WestVirginiaUniversity
THE RESEARCH REPOSITORY @ WVU

Theory and Methods RRI Input-Output Archive

1975

Estimating and Projecting Input-Output Coefficients

R. I.G. Allen

W. F. Gossling

A. G. Armstrong

T. S. Barker

H. W. Fisher

See next page for additional authors

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

Authors

R. I.G. Allen, W. F. Gossling, A. G. Armstrong, T. S. Barker, H. W. Fisher, and J. R. C. Lecomber

ESTIMATING AND PROJECTING INPUT-OUTPUT COEFFICIENTS

Edited by
R.I.G.ALLEN & W.F.GOSSLING

with papers by:
R.I.G.ALLEN
A.G.ARMSTRONG
T.S.BARKER
H.W.FISHER
J.R.C.LECOMBER

Foreword by:
PROFESSOR I.G.STEWART

Randall W. Jackson

Randall W. Jackson

ESTIMATING AND PROJECTING INPUT–OUTPUT COEFFICIENTS

Edited by

R. I. G. ALLEN & W. F. GOSSLING

with papers by

R. I. G. ALLEN
A. G. ARMSTRONG
T. S. BARKER
H. W. FISHER
J. R. C. LECOMBER

Foreword by

PROFESSOR I. G. STEWART
(*Edinburgh University*)

INPUT–OUTPUT PUBLISHING COMPANY
3, WYNDHAM PLACE, LONDON W1H 1AP, ENGLAND
1975

ii and iii are missing

Copyright © 1975 Input–Output Publishing Company

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, electrical or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the publishers.

First published 1975

MADE AND PRINTED BY
PAGE BROS (NORWICH) LTD.

FOREWORD

Input–output analysis, whatever its earlier antecedents, is very much the creation of one man, the very distinguished American economist Professor Wassily Leontief. His bold and pioneering study *The Structure of American Economy* published some thirty five years ago has provided an instrument of analysis which has gradually been adopted in many parts of the world. It is particularly fitting that in a foreword to a collection of seminar papers devoted to improving the art of inter-industry projections a tribute to the founder should have pride of place.

The reactions of economists trained in a neo-classical tradition to the simplifying assumptions upon which input–output analysis rests have been those of the tolerant sceptic. Sceptical because an industrial world in which price elasticities, substitutions between inputs and increasing returns to scale are taken to have very restricted scope flatly contradicts the instinctively plausible tenets of marginal analysis and in the long run does not, indeed cannot correspond to historical reality. Tolerant because, perhaps following Professor Milton Friedman's exhortation to look for fruitful results rather than realistic assumptions in our theorising, economists can be persuaded that the Leontief model has great merit "if it works".

"If it works" means if more accurate projections can flow from using this approach to inter-industry problems than can be obtained by other practicable methods of investigation. The tests therefore have become a matter of sensing the stability of input–output coefficients and, where possible, of revising and up-dating parts of the model without having to go to all the labour and expense, not to mention delay, in recalculating a fresh set of input–output coefficients from a new array of inter-industry sales and purchases.

The papers contained in this book represent a series of steps in this process of testing improvements in the Leontief model, improvements designed to facilitate the task of gaining more accurate (some may say less inaccurate) projections of inter-industry i.e. intermediate demands. However, critical one may be of the results and thence of the underlying model(s) one great virtue remains to the credit of input–output anaylsis. No other method of analysis permits simultaneous operational handling of inter-connected outputs and usages. If one is concerned to retain a sense of technological interdependence in handling economic issues, input–output analysis has unquestionable merit.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapters 1, 2, 3 and 5 in this volume were given originally as papers to two meetings of 'input–outputters'; Chapter 2, and R. I. G. Allen's portions of Chapter 3, and Chapter 5 were given at the Edinburgh University Department of Economics' Seminar on Input–Output, April 5th to 6th, 1971, by kind invitation of Professor I. G. Stewart; Chapter 1 and J. R. C. Lecomber's portions of Chapter 3, were given at the Input–Output Research Group's Conference on Input–Output and Throughput held at the University of East Anglia, Norwich, September 13th–16th, 1971, by kind invitation of Professor T. P. Hill (then Dean of the School of Social Studies) and Professor H. A. Lydall (Professor of Economics) University of East Anglia. The Authors and Editors are indebted not only for the hospitality, provided by these Universities, but also for a grant towards the expenses of the Edinburgh Seminar provided by the (U.K.) Social Science Research Council (who also supported Mr. R. I. G. Allen's research studies at Edinburgh University) and a similar grant by the Shell International Petroleum Company Limited in support of the Norwich Conference.

We are grateful to the editors and publisher (Basil Blackwell, Oxford) of the Oxford Bulletin of Economics and Statistics for permission to abstract from Mr. R. I. G. Allen's paper (August 1974) entitled "Some experiments with the RAS method of updating input–output coefficients" and to reprint certain sections and tables appended thereto. Furthermore, the editors are indebted to Dr. Halder W. Fisher of the Battelle Memorial Institute, Columbus, Ohio, U.S.A. for contributing Chapter 6 of this volume.

Various persons at diverse times have assisted with the stages of preparation of the final typescript, at Norwich, Bristol, Cambridge, Edinburgh and London; the editors are particularly indebted to Miss Frances Ware (on the staff of the National Institute of Economic and Social Research) for final typing of the manuscript and for the preparation of the Index. To all of them we offer our sincere thanks.

Norwich R.I.G.A.
October 1974 W.F.G.

INTRODUCTION

In 1936, Leontief published his seminal article "Quantitative Input and Output Relations in the Economic System of the United States" in the Harvard *Review of Economics & Statistics* which included a *Tableau Economique* for the U.S.A. on the basis of the available statistical information. The striking contrast between the central position of an input–output table in economic theory as a means of explaining the interconnections between different sectors of the economy, and the relative poverty and unreliability of much of the available statistical data was later vividly described by Leontief in *The Structure of American Economy* as follows:

> 'One hundred and fifty years ago, when Quesnay first published his famous scheme, his contemporaries and disciples acclaimed it as the greatest discovery since Newton's laws. The idea of general interdependence among the various parts of the economic system has become by now the very foundation of economic analysis. Yet, when it comes to the practical application of this theoretical tool, modern economists must rely exactly as Quesnay did upon fictitious numerical examples ..... Despite the remarkable increase in primary statistical data, the proverbial boxes of theoretical assumptions are in this respect as empty as ever.' ([30], p. 9)

The last thirty years have witnessed a remarkable flowering of official national income and production statistics in the United Kingdom as well as in the United States, with the result that we are now able to fill many of Leontief's 'empty boxes' with data that are for the most part reasonable approximations to the economist's theoretical requirements. The estimation and development of input–output statistics is now an important feature of the government statistical service and input–output relationships form an integral part of national accounts statistics. However, the *ex post* statistical basis of most input–output relationships lead to several perennial sources of difficulty. Although, from the 1970 census onwards, the marginal data of the input–output tables may be compiled annually, the majority of cells within the interindustry matrix, particularly those relating to manufacturing industries, can only be accurately determined on the basis of the full quinquennial Census of Production which details both input and output by industry. Given a delay of some five to six years in the collection and processing of the census material and in the preparation of the input–output tables, a benchmark table is

therefore likely to be at least one decade old before it can be supplanted by a second firmly based table. During this considerable period, many important input–output relationships may have changed substantially as a result of the changing character of productive output, of price substitution or of new techniques of production. Serious bias is likely to be introduced into input–output analysis unless these sources of changes can be accounted for.

The RAS or biproportional method developed by Professor Stone and his colleagues at Cambridge during the early sixties is perhaps the best known and most widely used technique for revising or projecting input–output relationships given only the bare minimum of information —the marginal totals of intermediate output and input by industry for the year in which the adjusted table is required. Subsequently, the method has played an important role in the development of the Cambridge Growth Project and has been widely adopted elsewhere in the United Kingdom and in many other countries. The formal properties of RAS have been explored in considerable detail by Michael Bacharach in his 1970 Cambridge monograph, *Biproportional Matrices and Input–Output Change* [5]. However, since the RAS method was essentially devised as an operational technique and not as a theoretical construct, it is surprising that so little attention has been paid to testing and evaluating its performance.

