Theory and Methods

RRI Input-Output Archive

4-1-1974

Comparative Analysis of Mathematical Methods for Updating Input-Output Tables
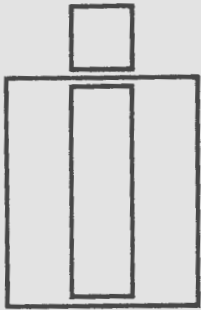
Joachim Lamel

Josef Richter

Werner Teufelsbauer

Karl Zelle

Recommended Citation

C58nn

SIXTH INTERNATIONAL CONFERENCE
ON input-output techniques

(Vienna, 22–26 April 1974)

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

CONTENTS

COMPARATIVE ANALYSIS OF MATHEMATICAL METHODS FOR UPDATING INPUT-OUTPUT TABLES[1]

by

Joachim Lamel, Josef Richter, Werner Teufelsbauer,
Karl Zelle [2]

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).

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 in-
troduce 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 "Rechen-
zentrum 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 how to select such information and how to chose 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 $x_{ij} \leqslant 30$ 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 techno-
logically 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 Norwegean 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 forcasts based on current price

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 preceeding 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:

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

$$ABSGM(F) = \sum \left| k_{ij} \right|, \quad k_{ij} = \frac{d_{ij}}{10 x_{ij}^{0,4}}$$

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 ABSGM(F) implies that a deviation of e.g. $\pm$ 400 for a flow of 10.000 is as tolerable as a deviation of $\pm$ 63 for a flow of size 100 and as tolerable as a deviation of $\pm$ 160 for a flow of size 1.000. 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.
Usally 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 clearity.

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

[1]

$$x^+_{ij} = \sqrt{\frac{g^{(1)}_i}{g^{(0)}_i} \frac{h^{(1)}_j}{h^{(0)}_j}} \; x^o_{ij}, \text{ where } g_i \text{ and } h_j \text{ are the row and column margins respectively}$$

Graph 1   RAS: Error performance of different PRECO selections

ABSA(F)

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 clm.
8. Indirect coefficients
9. Indirect requirements

ABSA(F) (1.000)   NQ →

Graph 2 — RAS: Error performance of different PRECO selections
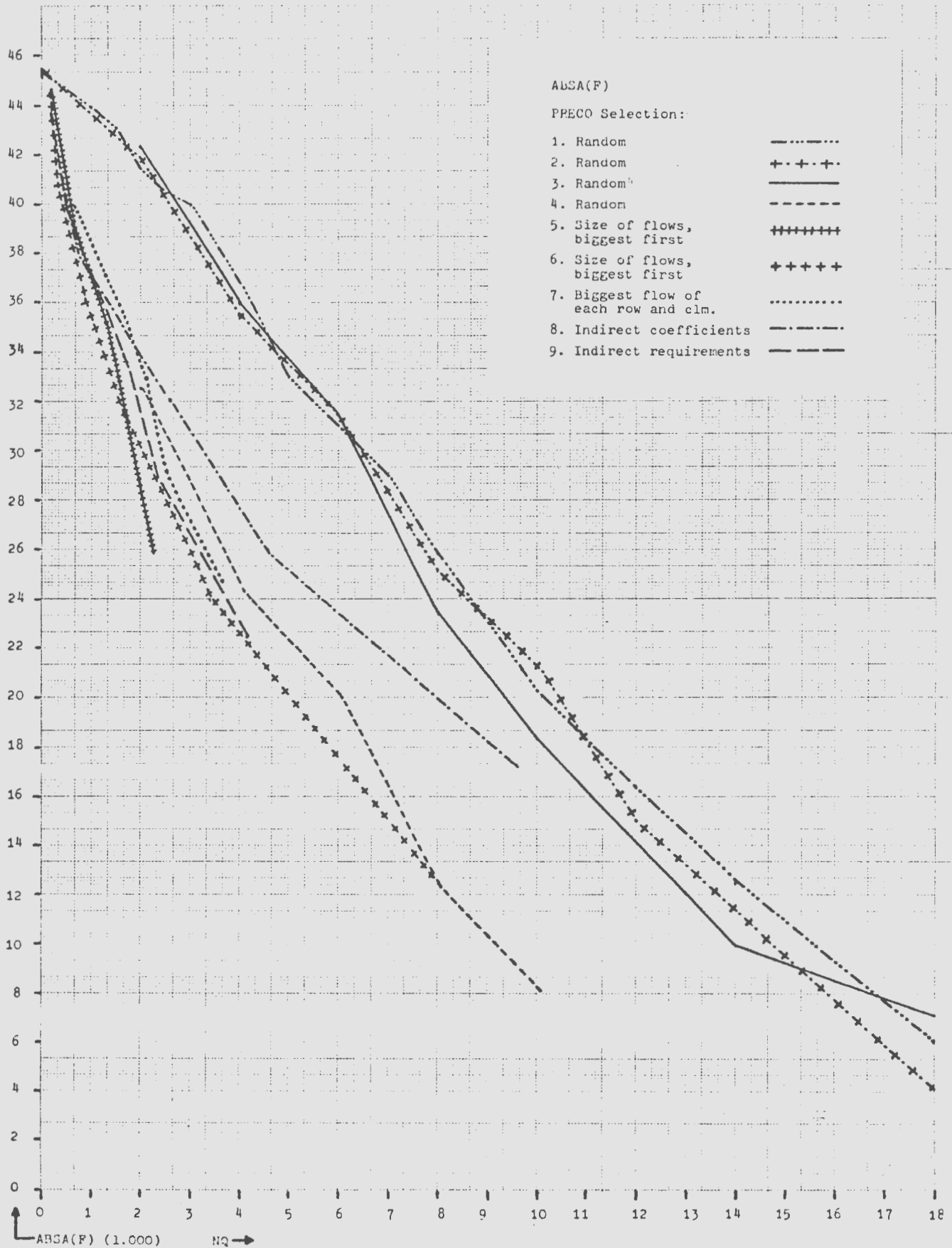
5 and 6 show a very good performance.

