th>
<th>VY</th>
<th>( ABSA(F) )</th>
<th>( ABSSGM(F) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Random</td>
<td>10.07</td>
<td>90.72</td>
<td>19.396</td>
<td>252.1</td>
</tr>
<tr>
<td>5. Size of flows, biggest first, incl. main diag.</td>
<td>2.27</td>
<td>69.86</td>
<td>34.109</td>
<td>394.7</td>
</tr>
<tr>
<td>7. Biggest flow of each row and column</td>
<td>3.69</td>
<td>69.92</td>
<td>32.937</td>
<td>387.6</td>
</tr>
<tr>
<td>9. Indirect requirements</td>
<td>4.23</td>
<td>69.90</td>
<td>30.978</td>
<td>357.3</td>
</tr>
</tbody>
</table>
Graphs 5 and 6 give a comparison of the different error performance of RAS and ALMON for the same PRECO selections.

**Graphs 5 and 6**  Comparison of RAS and ALMON error performance of different PRECO selections

As Graphs 5 and 6 show, for each PRECO selection the ALMON error performance gives worse results than the corresponding RAS error performance because perfect knowledge of the magnitude of PRECOs was assumed for the latter case. In reality better knowledge is mainly a question of more research, e.g. more costs. Within a given cost limit more imperfect PRECOs NQA can be found than perfect PRECOs NQR. The question now is whether it is preferable to have less PRECOs of reliability class O or to
have more PRECOs with less reliability. The answer to this question depends on the overall influence of the degree of reliability of PRECOs on the error performance of the updating procedure and on the cost ratio between PRECOs of different reliability classes. If, in the ALMON case, we distinguish between at least three different reliability classes:

- perfectly known PRECOs
- less well known PRECOs
- unknown flows

Graphs 7 and 8 give an impression about the comparative preference of PAS and ALMON. Both the ALMON and the PAS performance curve in Graph 7 are derived from the tests with Norwegian data (selection: indirect requirements; indicator: ABSA(P)). The curves show the level of error depending on NQ.

For updating the 64 table to 68 with an error E₁ PAS needed NQ₁ (perfectly known) PRECOs. The same record was achieved by ALMON with NQ₁ PRECOs of less reliability where

\[ NQ₁^P < NQ₁^A \]  \[ (1) \]

Inspection of the two performance curves shows that the differences

\[ NQ₁^A - NQ₁^P \]

are increasing with increasing NQ₁^P. Graph 8 gives a cross-section plotting of equal-error NQ₁^A and NQ₁^P. The resulting curve shall be called PAS-ALMON PERFORMANCE INDIFFERENCE CURVE.

With respect to costs it seems reasonable to assume that a perfectly known PRECO is more expensive than the average ALMON PRECO. The specification of this cost relation depends on the particular situation in any country and period. For the example's sake Graph 7 and 8 assume that

- PAS PRECOs are more expensive than ALMON PRECOs and that
- with each additional PRECO the cost gap between PAS and ALMON is getting more narrow.

This assumption reflects the fact that at the beginning there is

---

1) \[ NQ₁^A \]: quantity of ALMON PRECOs necessary to yield the same error E₁ as with NQ₁ PAS PRECOs.
Graph 7  Error performance of cost indifferent RAS and ALMON PRECO selections

Selection: Indirect Requirements, ABSA(f)

Almon Performance Curve
RAS Performance Curve
RAS-Almon Cost Indifference Curve
Graph 8  Determination of RAS and ALMON preferability zones

Selection:
Indirect Requirements, ABSACF
almost always a lot of information with less reliability at almost no costs. As the stock of these PRECOs is running out, additional less reliable knowledge is getting exponentially more expensive. On the other hand, perfect knowledge of flows is already scarce and comparatively expensive from the beginning. These considerations lead to the RAS-ALMON COST INDIFFERENCE CURVE in Graphs 7 and 8 which give a cross section plotting of \( NQ_1^R \) and \( P_{NQ_1^A} \).\(^1\) Depending on any country's particular situation other curves must be introduced.

Using Graph 7 it can be demonstrated that for a cost level which allows \( NQ_1^R \) RAS PRECOs or \( P_{NQ_1^A} \) ALMON PRECOs respectively, it is preferable to use ALMON, since \( P_{NQ_1^A} \) ALMON PRECOs yield a better error performance \( P_{E_1} \) than \( NQ_1^R \) RAS PRECOs (error performance \( E_1 \)). The situation is quite different at a cost-level which allows \( NQ_2^R \) RAS PRECOs or \( P_{NQ_2^A} \) ALMON PRECOs. \( P_{NQ_2^A} \) ALMON PRECOs have an error performance of \( P_{E_2} \) which is worse than the error performance \( E_2 \) of the corresponding RAS selection. To get the same error performance with ALMON, \( NQ_2^A \) (\( NQ_2^A > P_{NQ_2^A} \)) would be necessary.