The main group of papers in this small volume (Chapters 1 to 4) focus attention on this question, while in addition they suggest and explore new lines of departure. Four of the contributors have been involved at various stages in the development of the Cambridge Growth Project. Richard Lecomber in particular was involved in much of the initial development of the RAS method and in its early applications, e.g. to the projections for the 1965 national plan. His first contribution to this volume (Chapter 1) is intended as a general review and critique of the 'state of the art' in updating methods. After summarising the basic principles and mathematics of the RAS adjustment and related procedures, these methods are then contrasted with other solutions of the same limited information problem. He concludes that none of the rival methods are particularly successful but that this is largely attributable to their extremely slender informational basis. There is therefore a strong *a priori* case for incorporating additional information provided that this is sufficiently reliable.

A similar conclusion is reached by Terence Barker in Chapter 2. Barker sets out a series of experiments in which input–output matrices are applied to the problem of projecting intermediate demands for a non-base year. He then compares the performance of RAS with various other approaches that account for changes in input–output relationships in entirely different ways: introducing trends in coefficients, allowing for price substitution and incorporating non-homogeneous production functions. Barker demonstrates that it may be highly inefficient to apply a single assumption to all cells in the interindustry matrix because certain groups of coefficients behave in individualistic ways: we should therefore aim for greater flexibility in the choice and use of our updating models.

In Chapter 3 Richard Allen and Richard Lecomber also develop the case for models of broader scope and greater generality. After demonstrating that projections of intermediate demand are heavily dependent on the reliable estimation of a relatively small number of major cells within the input–output matrix, they argue that resources should be concentrated on providing accurate annual estimates of these coefficients, the RAS method (or other mechanical techniques) being employed as devices for balancing the residual elements of the matrix.

A further problem emphasised in Chapter 3 and also by Barker in Chapter 4 concerns the quality of the exogenous information used in preparing updated input–output tables. Tests on the C.S.O.'s updated tables for 1963 emphasise not only the weaknesses of the naive RAS adjustment but also suggest that inaccuracy in the exogenous estimates (especially of the row and column totals) may be a major source of error in the updated coefficient matrix. Lecomber and Allen therefore propose and test a modified version of RAS which allows for a greater range of exogenous information and which may also reflect judgments about the relative reliability of this information.

The inclusion of judgmental factors into economic models has tended to be regarded with some suspicion in scientific circles but there are precedents elsewhere in the input–output field. For example, the Battelle Memorial Institute of Columbus, Ohio, U.S.A., has pioneered a technique (the '*ex ante*' approach) for estimating and projecting U.S. coefficients which largely eschews the conventional use of *ex post* industrial survey statistics. While this novel approach should be used with care, there is clearly considerable scope for the discriminating application of industrial and technical expertise to the estimation problems faced by input–output statisticians; we append a brief introductory paper on the *ex ante* approach to forecasting input–output coefficients (mostly from technological data) by Dr. Halder W. Fisher (Chapter 6).

Chapters 1 to 4 are primarily concerned with problems of adjusting or projecting input–output relationships through time given—as a starting point—a firmly based benchmark table. In practice, the estimation of a benchmark matrix from the basic census data provides enormous difficulties, particularly if the table is required in its 'pure' form showing the input of products into products. The above *ex-ante* method provides a somewhat extreme solution to the problem, a 'practical' alternative which avoids using Census data and depends directly on industrial information. Alan Armstrong who participated in the construction of the

1963 matrix maintains the conventional use of Census data and shows in Chapter 5 that the crucial problem was the treatment of secondary production. The results of his tests are encouraging. Apart from a few cells, there are only small differences between tables calculated according to different assumptions about the secondary product, and increasing the dimensions of the table tends to reduce these differences still further. There would therefore seem to be little gain from developing more sophisticated models.

London & Norwich
September–October 1974

R. I. G. ALLEN
W. F. GOSSLING

CHAPTER 1

A Critique of Methods of Adjusting, Updating and Projecting Matrices

J. R. C. LECOMBER
University of Bristol

1.1 INTRODUCTION

Input–output workers, faced with large matrices and scanty data, must often resort to mechanical techniques of adjustment, updating and projection, among which the best-known is probably the biproportional or 'RAS' method in which a matrix is adjusted to sum to given row and column totals by successive pro-rating of its rows and columns. Since the early application by Deming and Stephan [14][1] in the demographic field, the RAS method has been extended in many directions. The first important economic application of RAS in the United Kingdom was by the Cambridge Growth Project [11] in which an input–output matrix for 1960 was estimated given the row and column totals for 1960 and a comparable matrix for 1954. This updating problem was formally equivalent to the adjustment problem just described. The row and column multipliers (respectively r and s) were however given an econometric interpretation and were used to project the matrix to 1966.

Since the Growth Project application, the RAS method has been widely used in the input–output field (e.g. Department of Economic Affairs [51], Johansen [26], Fontela [16], Upton [53] and Central Statistical Office [48] [49] [50]). It has also been applied by Stone and Leicester [39] and Evans and Lindley [15] to analyse the changes in employment cross-classified by occupation and industry, and by a number of researchers in the international trade field (e.g. Waelbroeck [54] [55], Bénard [8], Kouevi [27], and Grandville [22]). Bacharach [5] has pioneered an application of RAS to the analysis of the Markov process.[2]

This chapter compares RAS with alternative methods of solving the same minimal-information problems and the accumulating evidence on their reliability is briefly reviewed. The extension of the RAS method to situations where additional information is available is then considered.

[1] All footnotes are at the end of the Chapter on pp. 22–4. Likewise for the remaining Chapters.

1.2 THE ADJUSTMENT OF MATRICES

Consider the construction of an input–output table, or matrix, **X**, from Census returns. Since these returns are incomplete and also somewhat inaccurate, much judgement must be exercised and it is hardly surprising if the initial table, ${}_0\mathbf{X}$,[3] fails to conform with other information. In particular, estimates *may* be derived of total inputs and total intermediate sales of each industry, on the basis of Census and other information [see Cambridge D.A.E. [11], Paelinck and Waelbroeck [34], Central Statistical Office [46]]. These totals are of course (estimates of) the row and column sums of **X**. If these are regarded as reliable, then the next task must be to adjust the initial matrix ${}_0\mathbf{X}$ to satisfy the (supposedly) known row and column sums.

Much of the literature has been devoted to alternative means of accomplishing this task. For there is not one but an infinity of matrices, $\mathbf{X}^*$, satisfying the constraints $\mathbf{X}^*\,\mathbf{i} = \mathbf{u}$, $\mathbf{X}^{*\prime}\mathbf{i} = \mathbf{v}$. Manifestly, since ${}_0\mathbf{X}$ is the best available estimate of **X**, we seek to minimise the adjustments, that is to find a matrix $\mathbf{X}^*$ which is in some sense as close as possible to ${}_0\mathbf{X}$. Unfortunately, however, 'closeness' is not a very clearly defined concept. No fewer than five criteria have been suggested and these are tabulated in Table 1.1 opposite. Each entails a different procedure for estimating **X** and in general each results in a different estimate. Since this paper is concerned less with the rival merits of different minimands (see Bacharach[5]) than with features and problems common to all, it will be sufficient to concentrate on the two most commonly used—the closely related Cambridge (RAS) and Friedlander methods. The adjustment processes corresponding to these two methods are detailed in Table 1.2, overleaf. The non-similarity of the methods is apparent.

The Friedlander minimand is perhaps the most appealing, but unfortunately the adjustment (in common with all but RAS) fails to preserve signs. In many applications, negative elements are *a priori* impossible, but even where ${}_0\mathbf{X}$ is non-negative, negative elements can, and do, appear in $\mathbf{X}^*$; though generally small and few (Omar, [33] Bacharach [5]), these negatives must somehow be removed. If additional constraints $x^*_{ij} \geqslant 0$ are imposed, the simple iterative adjustment is no longer available and $\mathbf{X}^*$ must be found by quadratic programming. However the resultant problem makes heavy demands on computer time and programming ingenuity (Omar [33]).[4] It is therefore an important advantage of the similar RAS adjustment, that signs are automatically preserved. Available evidence (Schneider [35], Omar [33]) suggests that, provided the initial discrepancies are not large, the results are not very sensitive to the method used. The comparative simplicity of RAS over the other methods that preserve signs is an overriding asset.[5]

Such procedures are appropriate in a wide range of situations when matrices (or, more generally *n*-dimensional arrays) require adjustment to