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 1   Best updating performance per selection

PRECO selection	NQ	VY	ABSA(F)	ABSGM(F)
1. Random	18,00	68,92	6.118	124,2
2. Random	18,00	85,51	4.312	106,0
3. Random	17,98	81,70	7.821	118,4
4. Random	10,07	90,72	8.295	188,2
5. Size of flows, biggest first, incl. main diag.	2,27	69,86	25.807	362,6
6. Size of flows, biggest first, excl. main diag.	7,96	70,07	12.860	243,7
7. Biggest flow of each row and column	3,69	69,92	24.709	349,2
8. Indirect coefficients	9,61	64,55	17.591	261,9
9. Indirect requirements	4,23	69,90	22.607	319,7

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 2   Correlation (r) between ABSA(F) and NQ, VY

PRECO selection	NQ	VY
1. Random	- 0,996	- 0,984
2. Random	- 0,997	- 0,987
3. Random	- 0,941	- 0,988
4. Random	- 0,882	- 0,990
5. Size of flows, biggest first, incl. main diag.	- 0,987	- 0,964
6. Size of flows, biggest first, excl. main diag.	- 0,958	- 0,963
7. Biggest flow of each row and column	- 0,991	- 0,981
8. Indirect coefficients	- 0,982	- 0,991
9. Indirect requirements	- 0,987	- 0,989

III.2 <u>BIG PRECOS VERSUS MANY PRECOS</u>

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 coeffients, 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 can
not 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 move-
ment 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 paralell 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

VY

C = cost indifference curve
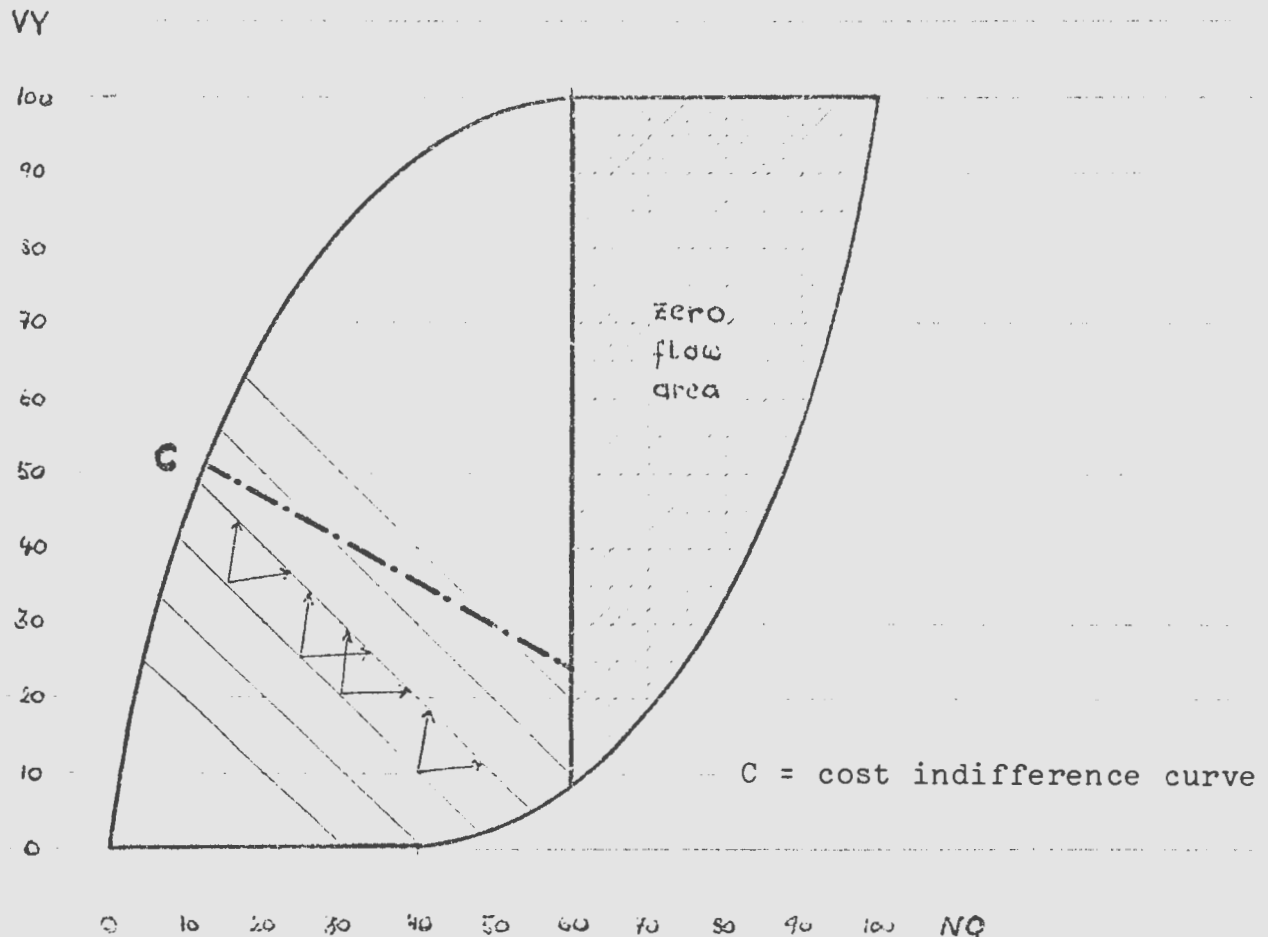
In order to obtain empirical estimates for the slopes of the
isoquants the results reported above were used. Since these
tests with the Norwegian IO tables were not done in such a
detail that NQ/VY combinations with identical updating per-
formance could be picked out, in many cases estimates have been
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 re-
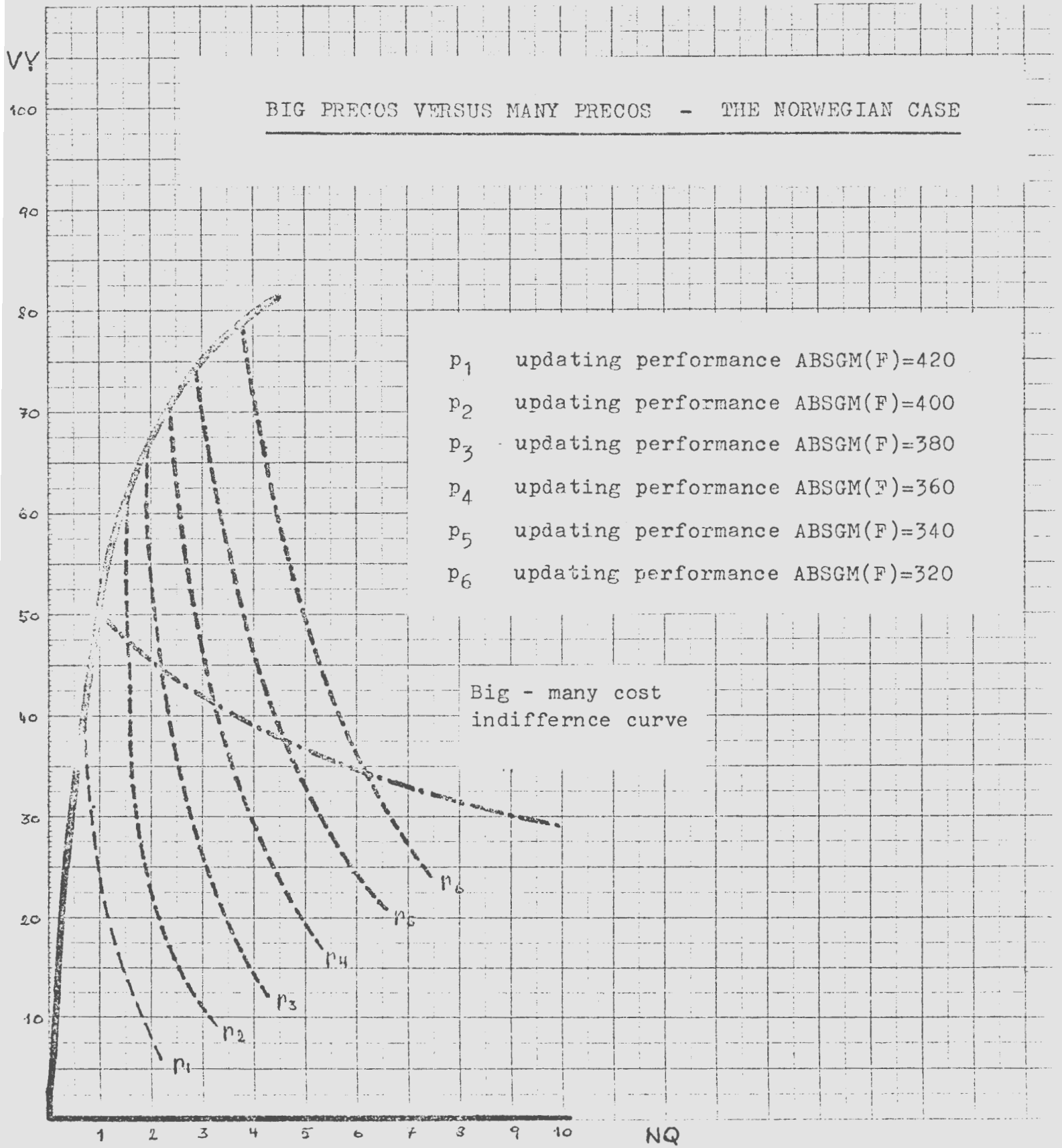spectively.