Graph 8 shows both the RAS-ALMON PERFORMANCE INDIFFERENCE CURVE and the RAS-ALMON COST INDIFFERENCE CURVE. At their intersection \( X \) RAS and ALMON are equally preferable. If \( P_{NQ_1^A} < P_{NQ_1^A} \) then RAS should be preferred to ALMON. If \( P_{NQ_1^A} > P_{NQ_1^A} \) ALMON should be preferred to RAS. Assuming a cost indifference Curve like in Graphs 7 and 8 ALMON seems to be preferable if \( NQ \) is small. RAS is beginning to pay only starting from a relatively high \( NQ \) level. Size and location of the RAS and ALMON preferability zone will of course change depending on a changed specification of the cost indifference curve.

---

\(^1\) \( P_{NQ_i^A} \): quantity of ALMON PRECOs which can be found with the same cost level which allowed to find \( NQ_i^R \) RAS PRECOs.
The DyMod Procedure for Updating Input Output Tables

In this chapter the well-known problem of estimating a non-negative matrix from given margins is reconsidered in some revised way, taking into account that besides the margins sometimes also the sums of particular sets of flows are known with a high degree of certainty (see Graph 9). The simplest case is the availability of the value of a single flow, then the prescribed margins are reduced by this value. A further alternative arises if the sum of elements of a submatrix equals a definite value. Sums of this kind often are available from unsufficiently disaggregated statistics. Further information which can be dealt with in the estimation procedure are either details on the quality of flows (2, 3) or details on the limits within which a flow may vary. The proposed non-iterative least squares approach thus tries to incorporate this kind of information into the estimation and allows interventions by the user during the estimation process.

Graph 9 Examples for subset constraints

m rows

\[ \begin{array}{c|c|c|c} 
U_1 & U_2 & U_3 \\
\end{array} \]

n columns
Definitions:

\[ \hat{X}_{mn} = (x_{ij}) \quad \text{m by n matrix with margins} \]
\[ g_1^m, h_1^n \]

\[ X_{mn}^+ = (x_{ij}^+) \quad \text{a preliminary estimation of} \]
\[ X_{mn}^1, \text{generally not consistent} \]
\[ \text{with } g_1^m, h_1^n \]

\[ T \quad \text{is the set of all pairs of indices} \]
\[ \text{of all nonnegative elements of} \]
\[ \text{matrix } X_{mn} \]

\[ U_k \subseteq T \quad \text{is a subset of } T \]

The elements of matrix \( X^+ \) are either independent estimations or are computed by a mathematical model e.g. the model of biproportionality (8, 19).

The different types of information mentioned above can now be written as follows:

\[ (R1) \quad \sum_j \hat{x}_{ij} = e_i \quad i = 1, \ldots, m \quad \text{and} \]
\[ \sum_i \hat{x}_{ij} = h_j \quad j = 1, \ldots, n-1 \]

\( \hat{x}_{ij} \)

represents the \( m+n-1 \) linear independent marginal constraints.

\[ (R2) \quad \sum_{(i,j) \in U_k} \hat{x}_{ij} = f_k \quad k = 1, \ldots, l \quad \text{are } l \text{ linear independent} \]
\[ \text{subset constraints. The different} \]
\[ \text{sets } U_k \text{ may overlap each other} \]
\[ \text{(see graph 9).} \]
The least squares approach with different forms of weighting factors subject to constraints (R1) has already been mentioned several times in some papers (4, 11, 18).

Also in this contribution a similar approach was chosen, this time, however, with regard to constraints of type (R1) and (R2).

We minimize the Lagrangian

$$
(4) \quad \mathcal{L} = \frac{1}{2} \sum_{(i,j) \in \Gamma} \left( \frac{\hat{x}_{ij} - x_{ij}^+}{d_{ij}} \right)^2 - \sum_i \kappa_i \left( \sum_{i,j} \hat{x}_{ij} - g_i \right) - \sum_j \lambda_j \left( \sum_i \hat{x}_{ij} - h_j \right) - \sum_k \mu_k \left( \sum_{(i,j) \in U_k} \hat{x}_{ij} - f_k \right) + \min \kappa_i, \lambda_j \text{ and } \mu_k \text{ are the well known Lagrange multipliers.}
$$

\( \sum' \) denotes only a summation over columns \( j \) until \( n' = n - 1 \).