Table 1.1 Adjustment Procedures

No.	*Minimand*	*Proposed by*	*Further application or discussion*	*Functional relationship*	*Adjustment procedure*	*Notes*
1	$\sum(x^*_{ij} - {}_0x_{ij})^2$	Almon [3]	Bacharach [5] Geary [18]	$\mathbf{X}^* = \hat{\mathbf{r}}\mathbf{H} + {}_0\mathbf{X} + \mathbf{H}\hat{\mathbf{s}}$ where $\mathbf{H} = \mathbf{ii}'$	Iterative allocation of row and column discrepancies equally to elements concerned, or, direct solution	Large proportional changes to small elements
2	$\sum \frac{(x^*_{ij} - {}_0x_{ij})^2}{{}_0x_{ij}}$	Friedlander [17]	Stephan [36] Omar [33] Bacharach [5] Henry [24] [25]	$\mathbf{X}^* = \hat{\mathbf{r}}\,{}_0\mathbf{X} + {}_0\mathbf{X} + {}_0\mathbf{X}\hat{\mathbf{s}}$	Iterative allocation of row and column discrepancies pro-rata with elements of ${}_0\mathbf{X}$, or, direct solution	
3	$\sum \frac{\lvert x^*_{ij} - {}_0x_{ij}\rvert}{{}_0x_{ij}}$	Matuszewski *et al.* [32]	Schneider [35] Bacharach [5]	None	Linear programming	Computationally heavy
4	$\sum\left(x^*_{ij}\log\frac{x^*_{ij}}{{}_0x_{ij}}\right)$	Cambridge DAE [11] RAS or biproportional	Bacharach [5] Omar [33] Waelbroeck [54] Upton [53]	$\mathbf{X}^* = \hat{\mathbf{r}}\,{}_0\mathbf{X}\hat{\mathbf{s}}$	Iterative allocation or row and column discrepancies pro-rata with elements obtained at previous iteration	Signs of ${}_0\mathbf{X}$ preserved
5	$\sum\left({}_0x_{ij}\log\frac{{}_0x_{ij}}{x_{ij}}\right)$	Theil [40]	Bacharach [5]	None	Intractable. Theil suggests approximating by procedure 2	

Table 1.2 Friedlander and RAS Adjustment Processes

	Friedlander[1]	*RAS*
nth row adjustment	$\mathbf{X}_{2n-1} = \mathbf{X}_{2n-2} + \hat{\mathbf{r}}_n\, {}_0\mathbf{X}$ where $\mathbf{r}^n = (\widehat{\mathbf{u} - \mathbf{X}_{2n-2}\mathbf{i}})^{-1}\, {}_0\mathbf{X}\mathbf{i}$ ensuring $\mathbf{X}_{2n-1}\mathbf{i} = \mathbf{u}$	$\mathbf{X}_{2n-1} = \mathbf{X}_{2n-2} + \hat{\mathbf{r}}_n \mathbf{X}_{2n-2}$ $= (\hat{\mathbf{r}}_n + \mathbf{I})\, \mathbf{X}_{2n-2}$ where $\mathbf{r}_n = (\widehat{\mathbf{u} - \mathbf{X}_{2n-2}\mathbf{i}})^{-1}\, \mathbf{X}_{2n-2}\, \mathbf{i}$ ensuring $\mathbf{X}_{2n-1}\, \mathbf{i} = \mathbf{u}$
nth column adjustment	$\mathbf{X}_{2n} = \mathbf{X}_{2n-1} + {}_0\mathbf{X}\hat{\mathbf{s}}_n$ where $\mathbf{s}_n = (\widehat{\mathbf{v} - \mathbf{X}_{2n-1}'\mathbf{i}})^{-1}\, {}_0\mathbf{X}'\mathbf{i}$ ensuring $\mathbf{X}_{2n-1}'\, \mathbf{i} = \mathbf{v}$	$\mathbf{X}_{2n} = \mathbf{X}_{2n-1} + \mathbf{X}_{2n-1}\, \hat{\mathbf{s}}_n$ $= \mathbf{X}_{2n-1}(\hat{\mathbf{s}}_n + \mathbf{I})$ where $\mathbf{s}_n = (\widehat{\mathbf{v} - \mathbf{X}_{2n-1}'\mathbf{i}})^{-1}\, \mathbf{X}_{2n-1}'\, \mathbf{i}$ ensuring $\mathbf{X}_{2n}'\, \mathbf{i} = \mathbf{v}$
Form of final matrix[2]	$\mathbf{X}^* = {}_0\mathbf{X} + (\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2 \ldots)\, {}_0\mathbf{X}$ $+ {}_0\mathbf{X}(\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2 \ldots)$ which is of form $\mathbf{X}^* = {}_0\mathbf{X} + \hat{\mathbf{r}}\, {}_0\mathbf{X} + {}_0\mathbf{X}\hat{\mathbf{s}}$	$\mathbf{X}^* = (\mathbf{I} + \hat{\mathbf{r}}_1)(\mathbf{I} + \hat{\mathbf{r}}_2) \ldots$ ${}_0\mathbf{X}(\mathbf{I} + \hat{\mathbf{s}}_1)(\mathbf{I} + \hat{\mathbf{s}}_2) \ldots$ which is of form $\mathbf{X}^* + \hat{\mathbf{r}}\, {}_0\mathbf{X}\hat{\mathbf{s}}$

Notes: 1. The Friedlander adjustment may also be obtained directly by solving a set of simultaneous equations (Henry [24]).

2. Assuming convergence—see Bacharach [5].

conform with conflicting estimates of row and column sums. For example, the original application (Deming and Stephan [14], Stephan [36]) involved the adjustment of a cross-tabulation of population characteristics, derived from a sample census, to fit marginal totals from complete enumeration. In another well-known application (Cambridge D.A.E. [11]) ${}_0\mathbf{X}$ was an input–output table (or it might be a cross-tabulation of international trade flows or if labour skills by industry) on an out-of-date classification and $\mathbf{u}$, $\mathbf{v}$, the marginal totals on a new classification.

The strength of these procedures is their minimal data requirements; ${}_0\mathbf{X}$, $\mathbf{u}, \mathbf{v}$, only. Indeed where more information is available, for example about the sources of error in ${}_0\mathbf{X}$, they cease to be satisfactory.

Consider again the original example of constructing an input-output table from Census returns. Suppose the initial estimate, ${}_0\mathbf{X}$, was obtained by summing those purchases which are identifiable in terms of the headings of the matrix. If purchases recorded under vague headings such as 'materials not elsewhere specified' and 'replacements parts' are omitted, the elements of ${}_0\mathbf{X}$ will generally fall short of the true values of the intermediate demand matrix, $\mathbf{X}$. ${}_0\mathbf{X}$ is thus biased but, if nothing is known about the relative bias of different elements, a mechanical adjustment may still be appropriate. However this is rarely the case. For example, 'materials not elsewhere specified', though vague, at least excludes fuels

and services; moreover materials 'not elsewhere specified' by an iron foundry may include drawing pins and paint but are unlikely to include iron ore. Nor will all industries show an equal tendency to include paint under some vague head. For the motor industry, this is a major input likely to be specifically recorded. If ${}_0\mathbf{X}$ is adjusted using RAS, most of the unrecorded paint will be allocated to major users, such as motors, precisely the places where unrecorded paint is least likely to go. It is thus vital that ${}_0\mathbf{X}$ be corrected for bias (other than biproportional bias) before a mechanical adjustment procedure is applied. ${}_0\mathbf{X}$ will henceforth denote a matrix that has already been so corrected.

Secondly, some elements of ${}_0\mathbf{X}$ are likely to be more accurate than others. In these circumstances, Stephan [36] suggested minimising $\sum_{ij}[(x_{ij}^* - {}_0x_{ij})^2/e_{ij}]$ instead of $\sum[(x_{ij}^* - {}_0x_{ij})^2/{}_0x_{ij}]$ where $\mathbf{E}$ is a matrix of standard of errors attached to elements of ${}_0\mathbf{X}$ and where e_{ij} is the ijth element of matrix $\mathbf{E}$. It is easily shown that

$$\mathbf{X}^* = \hat{\mathbf{r}}\mathbf{E} + {}_0\mathbf{X} + \mathbf{E}\hat{\mathbf{s}} = \hat{\boldsymbol{\rho}}\mathbf{E} + ({}_0\mathbf{X} - \mathbf{E}) + \mathbf{E}\hat{\boldsymbol{\sigma}} \tag{1.2.1}$$

${}_0\mathbf{X}$ was, in this case, obtained from random samples, so that standard errors were readily available. However, in many applications, such as the construction of input–output tables, no formal error estimates can be made: while it is known that industries comprising a few large firms with well developed accounting systems are likely to provide better figures than more fragmented industries and that elements including a large allocation of 'materials not elsewhere specified' are likely to be particularly inaccurate, it is impossible to find any fully objective basis for quantifying this information. Many econometricians display considerable reluctance to introduce subjective elements; some would doubtless prefer to use the standard procedures as in some sense objective, although to do so would be to assume (implicitly) uniform reliability throughout ${}_0\mathbf{X}$. If nothing is known about the relative accuracy of individual elements, this is reasonable enough as a neutral assumption. But, in general, the necessity of guessing relative[6] errors cannot be avoided. The assumption that all are equal is generally nothing more than an unnecessary bad guess, likely to give inaccurate results.