Fitting a curve of the specification $VY = A \cdot NQ^{-\beta}$ to the ob-
servations 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 3   Elasticities of VY with respect to NQ

ABSGM(F)	Elasticity	t-value
420	- 1.72	4.5
400	- 2.08	4.4
380	- 2.01	5.1
360	- 1.78	5.1
340	- 1.57	3.9
320	- 1.78	5.3

Graph 4   Big PRECOs versus many PRECOs - the Norwegian case

BIG PRECOS VERSUS MANY PRECOS — THE NORWEGIAN CASE

p_1 updating performance ABSGM(F)=420

p_2 updating performance ABSGM(F)=400

p_3 updating performance ABSGM(F)=380

p_4 updating performance ABSGM(F)=360

p_5 updating performance ABSGM(F)=340

p_6 updating performance ABSGM(F)=320

Big - many cost
indiffernce curve

IV THE INFLUENCE OF PRECOS ON ALMON PERFORMANCE [1]

PRECOS remain unchanged by RAS. This implies that the exact
value of the respective flows must be known, otherwise RAS
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 PPECOs with
less than absolut 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 expli-
citly 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 inter-
mediate 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 PPECO-
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"

$$y_i^{(1)} - T_i = U_i r_i + V_i r_i^{\alpha_1} + W_i r_i^{\alpha_2}$$

$y_i^{(1)}$: i-th row sum for the target year
T_i : sum of flows of reliability class 0
U_i : sum of flows of reliability class 1
V_i : sum of flows of reliability class 2
W_i : sum of flows of reliability class 3

$$\alpha_1 = \frac{U_i}{(U_i + V_i + W_i)} \qquad \alpha_2 = \frac{U_i + V_i}{(U_i + V_i + W_i)}$$

$$0 < \alpha_1 \leq \alpha_2 \leq 1$$

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 α 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 RAS is used to balance the remaining minor deviations from the margins.