\( d_{ij} \) represents a measure to be compared with the deviation \( \hat{x}_{ij} - x_{ij}^+ \) and is set \( d_{ij} = \hat{x}_{ij} - \bar{x}_{ij} \). In cases where no upper and lower bounds are given one can define a deviation \( d_{ij} = 10 \cdot x_{ij}^+ O^4 \) as analogon to the statistical indicator \( \text{ABSCSM}(F) \) mentioned above.
This kind of definition of the weighting factors in [4] incorporates in some sense restrictions of type (R3). Of course the matrix calculated from [4] will satisfy (R1) and (R2) but this does not apply for all constraints (R3). In particular there may be some \( \hat{x}_{ij} < 0 \).

The user can now gradually remove such inconsistencies by adding further constraints of type (R2) without being forced to repeat the whole procedure from the beginning. Thus it becomes evident that not a rigid model was built, the solution of which is more or less binding for the user as it is e.g. the case with iterative procedures. There exists, however, a real opportunity of controlling the calculating process in order to correct some weaknesses of the weighted least squares approach defined in [4]. The procedure is presented schematically by a "flowchart" before running into mathematical details:

- Set up first estimate \( \hat{x}^+ \);
- find first solution of the least squares problem \( \hat{x} \).
- Matrix rounding procedure with regard to constraints (R1).
- Are there any constraints (R3) violated?
- Yes: Add one constraint of type (R2) in order to remove some violated constraints (R3); solution \( \hat{x} \) is modified.
- No: Stop.
From (4) we have for all \((i,j) \in \Omega\)

\[
\frac{\partial \mathcal{L}}{\partial y_{ij}} = \frac{\hat{x}_{ij} - x_{ij}^+}{d_{ij}} - \kappa_i - \lambda_j - \sum_k u_{ik} \delta_{ij,k} = 0; \quad \lambda_n = 0
\]

where \(\delta_{ij,k} = \begin{cases} 0 & \text{for } (i,j) \in \Omega_k, \\ 1 & \text{otherwise} \end{cases}\)

that means \(\delta_{ij,k}\) is an indicator, whether the particular pair \((i,j)\) is an element of \(\Omega_k\) or not.

With the abbreviation \(e_{ij} = d_{ij}^2\) we write

\[
(6) \quad x_{ij} - x_{ij}^+ = (\kappa_i + \lambda_j) e_{ij} + \sum_k u_{ik} \delta_{ij,k} e_{ij}
\]

Insertion in (R1) and (R2) yields following system of linear equations:

\[
(7) \quad p_i \kappa_i + q_j e_{ij} \lambda_j + \sum_k u_{ik} u_k = r_i
\]

\[
(8) \quad \sum_i u_{ik} \kappa_i + \sum_j v_{jk} \lambda_j + \sum_k w_{ak} u_a = t_k
\]

The variables used in (7) are defined as follows:

\[
(8) \quad p_i = \sum_j e_{ij}; \quad q_j = \sum_i e_{ij};
\]

\[
u_{ik} = \sum_j \delta_{ij,k} e_{ij}; \quad v_{jk} = \sum_i \delta_{ij,k} e_{ij}; \quad \sum_k \delta_{ij,k} a e_{ij};
\]

\[
r_i = g_i - \sum_j x_{ij}^+; \quad s_j = h_j - \sum_i x_{ij}^+; \quad t_k = f_k - \sum_{(i,j) \in \Omega_k} x_{ij}^+;
\]
In matrix notation (7) becomes:

\[
\begin{pmatrix}
P_{m \times m} & E_{m \times n'} & U_{m \times i} \\
E_{n' \times m}^T & Q_{n' \times n'} & V_{n' \times i^2} \\
U_{n' \times m}^T & V_{n' \times m'} & W_{n' \times i^2}
\end{pmatrix}
\begin{pmatrix}
\kappa_m \\
\lambda_{n'} \\
\mu_i
\end{pmatrix}
= 
\begin{pmatrix}
r_m \\
s_{n'} \\
t_i
\end{pmatrix}
\]

\(n' = n - 1\)

\(P\) und \(Q\) are diagonal matrices. The dimension subscripts will be omitted in subsequent formulas. If there are no constraints of type (R2) during the first step (i.e. \(l = 0\)), system (9) is solved easily:

\[
(10)
\begin{pmatrix}
\kappa \\
r'
\end{pmatrix}
= 
\begin{pmatrix}
A & -C^T \\
-C & B
\end{pmatrix}
\begin{pmatrix}
r' \\
s
\end{pmatrix}
\]

where use is made of the abbreviations

\[
B = (Q - E^T P^{-1} E)^{-1} = B^T \\
C = B E^T P^{-1} \\
A = P^{-1} (I + EC) = A^T
\]