The above procedure, like the simple Friedlander adjustment, fails to preserve signs. This suggests the possibility of using an analogous RAS adjustment[7]

$$\mathbf{X}^* = ({}_0\mathbf{X} - \mathbf{E}) + \hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}} \tag{1.2.2}$$

Provided $0 \leqslant e_{ij} \leqslant {}_0x_{ij}$, for all i and j, and, further, that the elements of $\mathbf{E}$ are sufficiently large that the control vectors, $\mathbf{u} - ({}_0\mathbf{X} - \mathbf{E})\mathbf{i}$ and $\mathbf{v} - ({}_0\mathbf{X} - \mathbf{E})'\mathbf{i}$ are non-negative, an appeal to the standard RAS results (Bacharach [5]) yields the following conclusions:

(*1*) ${}_0\mathbf{X} - \mathbf{E}$, $\mathbf{E}$, $\hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}$ and hence $\mathbf{X}^*$ are non-negative;

(*2*) Zeros in ${}_0\mathbf{X}$ will be preserved in $\mathbf{E}$, $\hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}$ and hence $\mathbf{X}^*$.

Table 1.2 Friedlander and RAS Adjustment Processes

	Friedlander[1]	*RAS*
nth row adjustment	$\mathbf{X}_{2n-1} = \mathbf{X}_{2n-2} + \hat{\mathbf{r}}_n\,{}_0\mathbf{X}$ where $\mathbf{r}^n = (\widehat{\mathbf{u} - \mathbf{X}_{2n-2}\mathbf{i}})^{-1}\,{}_0\mathbf{X}\mathbf{i}$ ensuring $\mathbf{X}_{2n-1}\mathbf{i} = \mathbf{u}$	$\mathbf{X}_{2n-1} = \mathbf{X}_{2n-2} + \hat{\mathbf{r}}_n\mathbf{X}_{2n-2}$ $= (\hat{\mathbf{r}}_n + \mathbf{I})\mathbf{X}_{2n-2}$ where $\mathbf{r}_n = (\widehat{\mathbf{u} - \mathbf{X}_{2n-2}\mathbf{i}})^{-1}\mathbf{X}_{2n-2}\mathbf{i}$ ensuring $\mathbf{X}_{2n-1}\mathbf{i} = \mathbf{u}$
nth column adjustment	$\mathbf{X}_{2n} = \mathbf{X}_{2n-1} + {}_0\mathbf{X}\hat{\mathbf{s}}_n$ where $\mathbf{s}_n = (\widehat{\mathbf{v} - \mathbf{X}_{2n-1}'\mathbf{i}})^{-1}\,{}_0\mathbf{X}'\mathbf{i}$ ensuring $\mathbf{X}_{2n-1}'\mathbf{i} = \mathbf{v}$	$\mathbf{X}_{2n} = \mathbf{X}_{2n-1} + \mathbf{X}_{2n-1}\hat{\mathbf{s}}_n$ $= \mathbf{X}_{2n-1}(\hat{\mathbf{s}}_n + \mathbf{I})$ where $\mathbf{s}_n = (\widehat{\mathbf{v} - \mathbf{X}_{2n-1}'\mathbf{i}})^{-1}\mathbf{X}_{2n-1}'\mathbf{i}$ ensuring $\mathbf{X}_{2n}'\mathbf{i} = \mathbf{v}$
Form of final matrix[2]	$\mathbf{X}^* = {}_0\mathbf{X} + (\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2 \ldots)\,{}_0\mathbf{X}$ $+ {}_0\mathbf{X}(\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2 \ldots)$ which is of form $\mathbf{X}^* = {}_0\mathbf{X} + \hat{\mathbf{r}}\,{}_0\mathbf{X} + {}_0\mathbf{X}\hat{\mathbf{s}}$	$\mathbf{X}^* = (\mathbf{I} + \hat{\mathbf{r}}_1)(\mathbf{I} + \hat{\mathbf{r}}_2)\ldots$ ${}_0\mathbf{X}(\mathbf{I} + \hat{\mathbf{s}}_1)(\mathbf{I} + \hat{\mathbf{s}}_2)\ldots$ which is of form $\mathbf{X}^* + \hat{\mathbf{r}}\,{}_0\mathbf{X}\hat{\mathbf{s}}$

Notes: 1. The Friedlander adjustment may also be obtained directly by solving a set of simultaneous equations (Henry [24]).
2. Assuming convergence—see Bacharach [5].

conform with conflicting estimates of row and column sums. For example, the original application (Deming and Stephan [14], Stephan [36]) involved the adjustment of a cross-tabulation of population characteristics, derived from a sample census, to fit marginal totals from complete enumeration. In another well-known application (Cambridge D.A.E. [11]) ${}_0\mathbf{X}$ was an input–output table (or it might be a cross-tabulation of international trade flows or if labour skills by industry) on an out-of-date classification and $\mathbf{u}$, $\mathbf{v}$, the marginal totals on a new classification.

The strength of these procedures is their minimal data requirements; ${}_0\mathbf{X}$, $\mathbf{u}$, $\mathbf{v}$, only. Indeed where more information is available, for example about the sources of error in ${}_0\mathbf{X}$, they cease to be satisfactory.

Consider again the original example of constructing an input-output table from Census returns. Suppose the initial estimate, ${}_0\mathbf{X}$, was obtained by summing those purchases which are identifiable in terms of the headings of the matrix. If purchases recorded under vague headings such as 'materials not elsewhere specified' and 'replacements parts' are omitted, the elements of ${}_0\mathbf{X}$ will generally fall short of the true values of the intermediate demand matrix, $\mathbf{X}$. ${}_0\mathbf{X}$ is thus biased but, if nothing is known about the relative bias of different elements, a mechanical adjustment may still be appropriate. However this is rarely the case. For example, 'materials not elsewhere specified', though vague, at least excludes fuels

and services; moreover materials 'not elsewhere specified' by an iron foundry may include drawing pins and paint but are unlikely to include iron ore. Nor will all industries show an equal tendency to include paint under some vague head. For the motor industry, this is a major input likely to be specifically recorded. If ${}_0\mathbf{X}$ is adjusted using RAS, most of the unrecorded paint will be allocated to major users, such as motors, precisely the places where unrecorded paint is least likely to go. It is thus vital that ${}_0\mathbf{X}$ be corrected for bias (other than biproportional bias) before a mechanical adjustment procedure is applied. ${}_0\mathbf{X}$ will henceforth denote a matrix that has already been so corrected.

Secondly, some elements of ${}_0\mathbf{X}$ are likely to be more accurate than others. In these circumstances, Stephan [36] suggested minimising $\sum_{ij}[(x_{ij}^* - {}_0x_{ij})^2/e_{ij}]$ instead of $\sum[(x_{ij}^* - {}_0x_{ij})^2/{}_0x_{ij}]$ where $\mathbf{E}$ is a matrix of standard of errors attached to elements of ${}_0\mathbf{X}$ and where e_{ij} is the ijth element of matrix $\mathbf{E}$. It is easily shown that

$$\mathbf{X}^* = \hat{\mathbf{r}}\mathbf{E} + {}_0\mathbf{X} + \mathbf{E}\hat{\mathbf{s}} = \hat{\boldsymbol{\rho}}\mathbf{E} + ({}_0\mathbf{X} - \mathbf{E}) + \mathbf{E}\hat{\boldsymbol{\sigma}} \qquad (1.2.1)$$

${}_0\mathbf{X}$ was, in this case, obtained from random samples, so that standard errors were readily available. However, in many applications, such as the construction of input–output tables, no formal error estimates can be made: while it is known that industries comprising a few large firms with well developed accounting systems are likely to provide better figures than more fragmented industries and that elements including a large allocation of 'materials not elsewhere specified' are likely to be particularly inaccurate, it is impossible to find any fully objective basis for quantifying this information. Many econometricians display considerable reluctance to introduce subjective elements; some would doubtless prefer to use the standard procedures as in some sense objective, although to do so would be to assume (implicitly) uniform reliability throughout ${}_0\mathbf{X}$. If nothing is known about the relative accuracy of individual elements, this is reasonable enough as a neutral assumption. But, in general, the necessity of guessing relative[6] errors cannot be avoided. The assumption that all are equal is generally nothing more than an unnecessary bad guess, likely to give inaccurate results.