IV.1 MAIN RESULTS

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

Table 4   Best ALMON updating performance per selection

PRECO selection	NQ	VY	ABSA(F)	ABSGM(F)
4. Random	10,07	90,72	19.396	252,1
5. Size of flows, biggest first, incl. main diag.	2,27	69,86	34.109	394,7
7. Biggest flow of each row and column	3,69	69,92	32.937	387,6
9. Indirect requirements	4,23	69,90	30.978	357,3

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
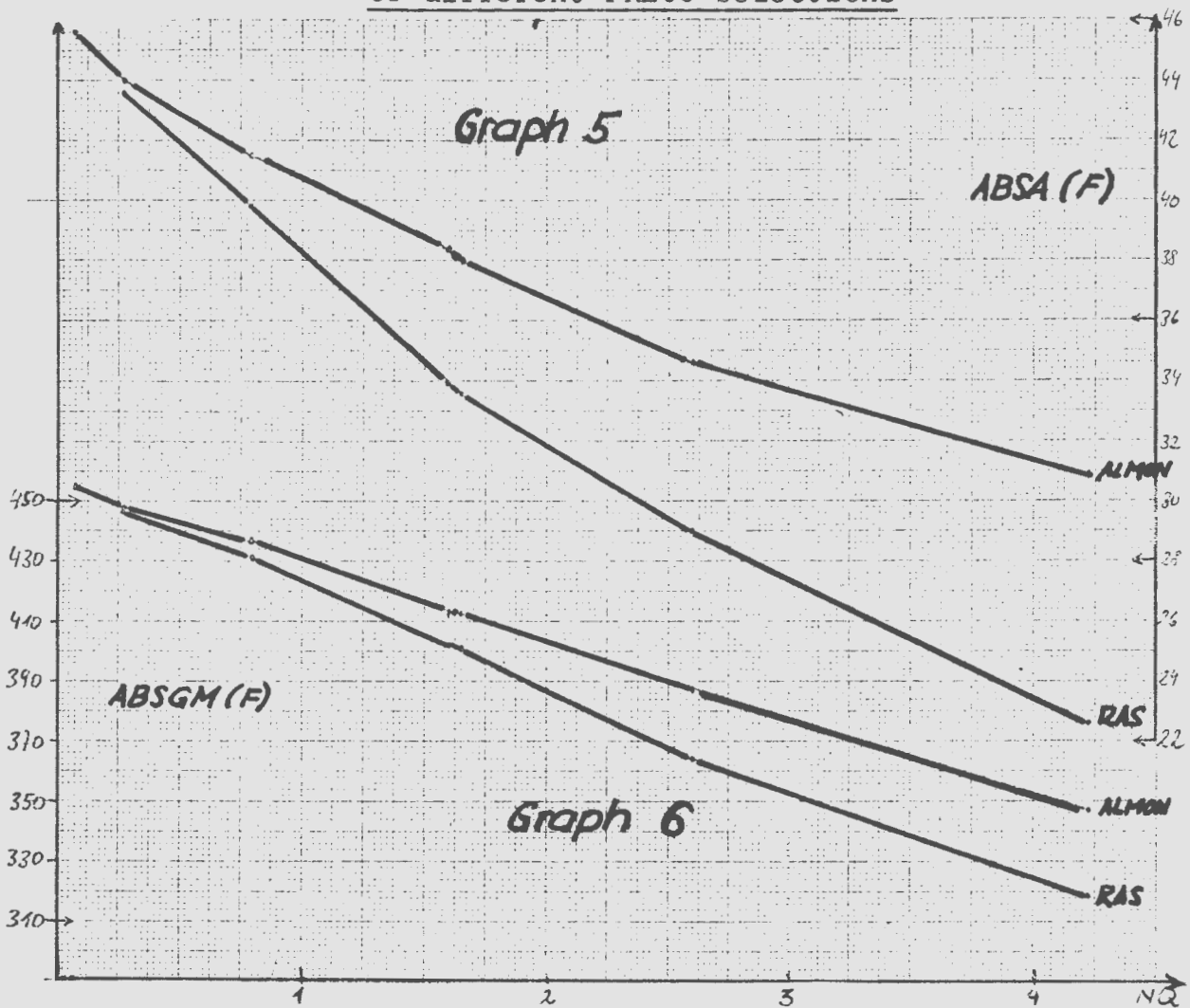
IV.2   ALMON VERSUS RAS

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 magni-
tude 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 NQ^A can be
found than perfect PRECOs NQ^R. The question now is whether it is
preferable to have less PRECOs of reliability class 0 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 beween 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 pre-
ferability of RAS and ALMON. Both the ALMON and the RAS perfor-
mance curve in Graph 7 are derived from the tests with Norwegian
data (selection:indirect requirements;indicator:ABSA(F)). The
curves show the level of error depending on NQ.

For updating the 64 table to 68 with an error E_1 RAS needed NQ_1^P
(perfectly known) PRECOs. The same record was achieved by ALMON
with $^I NQ_1^A$ PRECOs of less reliability where

$$NQ_1^P < {^I NQ_1^A} . \qquad 1)$$

Inspection of the two performance curves shows that the differences

$$^I NQ_i^A - NQ_i^P$$

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

With respect to costs it seems reasonable to assume that a
perfectly known PRECO is more expensive than the averag~ ``MON
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

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

This assumption reflects the fact that at the beginning there is

---

1) $^I NQ_i^A$: quantity of ALMON PRECOs necessary to yield the same
      Error E_i as with NQ_i^P RAS PRECOs.