(10) states clearly, that we have to perform one matrix inversion for a \(n'\) by \(n'\) matrix. In order to write the following formulas more clearly we define some new matrices
Further

\[
(12) \quad F = \begin{pmatrix} U \\ V \end{pmatrix}, \quad K = (W - F^T Z F)^{-1} = K^T \\
L = K F^T Z \\
M = Z (I + FL) = M^T
\]

With these definitions the general solution of (9) can be written

\[
(13) \quad \begin{pmatrix} \psi \\ \mu \end{pmatrix} = \begin{pmatrix} M & -L^T \\ -L & K \end{pmatrix} \begin{pmatrix} \omega \\ t \end{pmatrix} \quad \text{and is easy to verify.}
\]

(Compare with (10)!). More clearly arranged

\[
(14) \quad \begin{pmatrix} \psi \\ \mu \end{pmatrix} = \begin{pmatrix} \psi_o + Z F \rho \\ -\rho \end{pmatrix} \quad \text{with}
\]

\[
\psi_o = Z \omega; \quad \rho = K (F^T \psi_o - t)
\]

This formula shows the influence of the constraints (R2). Without these constraints \( \psi = \psi_o = Z \omega \) which is equivalent to (10). As stated above this solution will possibly violate some constraints (R3). Now the user has the opportunity of adding one (or more) constraint of Type (R2) in order to modify the first solution slightly and so on.
In particular he has to define some set $U_{l+1}$ for which

$$[15] \sum_{U_{l+1}} \hat{x}_{ij} = \tau ;$$

this new constraint has to be linear independent of the old [R1] and [R2]. $U_{l+1}$ contains some of that flows, which have been altered too much or to the wrong side according to the user's opinion. is still an unknown parameter. In complete analogy to [11] till [14] the new system (enlarged by one equation) is solved, that implies the "inversion of a 1 by 1 matrix". The new Lagrange multiplicators are now linear functions of parameter $\tau$ and from [6] we can derive therefore

$$[16] \hat{x}_{\text{new}}(\tau) = \hat{x}_{\text{old}} + \tau \tilde{X}.$$

$\hat{x}_{\text{old}}$ means the preceding approximation of $X^1$ and $\tilde{X}$ a constant matrix. [16] is the equation of a straight line, all points on it are feasible solutions with regard to all preceding constraints [R1] and [R2]. $\tau$ should now be defined in a way, that $\hat{x}_{\text{new}}$ will contain improved flow estimations for $U_{l+1}$. 

SUMMARY

The aim of this paper is to present the results of empirical tests carried out to check the error performance of the RAS and Almon algorithms for updating input-output tables from a base year to a target year. Special consideration was given to the influence of introducing exogenous information about intermediate flows of the target year.

The tests were carried out on the basis of a time series of Norwegian input-output tables by updating the base year table (1964) of this time series to the target year (1968) and comparing the results with the genuine target year table. Margins and exogenous information were taken from the genuine target year table.

With respect to different possibilities of choosing exogenous information it was found that certain selections as e.g. the selection according to the size of the flow, selection according to the largest indirect requirements, etc., are preferable to other selections.

Considering a given cost limit for research on additional exogenous information, a special analysis of individual results was done to show that exogenous knowledge on many flows should be preferred to big flows or vice versa depending on the slope of a given cost indifference curve. In the usual case the selection of many exogenous flows yields better results than the selection of big flows.

Comparing RAS with Almon's method application preferability zones defined by the number of available exogenous flows were found, also depending on the slope of a cost indifference curve.

A new updating procedure DYMOD which allows to introduce information on submatrices and to consider subset constraints and which enables users to intervene during the updating procedure is proposed.
REFERENCES


7. H. Birg, "Probleme der Schätzung von Matrixelementen aus bekannten Randverteilungen (Dargestellt am Beispiel einer Wanderungsmatrix auf Länderebene für die Bundesrepublik Deutschland)", Deutsches Institut für Wirtschaftsforschung, Vierteljahreshefte zur Wirtschaftsforschung Nr. 1/2 1972, pp. 79 - 85


14. J. Schintke, R. Stäglin, H. Wessels, "Die Bedeutung der Vorgabe einzelner Feldwerte für die modellmäßige Erstellung von Input-Output-Tabellen", Deutsches Institut für Wirtschaftsforschung, Vierteljahreshefte zur Wirtschaftsforschung, Jg. 1969; Drittes Heft, pp. 298 - 312