The above procedure, like the simple Friedlander adjustment, fails to preserve signs. This suggests the possibility of using an analogous RAS adjustment[7]

$$\mathbf{X}^* = ({}_0\mathbf{X} - \mathbf{E}) + \hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}} \qquad (1.2.2)$$

Provided $0 \leqslant e_{ij} \leqslant {}_0x_{ij}$, for all i and j, and, further, that the elements of $\mathbf{E}$ are sufficiently large that the control vectors, $\mathbf{u} - ({}_0\mathbf{X} - \mathbf{E})\mathbf{i}$ and $\mathbf{v} - ({}_0\mathbf{X} - \mathbf{E})'\mathbf{i}$ are non-negative, an appeal to the standard RAS results (Bacharach [5]) yields the following conclusions:

(*1*) ${}_0\mathbf{X} - \mathbf{E}$, $\mathbf{E}$, $\hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}$ and hence $\mathbf{X}^*$ are non-negative;
(*2*) Zeros in ${}_0\mathbf{X}$ will be preserved in $\mathbf{E}$, $\hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}$ and hence $\mathbf{X}^*$.

The formula may be used even when the initial matrix includes negative elements; it is necessary only that the elements of **E** (and the control totals $\mathbf{u} - {}_0\mathbf{Xi}$ and $\mathbf{v} - {}_0\mathbf{X'i}$) be non-negative.

This adjustment includes the special case where certain elements (${}_0x_{ij}$) are known to be accurate, but nothing is known about the relative accuracy of the other elements. The corresponding e_{ij} are set equal to zero and all other elements of **E** equal to the corresponding elements of ${}_0\mathbf{X}$.

More generally e_{ij} may vary from 0, for an element known with certainty, to ${}_0x_{ij}$, as the putative accuracy decreases. However, if **E** is viewed as a matrix of standard errors, then a certain arbitrariness becomes apparent, in that $\mathbf{X^*}$ is not invariant with the multiplication of **E** by a scalar. In general, the smaller the scalar, the further $\mathbf{X^*}$ moves from the matrix obtained by simple RAS. This arbitrariness is scarcely an argument for preferring the latter,[8] but it is perhaps an argument in favour of the modified Friedlander adjustment, which avoids this difficulty.

Finally, it may be appropriate to question the accuracy of **u** and **v**. The row totals of the input–output matrix, generally obtained by deducting estimates of final demands from estimates of total supplies (home output plus imports), are particularly unreliable. The Friedlander method may readily be generalised to take account of errors (${}_u\mathbf{e}$ and ${}_v\mathbf{e}$) in the estimates of the row and column totals (${}_0\mathbf{u}$ and ${}_0\mathbf{v}$). $\mathbf{X^*}$, $\mathbf{u^*}$ and $\mathbf{v^*}$ are then given by

$$\mathbf{X^*} = ({}_0\mathbf{X} - \mathbf{E}) + \hat{\mathbf{r}}\mathbf{E} + \mathbf{E}\hat{\mathbf{s}} \qquad (1.2.3)$$

$$\mathbf{u^*} = ({}_0\mathbf{u} - {}_u\mathbf{e}) - \mathbf{r}\,{}_u\mathbf{e} \qquad (1.2.4)$$

$$\mathbf{v^*} = ({}_0\mathbf{v} - {}_v\mathbf{e}) - \mathbf{s}\,{}_v\mathbf{e} \qquad (1.2.5)$$

where

$$\mathbf{X^*i} = \mathbf{u^*}, \text{ and } \mathbf{X^{*\prime}\,i} = \mathbf{v^*} \qquad (1.2.6)$$

This adjustment may be shown to minimise

$$\sum_{ij} \frac{(x^*_{ij} - {}_0x_{ij})^2}{e_{ij}} + \sum_i \frac{(u^*_i - {}_0u_i)^2}{{}_ue_i} + \sum_j \frac{(v^*_j - {}_0v_j)^2}{{}_ve_j}$$

An analogous adjustment is available, which has the usual advantage of preserving signs both in **X** and in **u** and **v** (see Chapter 3 below).

1.3 THE UPDATING OF MATRICES

Now consider the problem of updating a matrix relating to one year (0) to a later year (θ), for which only the marginal totals (${}_\theta\mathbf{u}$, ${}_\theta\mathbf{v}$) are available. This is an important problem for input–output analysis, since full Censuses of Production are held infrequently and take many years to process. When, for example, the U.K. national plan was prepared in 1964–5, the latest Census-based input–output tables related to 1954. Meanwhile input–output relationships had changed substantially and the 1954 tables could not be used without due allowance for such change.

The problem is by no means confined to the input-output field: it has also been found necessary to update matrices of the labour force (cross-classified by occupation and industry) [Stone and Leicester [39]], and of international trade flows (by origin and destination) [Waelbroeck [54]].

Formally, the updating problem may be viewed as a special case of the adjustment problem just considered. In the absence of further information ${}_0\mathbf{X}$ is the best available estimate of ${}_\theta\mathbf{X}$ and the simple Friedlander or RAS adjustment is appropriate. But the row and column multipliers may be endowed with greater significance, as describing and perhaps helping to explain temporal change in **X**. They may then be tested for economic plausibility and used to project the matrix to a later year (see section 1.4 below).

Consider briefly the possible economic significance of the multipliers in an input–output update. First suppose that all the data relate to current price flows. Then, if average input–output coefficients and the relative prices at which a given commodity is sold to different users are both invariant through time we obtain

$${}_\theta\mathbf{X} = \hat{\mathbf{p}}\,{}_0\mathbf{X}\hat{\mathbf{q}} \qquad (1.3.1)$$

where ${}_t\mathbf{p}$, ${}_t\mathbf{q}$ are vectors of commodity prices and outputs in year t ($t = 0, \theta$) and $\mathbf{p} = {}_0\hat{\mathbf{p}}^{-1}\,{}_\theta\mathbf{p}$, $\mathbf{q} = {}_0\hat{\mathbf{p}}^{-1}\,{}_\theta\mathbf{q}$. This expression is biproportional in form and incorporates the main sources of change in **X**. Notice that in such a case, a Friedlander adjustment to **X** is, *a priori*, inappropriate.

In fact, relative prices at which any commodity is sold do change, partly because of varying demand conditions, partly because of variations in product mix. It is possible that prices are determined so that their changes appear to be governed by biproportional row-and-column adjustments, or that demand elasticities are such as to preserve biproportionality in the value matrix. While this last possibility could provide a justification for updating the matrix in value terms (see Tilanus [42]) it is generally preferable to make an explicit separation between price and volume changes. The first stage, which may itself be viewed as an updating problem, is then to re-express the basic matrix in the price of year θ. The second stage and the explicit concern of this section is to adjust the resultant matrix (henceforth ${}_0\mathbf{X}$) for changes in volume. Then it input–output coefficients are constant

$${}_\theta\mathbf{X} = {}_0\mathbf{X}\hat{\mathbf{q}} \qquad (1.3.2)$$

However, it is well known that coefficients are subject to substantial changes over time and it is observed that ${}_\theta\mathbf{X}$, if thus derived, fails to satisfy the row and column constraints, ${}_\theta\mathbf{u}$ and ${}_\theta\mathbf{v}$. Apart from errors of measurement, this inconsistency must necessarily be attributed to changes in the matrix of coefficients, $\mathbf{A} = \mathbf{X}\hat{\mathbf{q}}^{-1}$. Stone [37] advances the hypothesis that elements of **A** are subject to substitution effects (substitution of one input for another) and fabrication effects (more or less value added to the inputs) and that these effects act evenly over rows and columns, so that

$$_{\theta}\mathbf{A} = \hat{\mathbf{r}}\,{}_{0}\mathbf{A}\,\hat{\mathbf{s}} \tag{1.3.3}$$

$$_{\theta}\mathbf{X} = \hat{\mathbf{r}}\,{}_{0}\mathbf{X}\,\hat{\mathbf{s}}\,\hat{\mathbf{q}} \tag{1.3.4}$$

Both expressions are biproportional. However, the assumptions on which they are based, while not implausible, have no special economic justification. To assess the method, one must turn to the evidence.

Tests by Paelinck and Waelbroeck [34] (on Belgian data 1953–9) and Schneider [35] (U.S. 1947–58) show the RAS update to perform somewhat better than $_{0}\mathbf{A}$ as an estimate of $_{\theta}\mathbf{A}$.[9] For example, the number of errors in individual cells of over one per cent is reduced from 17 to 9 in the Belgian case, 121 to 103 in the American case. Schneider also shows the RAS method to perform rather better than Matuszewski's [32] linear programming method.