Graph 7   Error performance of cost indifferent RAS and ALMON PRECO selections
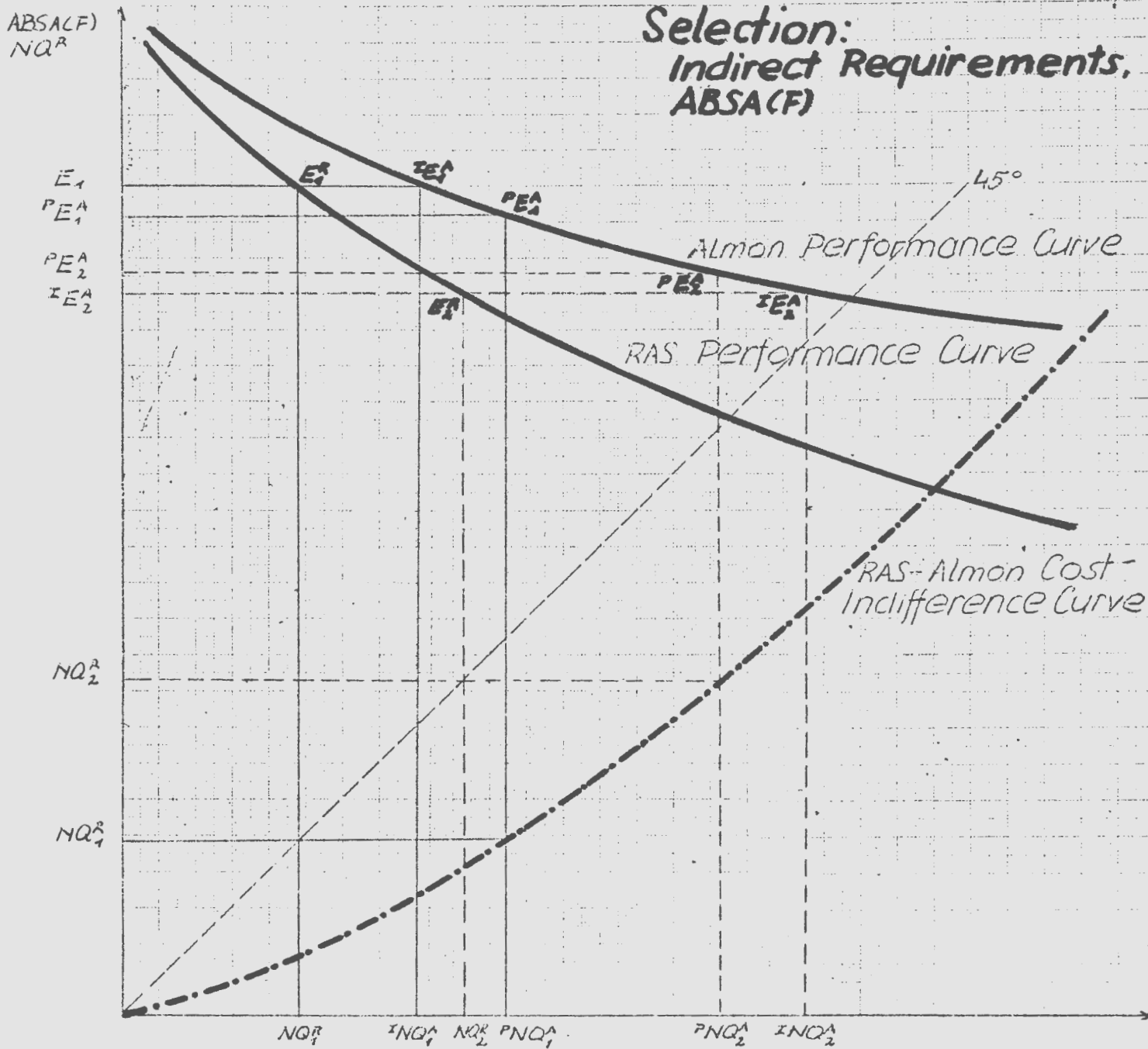
Graph 8   Determination of RAS and ALMON preferability zones

Selection:
Indirect Requirements, ABSA(F)