The Cambridge Growth Project updating exercise (Cambridge D.A.E. [11] (U.K. 1954–60)) is also instructive, for while no direct test has been made, certain implausibilities are striking. For example the column multiplier for aircraft is 0.34 and most coefficients in this column are accordingly reduced by two-thirds. This curious result may be explained by reference to the aircraft row, which comprises aircraft into aircraft (largely engines and parts transferred between establishments for embodiment in the final aircraft) and aircraft into transport. The latter coefficient probably rose sharply, reflecting the increased share of air transport in transport as a whole; any change in the former coefficient was probably small.[10] RAS managed to achieve some approximation to this result by combining a row multiplier of 3.65 with the low column multiplier previously quoted. Nevertheless both coefficients were almost certainly seriously misestimated (and have later been revised *ad hoc*), with serious repercussions on the whole aircraft column. Clearly in such circumstances any technical interpretation of the column multiplier would be out of place. A rather different problem is exemplified by the input of agriculture into food processing which falls by over fifty per cent as a direct consequence of an agricultural row multiplier of 0.37. These falls result from a probable misestimation of the row total of agriculture derived as the (small) difference between total supplies and final demands. This important source of error was not explored by Paelinck and Waelbroeck or Schneider.[11]

Some of these difficulties are magnified in the international trade applications of RAS. Bénard [8] updated matrices of international flows between industrial countries over the period 1953–57; the mean square relative error of the matrix obtained was ten per cent. A similar exercise

over the longer period 1953–60 produced larger errors, though in the main these could be explained (by the formation of the Common Market). Waelbroeck [54] [55] used departures from biproportionality as a measure of the effects of the Common Market. Kouevi [27] conducted a series of updating exercises and obtained corresponding time series of error for each element. Graphs of these showed trends and cycles indicating systematic departures from biproportionality.

The various shortcomings of RAS are shared by all mechanical procedures using the same information ($_{0}\mathbf{X}$, $_{\theta}\mathbf{u}$, $_{\theta}\mathbf{v}$). It is no surprise that the various methods yield very similar results (Schneider [35], and Bacharach [5]) and, in these circumstances, the convenience of RAS is an important advantage. However if the accuracy of RAS and similar procedures is to be improved more information must somehow be incorporated, information on the movements of individual elements and on the putative accuracy of estimates of row and column totals.

Paelinck and Waelbroeck [34] pointed out certain major sources of failure of biproportionality and showed that if certain coefficients generally identifiable in advance could be derived exogenously, a much improved estimate of the whole matrix could be obtained by a simple modification of the RAS routine. Denote by $_{\theta}\mathbf{C}$ a matrix including the exogeneously derived elements and zeros elsewhere and by $_{0}\mathbf{C}$ a corresponding matrix for the base year; estimate

$$_{\theta}\mathbf{X} = \hat{\mathbf{r}}({}_{0}\mathbf{X} - {}_{\theta}\mathbf{C})\,\hat{\mathbf{s}} + {}_{\theta}\mathbf{C} \tag{1.3.5}$$

This formula may be compared with equation (1.2.2). The most important cells related to the inputs of primary fuels into secondary fuels: for example coal, used as a fuel in most industries, is a *raw material* in the production of coke, gas and electricity; it is fortunate that reliable estimates of fuel elements are available in non-Census years in many countries.

This was the only extraneous information taken into account both in the Belgian experiment and in the Cambridge up-dating exercise although much more was available. Annual series existed for major inputs into several industries, notably agriculture and iron and steel. Some cells could be identified with particular sub-products for which annual production series were published. Such series are not entirely reliable, due to imperfect coverage and variations in product mix, but they frequently give a good enough idea of trends to invalidate the RAS estimate. Furthermore, equation (1.3.5) may be further generalised so that a wider variety of information can be taken into account and at the same time allowance made for its varying reliability. This possibility is explored and tested by Allen and Lecomber in Chapter 3 below.

1.4 THE PROJECTION OF MATRICES

Updating the matrix is frequently a prior stage to making projections.

Future changes in input–output relationships must be allowed for and often these can be gauged only from changes that have occurred in the past. If two tables are available on a comparable basis for two or more years then a comparison provides an estimate of past changes. Unfortunately, however, the most recent Census-based table is generally based on a different industrial classification from its predecessors and in any case is so dated that such a comparison would relate to a somewhat distant period. Accordingly, the procedure generally adopted (*e.g.* Cambridge D.A.E. [11], Barker [6]) is to update the most recent table and then use the estimated changes over the updating period as a basis for projection. Formally, then, the problem is to estimate a matrix for a future year (${}_t\mathbf{X}$) given the corresponding matrix in the Census year (${}_0\mathbf{X}$) and row and column totals in a more recent year (${}_\theta\mathbf{u}$, ${}_\theta\mathbf{v}$).

Suppose that the simple RAS updating procedure has been used. A convenient feature of the method is that the estimated matrix (${}_\theta\mathbf{X}^*$) is functionally related to the original matrix ${}_0\mathbf{X}$ by the row and column multipliers:

$$ {}_\theta\mathbf{X}^* = {}_{0\theta}\hat{\mathbf{r}}\,{}_0\mathbf{X}\,{}_{0\theta}\hat{\mathbf{s}} \qquad (1.4.1). $$

These may be used to derive further matrices for other years for which row and column totals are not available. The updating assumption may be generalised as:

$$ {}_t\mathbf{X}^* = {}_{0t}\hat{\mathbf{r}}\,{}_0\mathbf{X}\,{}_{0t}\hat{\mathbf{s}}\,{}_{0t}\hat{\mathbf{q}} \qquad (1.4.2) $$

where ${}_{0t}\mathbf{r}$ and ${}_{0t}\mathbf{s}$ are multipliers and ${}_{0t}\mathbf{q}$ are quantum indices of outputs connecting years 0 and t. Removing the output effects

$$ {}_t\mathbf{A}^* = {}_{0t}\hat{\mathbf{r}}\,{}_0\mathbf{A}\,{}_{0t}\hat{\mathbf{s}} \qquad (1.4.3) $$

If it is assumed that ${}_{0t}\mathbf{r}$ and ${}_{0t}\mathbf{s}$ are simple functions of the time interval t, then ${}_t\mathbf{X}$ may be found in terms of ${}_0\mathbf{X}$ and the updating multipliers. An obvious assumption is that

$$ {}_{0t}\hat{\mathbf{r}} = {}_{0\theta}\hat{\mathbf{r}}^{t/\theta} \qquad {}_{0t}\hat{\mathbf{s}} = {}_{0\theta}\hat{\mathbf{s}}^{t/\theta} \qquad (1.4.4) $$

which, it is easy to show, implies that individual elements of the matrix follow exponential trends.

Unfortunately, however, exponential projection suffers from a serious drawback. The sum of a set of variables each growing exponentially may be shown to grow at a rate of growth which increases through time approaching asymptotically the rate of growth of the fastest growing variables.[12] The column sums of matrices generated in this way will thus tend to rise through time at an increasing rate. The implausibility of unmodified exponential trends in practice is amply demonstrated by Omar [33] and Bacharach [5].

Bacharach proposes adjusting columns of the initial projection pro-

rata to satisfy independent projections of column sums, ${}_t\mathbf{v}^*$. Thus

$$ {}_t\mathbf{A}^* = {}_{0\theta}\hat{\mathbf{r}}^{t/\theta}\,{}_0\mathbf{A}\,{}_{0\theta}\hat{\mathbf{s}}^{t/\theta}\,\hat{\mathbf{w}} \qquad (1.4.5) $$

$\mathbf{w}$ being chosen to ensure ${}_t\mathbf{A}^*\mathbf{i} = {}_t\mathbf{v}^*$. The resultant matrix is functionally related to ${}_0\mathbf{A}$ (and ${}_\theta\mathbf{A}$) and signs of elements are preserved. However, elements of $\mathbf{w}$ are often substantially less than one, to counter the 'explosive' nature of the exponential projection, and hence elements exhibiting a small rise between years 0 and θ tend to fall between years θ and t, as may be seen from a cursory inspection of the modified projections (Cambridge D.A.E. [11], Bacharach [5], Omar [33]). Whether or not such changes sometimes or even frequently occur in the real world, this would not seem a desirable property of a general projection method. In addition both the unmodified and modified exponential method are affected by aggregation.

A linear projection of coefficients (as suggested by the Friedlander updating procedure) avoids this difficulty, but signs of elements are not preserved and the projections by Omar and Bacharach include a number of large negative elements. In general, it may be said that, to avoid negative elements, declining trends must be flattened and, to compensate for this, rising trends must be slowed down. Consider the formula

$$ {}_t\mathbf{A}^* = {}_0\mathbf{A} + \Delta\mathbf{A}f(t), \qquad (1.4.6) $$

where $\Delta\mathbf{A} = ({}_\theta\mathbf{A} - {}_0\mathbf{A})$. The purpose of the scalar $f(t)$ is to 'stretch out' the otherwise linear trends. Accordingly a function is chosen such that $f(\theta) = 1, f'(t) > 0$ and $f''(t) < 0$. This method has the following advantages.

(i) there are no reversals of trend;
(ii) the projection is unaffected by aggregation;
(iii) columns sums are given by an analogous formula (${}_t\mathbf{v}^* = \mathbf{v} + \Delta\mathbf{v}f(t)$) and hence satisfy (i) and (ii) above;
(iv) by suitable choice of $f(t)$ individual elements and columns sums can be prevented from turning negative or violating other *a priori* bounds.

Choice of $f(t)$ is arbitrary certainly, but projection is a somewhat arbitrary process, especially when based on such slender information.

Bacharach's projection method (as defined by equation 1.4.3) has not yet been tested against actual data, but two empirical studies are relevant. Beckerman [7] applied the 1966–70 input–output matrix estimated at Cambridge [11][13] to obtain industry output projections for 1975. A high proportion of these he rejected as implausible, substituting projections obtained by cruder methods.

Omar's study [33] provides a valuable test of Bacharach's projection methods. She worked from an input–output matrix for 1954 and control totals for 1957 through 1962 and used the RAS updating technique to estimate matrices for these years. She applied Bacharach's projection methods to 1954 (year 0) and 1960 (θ) to obtain a matrix for 1966, which may be devoted by ${}_{66}\mathbf{A}^{60}$. Alternative projections, ${}_{66}\mathbf{A}^{\theta}$, may be obtained, using other years in place of 1960 (1954 is, however, year 0, throughout). The projections of individual elements depend on the constraints, if any, imposed on the column sums in 1966; however *ratios* of elements in any column are unaffected, providing a test of more general relevance. Accordingly, the ratios of the two largest[14] elements in each column (R_j) have been calculated for the projections ${}_{66}\mathbf{A}^{57}$, ${}_{66}\mathbf{A}^{60}$ and ${}_{66}\mathbf{A}^{62}$, and the frequency distribution of the percentage differences, $R_j^{\theta} - R_j^{60} = D_j^{\theta}$, is shown in Table 1.3. The magnitude of these differences casts considerable doubt on the reliability of the method.

Table 1.3 Differences between Alternative RAS Projections for 1966 (ratios of largest[14] elements in each column)

Percentage difference, D_j^{θ} (as defined in text) Range:	*1957 and 1960*	*1962 and 1960*
under 1	0	4
1–	1	9
5–	3	4
10–	10	9
50 and over	17	5
Total	31	31

Omar also used the data for 1954 and 1958 to make projections for 1962 (${}_{62}\mathbf{A}^{58}$) and these may be compared with the RAS update (${}_{62}\mathbf{A}^{u}$) year, which satisfies the control totals. The percentage difference between the ratios of the largest elements in each column were again calculated. ($D_j = R_j^{58} - R_j^{u}$), and the results tabulated in Table 1.4 opposite.

Projection on the basis of two sets of observations, the second derived in part from the first, must inevitably be hazardous. These tests illustrate this and emphasise the importance of seeking out and utilising additional information.

1.5 RAS PROJECTIONS WHEN MORE INFORMATION IS AVAILABLE

The RAS method of projection was designed to meet a situation of minimal base information—namely one complete matrix in the series to be estimated and the row and column totals of a second. Several examples of the extension of RAS to the analysis and projection of trends when more

Table 1.4 Differences between RAS Projections and RAS Updates for 1962

Percentage difference, D_j (as defined in text) Range (per cent):	*Ratio of two largest elements in each column*
under 1	3
1–	12
5–	8
10–	7
50 and over	1
Total	31

information is available have been described in the literature. In this section it is suggested that these extensions have tended to stick too closely to the original minimum-information method and in doing so have failed to make efficient use of the available data. A number of stituations are examined in turn.

Firstly suppose a complete matrix is available, plus the row and column totals for a series of later years. This is a situation considered by Omar [33]. Having constructed RAS updates for the later years, she tested two methods of projection, one based on separate projections for each element, the other on projection of rows and column multipliers. The most powerful tests relate to 1962 where an RAS update is also available. She showed that both methods represented a striking improvement on the

Table 1.5 Differences between Alternative RAS Projections and RAS Updates for 1962

Percentage difference Range (per cent):	*Ratio of two largest elements in each column*		
	Bacharach method using data for 1954, 61	*Trends in multipliers*	*Trends in elements*
Under 1	3	5	7
1–	12	19	12
5–	8	2	2
10–	5	2	8
50 and over	3	3	2
Total	31	31	31

simple Bacharach method using data for 1954 and 1958. However her own methods employed not only *more* information but *more recent* information up to 1961). So, in Table 1.5 above her projections are compared with a Bacharach projection using data for 1954 and 1961. This confirms the value of additional data, but does not establish either of her two methods as superior. Omar expresses a preference for the trends-in-multipliers method as far less laborious.

It must however be emphasised that Omar is considering a simple updating process and purely mechanical projections either of elements or multipliers. The trends-in-multipliers approach would be seriously complicated by the more elaborate updating procedures described in section 1.3 above. Moreover it would be difficult to incorporate special information on individual elements (e.g. forecasts of technical developments) if a trends-in-multipliers approach is used. In short the more extraneous information can profitably be taken into account, the more worthwhile it becomes to adopt the more laborious trends-in-elements approach.

Next suppose that complete matrices are available for two years.[15] Many countries, including the United Kingdom, are in this situation in the input–output field.

Stone and Leicester [39] faced this problem in the analysis and projection of labour demand by industry and occupation, using data from two censuses of population. They derived the coefficient matrix $\mathbf{A} = \mathbf{X}\hat{\mathbf{q}}^{-1}$ where (as before) $\mathbf{q}$ is a vector of industry outputs, and estimated

$$ {}_\theta\mathbf{A}^* = {}_{0\theta}\hat{\mathbf{r}}\;{}_0\mathbf{A}\;{}_{0\theta}\hat{\mathbf{s}} \qquad (1.5.1)$$

subject to ${}_\theta\mathbf{X}^*\mathbf{i} = {}_\theta\mathbf{u}$, ${}_\theta\mathbf{X}^{*\prime}\mathbf{i} = {}_\theta\mathbf{v}$ and obtained ${}_t\mathbf{A}^*$ as

$$ {}_t\mathbf{A}^* = ({}_{0\theta}\hat{\mathbf{r}})^{(t-\theta)/\theta}\;{}_\theta\mathbf{A}\;{}_{\theta t}\hat{\mathbf{s}} \qquad (1.5.2)$$

The column (industry) multipliers used for projection (${}_{\theta t}\mathbf{s}$) were related not to time but to growth in output, a type of procedure considered in the appendix. The row multipliers were modified to allow for the different time span, and both sets of multipliers applied to the matrix for year θ. The formula

$$ {}_t\mathbf{A}^* = ({}_{0\theta}\hat{\mathbf{r}})^{t/\theta}\;{}_0\mathbf{A}\;{}_{0t}\hat{\mathbf{s}} \qquad (1.5.3)$$

would appear to be at least as valid. Equation (1.5.3) assumes failures in biproportionality in year t to be uncorrelated with those in year θ, so that divergences between ${}_\theta\mathbf{A}^*$ and ${}_\theta\mathbf{A}$ are ignored. It is difficult to find a comparable justification for equation (1.5.2).

Two further projections may be obtained by reversing the roles of years 0 and θ and thus deriving alternative multipliers ${}_{\theta 0}\mathbf{r}$ and ${}_{\theta 0}\mathbf{s}$. The four projections will coincide only if ${}_0\mathbf{A}$ and ${}_\theta\mathbf{A}$ are connected by an exact biproportionality relationship.