Almon Preferability Zone   RAS Preferability Zone

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
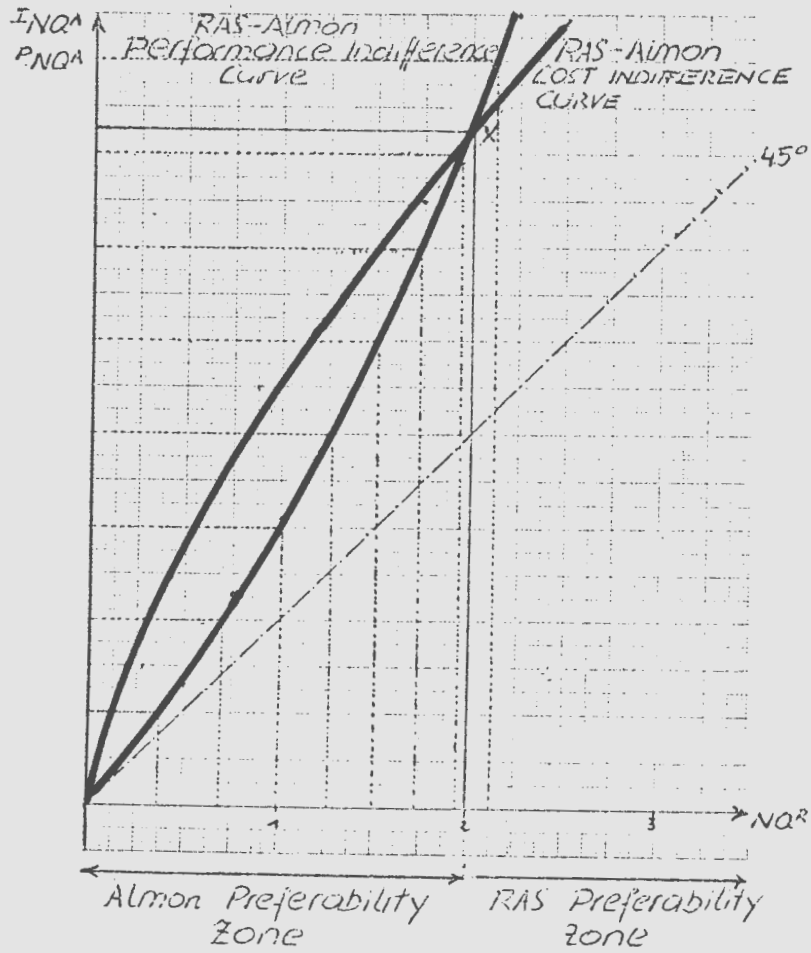more expensive. On the other hand, perfect knowledge of flows
is already scarce and comparatively expensive from the beginning.
These considerations lead to the RAS-ALMON COST INDIFFERENCE
CURVE in Graphs 7 and 8 which give a cross section plotting of
NQ_i^R and $^PNQ_i^A$.[1] Depending on any country's 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 $^PNQ_1^A$ ALMON PRECOs respectively, it
is preferable to use ALMON, since $^PNQ_1^A$ ALMON PRECOs yield a
better error performance $^PE_1$ 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 $^PNQ_2^A$ ALMON PRECOs.
$^PNQ_2^A$ ALMON PRECOs have an error performance of $^PE_2$ which is worse
than the error performance E_2 of the corresponding RAS selection.
To get the same error performance with ALMON, $^INQ_2^A$ ($^INQ_2^A > ^PNQ_2^A$)
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 $^PNQ_i^A < ^INQ_i^A$ then RAS
should be prefered to ALMON. If $^PNQ_i^A > ^INQ_i^A$ 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) $^PNQ_i^A$: quantity of ALMON PRECOs which can be found with the
same cost level which allowed to find NQ_i^R RAS PRECOs.

V 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 some-
times 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 pre-
scribed 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

n columns
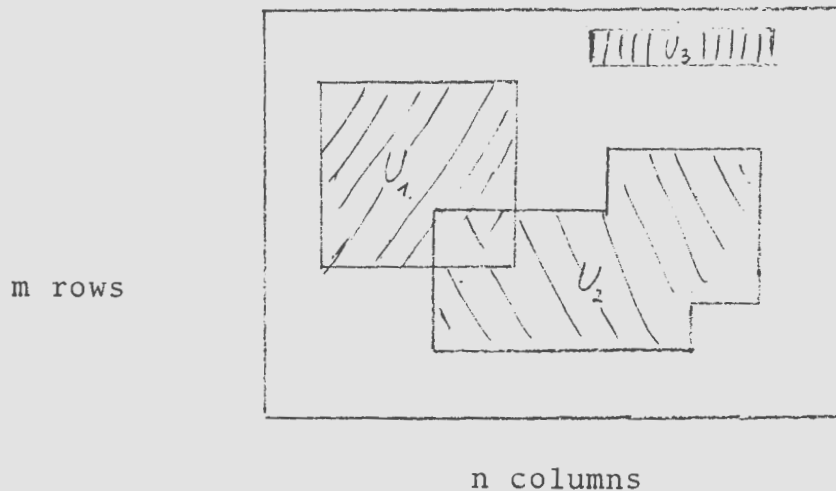
Definitions:

[1] $\hat{X}_{mn} = (x_{ij})$ m by n matrix with margins
                                      g_m^1, h_n^1

        $X_{mn}^+ = (x_{ij}^+)$ a preliminary estimation of
                                      X_{mn}^1, generally not consistent
                                      with g_m^1, h_n^1

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

        $U_k \subset T$ 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 bipro-
portionality (8, 19).

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

[3] (R1) $\sum_j \hat{x}_{ij} = g_i$ i = 1,...,m   and

           $\sum_i \hat{x}_{ij} = h_j$ j = 1,...,n-1

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

    (R2) $\sum_{(i,j)\epsilon U_k} \hat{x}_{ij} = f_k$ k = 1,...,ℓ are ℓ linear independent
                                          subset constraints. The different
                                          sets U_k may overlap each other
                                          (see graph 9).

(R3) $\underline{\eta}_{ij} \leq \overset{\grave{}}{x}_{ij} \leq \bar{\eta}_{ij}$ (in particular $0 \leq \hat{x}_{ij}$) for

some set of pairs of indices. Of course constraints

of type (R3) must not contradict (R1) or (R2).

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) <u>and</u> (R2).
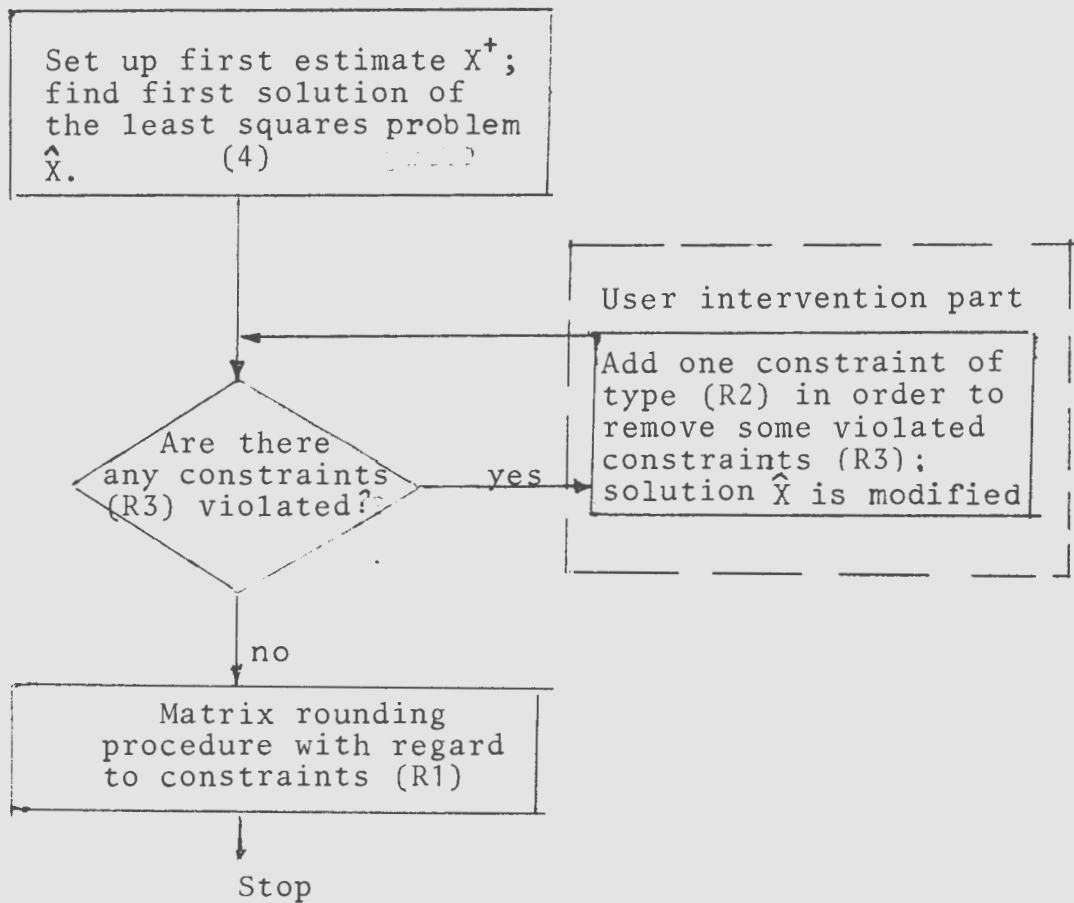
We minimize the Lagrangian

(4) $$\mathcal{L} = \frac{1}{2} \cdot \sum_{(i,j) \in T} \left(\frac{\hat{x}_{ij} - x^+_{ij}}{d_{ij}} \right)^2 - \sum_i \kappa_i \left(\sum \hat{x}_{ij} - g_i \right) -$$

$$- {\sum_j}' \lambda_j \left(\sum_i \hat{x}_{ij} - h_j \right) - \sum_k \mu_k \left(\sum_{(i,j) \in U_k} \hat{x}_{ij} - f_k \right) \rightarrow \min$$

κ_i, λ_j and μ_k are the well known Lagrange multiplicators.

${\sum_j}'$ denotes only a summation over columns 1 until $n' = n - 1$

d_{ij} represents a measure to be compared with the deviation $\hat{x}_{ij} - x^+_{ij}$ and is set $d_{ij} = \bar{\eta}_{ij} - \underline{\eta}_{ij}$. In cases where no upper and lower bounds are given one can define a deviation $d_{ij} = 10 \cdot x^{+^{0,4}}_{ij}$ as analogon to the statistical indicator ABSGM(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 beeing 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 X^+; find first solution of the least squares problem $\hat{X}$. (4)

User intervention part

Are there any constraints (R3) violated?

Add one constraint of type (R2) in order to remove some violated constraints (R3); solution $\hat{X}$ is modified

yes

no

Matrix rounding procedure with regard to constraints (R1)

Stop

From (4) we have for all $(i,j) \in \overline{T}$

(5) $\quad \dfrac{\partial \mathcal{L}}{\partial y_{ij}} = \dfrac{\hat{x}_{ij} - \bar{x}^{+}_{ij}}{d_{ij}^{2}} - \kappa_i - \lambda_j - \sum\limits_{k} \mu_k \, \delta_{ij,k} = 0; \quad \lambda_n = 0$

where $\delta_{ij,k} = \begin{cases} 0 & \text{for } (i,j) \notin U_k \\ 1 & \in \end{cases}$, that means $\delta_{ij,k}$ is an

indicator, whether the particular pair (i,j) is an element of U_k or not. With the abbreviation $\quad e_{ij} = d_{ij}^{2}$ we write

(6) $\quad x_{ij} - x^{+}_{ij} = (\kappa_i + \lambda_j) \, e_{ij} + \sum\limits_{k} \mu_k \, \delta_{ij,k} \, e_{ij}$

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

(7) $\quad p_i \kappa_i + \sum\limits_{j}{}' e_{ij} \lambda_j + \sum\limits_{k} u_{ik} \mu_k = r_i$

$\quad \sum\limits_{i} e_{ij} \kappa_i + q_j \lambda_j \qquad\qquad + \sum\limits_{k} v_{jk} \mu_k = s_j$

$\quad \sum\limits_{i} u_{ik} \kappa_i + \sum\limits_{j}{}' v_{jk} \lambda_j + \sum\limits_{a} w_{ak} \mu_a = t_k$

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

(8) $p_i = \sum\limits_{j} e_{ij}; \qquad q_j = \sum\limits_{i} e_{ij};$

$\quad u_{ik} = \sum\limits_{j} \delta_{ij,k} \, e_{ij}; \quad v_{jk} = \sum\limits_{i} \delta_{ij,k} \, e_{ij}; \quad w_{ak} = \sum\limits_{(i,j) \in U_k} \delta_{ij,a} \, e_{ij};$

$\quad r_i = g_i - \sum\limits_{j} x^{+}_{ij}; \quad s_j = h_j - \sum\limits_{i} x^{+}_{ij}; \quad t_k = f_k - \sum\limits_{(i,j) U_k} x^{+}_{ij};$

In matrix notation (7) becomes:

$$(9) \quad \begin{pmatrix} P_{m\,m} & E_{m\,n'} & U_{m\,\ell} \\ E^T_{n'm} & Q_{n'n'} & V_{n'\ell} \\ U^T_{\ell\,m} & V^T_{\ell\,m'} & W_{\ell\ell} \end{pmatrix} \begin{pmatrix} \kappa_m \\ \lambda_{n'} \\ \mu_\ell \end{pmatrix} = \begin{pmatrix} r_m \\ s_{n'} \\ t \end{pmatrix} \qquad n' = n - 1$$

P und Q are diagnonal matrices. The dimension subcripts will be omitted in subsequent formulas. If there are no constraints of type (R2) during the first step (i.e. $1 = 0$), system (9) is solved easily:

$$(10) \quad \begin{pmatrix} \kappa \\ \ell \end{pmatrix} = \begin{pmatrix} A & -C^T \\ -C & B \end{pmatrix} \begin{pmatrix} r' \\ s \end{pmatrix} \quad \text{where use is made of the}$$

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

(11) $\quad Z = \begin{pmatrix} A & -C^T \\ -C & B \end{pmatrix} = Z^T; \quad \psi = \begin{pmatrix} \kappa \\ \lambda \end{pmatrix}; \quad \omega = \begin{pmatrix} r \\ s \end{pmatrix};$

Further

(12) $\quad F = \begin{pmatrix} U \\ V \end{pmatrix}, \quad K = (W - F^T Z F)^{-1} = K^T$

$$L = KF^T Z$$

$$M = Z(I + FL) = M^T$$

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

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

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

(14) $\quad \begin{pmatrix} \psi \\ \mu \end{pmatrix} = \begin{pmatrix} \psi_o + ZF\rho \\ - \rho \end{pmatrix}$ 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 $U_{\ell+1}$ for which

$$[15] \quad \sum_{U_{\ell+1}} \hat{x}_{ij} = \tau \text{ ; this new constraint has to be linear}$$

independent of the old [R1] and [R2]. $U_{\ell+1}$ 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 τ and from [6] we can
derive therefore

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

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

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 prefered 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 indiffernce 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

1. C.Almon, "Recent Methodological Advances in Input-Output
      in the US and Canada", Paper presented to the
      4th International Conference on Input-Output-
      Techniques, Geneva, 1968

2. C.Almon, "The Maryland Interindustry Forecasting Model",
      Paper presented to the Symposion on Models of
      Economies held in Novosibirsk, June 1970

3. C.Almon, L.Atkinson, T.C.Reimbold, "Dynamic Interindustry
      Forecasting for Business Planning", Input-Output
      Techniques, eds. A. Brody and A.P.Carter, Amsterdam
      1972

4. M.Bacharach, "Biproportional Matrices & Input-Output-Change",
      Cambridge at the University Press, 1970

5. J. Bates and M.Bacharach, "Input-Output Relationships,
      1954 - 1956", Vol 3 of "A Programme for Growth",
      ed. R. Stone, Cambridge-London, 1963

6. G.A.C. Beerens and C.B. Tilanus, "Alternative input-output
      predictions for the Netherlands, 1948 - 1958,"
      Statistica Neerlandica 20 (1966), pp 19 -29

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

8. Department of Applied Economics, University of Cambridge, "A Computable Model of Economic Growth", _A Programme for Growth_, No. 1, London, 1962

9. E.Fontela , A.Duval, A.Gabus, M.Börlin and C.Velay, "Forecasting technical coefficients and relative prices", _Applications of Input-Output Analysis_, eds. A.P. Carter and A. Bródy, Amsterdam 1970

10. H.J.Jaksch and K. Konrad, "Zur Schätzung von Input-Output Tabellen aus ihren Reihensummen", _Jahrbücher für Nationalökonomie und Statistik_, Bd. 186 (1971) pp. 131 - 138

11. T.I.Matuszewski, P.R. Pitts and J.A. Swayer, "Linear Programming Estimates of Changes in Input Coefficients", _Canadian Journal of Economics and Political Science_ 30 (1964), pp. 203 - 210

12. J.Paelinck and J. Waelbroek, "Etude empirique sur l'évolution de coefficients input-output", _Economie Appliquée_, vol. 16, 1963, pp. 81 - 111

13. J. Paelnick et al., _"Etude Comparée des Tableaux d'Entrée et de Sorties des Communautés Européennes"_, Centre d'Etude et de Recherches Universitaire de Namur, Namur, 1966

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

15. R.Stäglin, "Estimating an Input-Output Matrix with incomplete Information", _Konjunkturpolitik_, Heft 2/3, 1970 pp. 145 - 153

16. R. Stäglin, "MODOP - Ein Verfahren zur Erstellung empiri-
    scher Transaktionsmatrizen", <u>Anwendung statistischer</u>
    <u>und mathematischer Methoden auf sozialwissenschaft-</u>
    <u>liche Probleme,</u> Würzburg 1972

17. R. Stone and J.A.C. Brown, " A Long-Term Growth Model for
    the British Economy," Chapter 10 of <u>"Europés</u>
    <u>Future in Figures",</u> ed. R.C. Geary, Amsterdam, 1962

18. H.Theil, <u>"Applied Economic Forecasting",</u> Amsterdam 1966

19. C.B. Tilanus and H. Theil, "The Information Approach to
    the Evaluation of Input-Output Forecasts, <u>Econometrica</u>
    32 (1965) pp. 847 - 862

20. C.B.Tilanus, <u>"Input-Output Experiments, The Netherlands</u>
    <u>1948 - 1961",</u> Rotterdam University Press, Rotterdam
    1966

21. P. Uribe, C.G. de Leeuw, H. Theil, "The Information
    Approach to the Prediction of Interregional Trade
    Flows", <u>Review of Economic Studies</u> 33 (1966),
    pp. 209 - 220