Johansen [26] and Evans and Lindley [15] argue that individual elements contain more information than the row and column sums and that this information should be taken into account in estimating ${}_\theta\mathbf{A}^*$. Accordingly they suggest a least squares minimisation procedure and propose the introduction of a stochastic term (e_{ij}) into (1.5.1). Taking the logarithmic transform, rewriting in suffic notation and rearranging:

$$ \log \frac{{}_\theta a_{ij}}{{}_0 a_{ij}} = \log r_i + \log s_j + \log e_{ij} \qquad (1.5.4)$$

The vectors r and s may be chosen to minimise $\sum_{ij} (\log e_{ij})^2$ [16].

This is equivalent to an analysis of variance of the matrix of percentage changes where $\log r$ and $\log s$ are main effects and $\log \mathbf{E}$ a matrix of unexplained disturbances, unexplained that is by the main effects. Since there is only one observation per cell, the possible explanation of part of $\mathbf{E}$ by interaction effects cannot be examined.

In projection, Johansen argues that disturbances in year θ should not be projected and proposes that ${}_t\mathbf{A}$ should be estimated by

$$ {}_t\mathbf{A}^* = \hat{\mathbf{r}}^{t/\theta}\;{}_0\mathbf{A}\;\hat{\mathbf{s}}^{t/\theta} \qquad (1.5.5).$$

Implicitly he assumes that

(1) ${}_0\mathbf{A}$ is free from disturbance (or errors in measurement);
(2) $\mathbf{A}$ is subject to random disturbances (e_{ij});
(3) Any disturbances in the matrix to be estimated (${}_t\mathbf{A}$) are independent of those in year θ.

Again the roles of ${}_0\mathbf{A}$ and ${}_\theta\mathbf{A}$ could be reversed;[17] while a compromise estimate could be obtained by taking the geometric mean, it would seem preferable to recognise explicitly that both ${}_0\mathbf{A}$ and ${}_\theta\mathbf{A}$ are subject to disturbance.

Thus

$$ {}_0 a_{ij} = a_{ij}\,{}_0 e_{ij} \qquad (1.5.6)$$

$$ {}_\theta a_{ij} = r_i\, a_{ij}\, s_j\,{}_\theta e_{ij} \qquad (1.5.7)$$

or, more generally,

$$ {}_\phi a_{ij} = {}_\phi r_i\, a_{ij}\,{}_\phi s_j\,{}_\phi e_{ij} \qquad (\phi = 0, \theta) \qquad (1.5.8)$$

where ${}_0\mathbf{r} = {}_0\mathbf{s} = \mathbf{i}$. Logarithms may be taken and $\mathbf{r}$, $\mathbf{s}$ and $\mathbf{A}$ estimated by least squares.

An advantage of this formulation is that it may readily be generalised to take account of complete matrices for a series of years, such as are frequently available in the international trade field. The multipliers are expressed as explicit functions of time; hence if, for example, ${}_\phi\mathbf{r} = \mathbf{r}^\phi$, ${}_\phi\mathbf{s} = \mathbf{s}^\phi$, then

$$ {}_\phi a_{ij} = r_i^\phi\, a_{ij}\, s_j^\phi\,{}_\phi e_{ij} \qquad (1.5.9)$$

While the two-year case is formally equivalent to an analysis of variance (divide equation (1.5.7) by equation (1.5.6) and take logarithms) the many-year case differs in its treatment of time. Equation (1.5.9) may be written

$$\log {}_{\phi}a_{ij} = \phi \log r_i + \phi \log s_j + \log a_{ij} + \log {}_{\phi}e_{ij} \qquad (1.5.10)$$

The corresponding analysis of variance would be of the form

$$\log {}_{\phi}a_{ij} = \log r_i + \log s_j + \underline{\log t_{\phi}} + \log a_{ij} + \underline{\log y_{j\phi}} + \underline{\log z_{\phi i}} + \log {}_{\phi}e_{ij} \qquad (1.5.11)$$

In an analysis of variance, the time dimension is treated in exactly the same way as the other two dimensions, and appears in the three sets of dummy variables underlined in equation (1.5.11); effects *specific* to particular times are distinguished; hence the results could not be applied to projections (different values of ϕ) without further analysis. By contrast equation (1.5.10) suppresses the time dummy variables, but shows the i and j effects changing steadily with time.

Many economic time series will be subject to trends plus fluctuations associated with the trade cycle. Equation (1.5.9) analyses the trends which is the main requirement for medium-term projections. The short-term fluctuations will be reflected in the time profile of the error terms. These may likewise be analysed using a biproportional approach. One method would be to subject each year's error matrix ${}_{\phi}\mathbf{E}$ to an analysis of variance, thus deriving time series of non-trend row and column effects. These effects could be estimated simultaneously with the trend multipliers by modifying equation (1.5.10)

$$\log {}_{\phi}a_{ij} = \phi \log r_i + \phi \log s_j + {}_{\phi}a_i + {}_{\phi}b_j + \log a_{ij} + {}_{\phi}e_{ij} \qquad (1.5.12)$$

Equation (1.5.6) is one of a wider class of relationships explaining the movement of **X** through time in terms of row and column effects, as may be seen by its decomposition. Omitting the error term:

$${}_{\phi}a_{ij} = a_{ij}\,{}_{\phi}m_{ij} \qquad (1.5.13)$$

$${}_{\phi}m_{ij} = {}_{\phi}r_i\,{}_{\phi}s_j \qquad (1.5.14)$$

$${}_{\phi}r_i = r_i^{\phi}, {}_{\phi}s_j = s_j^{\phi}. \qquad (1.5.15)$$

This particular combination suffers from the upward bias characteristic of exponential trends, and the substitution of linear and additive relationships for exponential and multiplicative ones may be advantageous. Empirical work indicates that projections are highly sensitive to the form of trend chosen. If data is available for a series of years, it may be possible to choose between alternative hypotheses on the grounds of goodness of fit and inter-correlation between residuals especially over time.

The final assumption underlying Johansen's proposed methodology was that any disturbances in the projection year are independent of those

in the base year. The equivalent assumption when a series of matrices are available is that disturbances in particular elements are free from serial correlation. This assumption is by no means universally valid. For example the change in inputs of coal into electricity generation cannot be analysed solely in terms of coal input effects and electricity industry effects. The divergence between the actual input and its RAS-estimate will in such a case increase with time and any projection will be downward biased. Further biases will be induced in other elements, especially in the coal row and the electricity column.

Kouevi [27], in his analysis of intra-European trade flows, assumes that ${}_{\theta}x_{ij}$ is related to ${}_{0}x_{ij}$ by a row effect, a column effect and an element effect. Thus

$${}_{\theta}x_{ij} = {}_{0\theta}r_i\,{}_{0}x_{ij}\,{}_{0\theta}s_j\,{}_{0\theta}c_{ij} \qquad (1.5.16)$$

The row and column effects are estimated according to the Stone–Leicester approach equation (1.5.1) to satisfy the row and column totals, and the matrix of element multipliers (**C**) calculated from equation (1.5.16). Kouevi [27] considers a time series of ten matrices, and obtains time series in ${}_{0\theta}c_{ij}$, which exhibit trends and strong cyclical fluctuations, indicating systematic departures from biproportionality.

If only two matrices are available, it will be impossible to establish whether divergences from biproportionality represent short-term disturbances or trends which should be projected. Further information, relating either to individual elements or to row and column sums, is vital, even if not of comparable accuracy.

1.6 IS IT NECESSARY TO PROJECT THE WHOLE MATRIX?

Frequently the projection of matrices is part of a wider exercise. In particular, input–output tables are often used to calculate the vector of intermediate demands and hence of gross outputs. This has led some to question whether laborious projection of the complete matrix is necessary. Tilanus' experiments [42] appear to suggest that much simpler methods work almost equally well. Tilanus works with Dutch input–output matrices for thirteen consecutive years, and compares predictions of intermediate demands derived from a number of different methods with each other and with the outcome. Of his many comparisons, two are of particular relevance.

The first is a comparison between a kind of RAS method, and his own 'Statistical Correction Method' (SCM) which takes no *explicit* account of changes within the matrix through time. His RAS method consists of updating a matrix relating to year 0 to year θ and applying the updated matrix without further adjustment to derive intermediate demand (**u**) in projection year, t:

$${}_{t}\mathbf{u}^* = (\mathbf{I} - {}_{\theta}\mathbf{A}^*)^{-1}\,{}_{\theta}\mathbf{A}^*\,{}_{t}\mathbf{f} \qquad (1.6.1)$$

where ${}_t\mathbf{f}$ is the vector of final demands in year t. The SCM method involves first applying ${}_0\mathbf{A}$ in years θ and t:

$$\mathbf{u}^x = (\mathbf{I} - {}_0\mathbf{A})^{-1}\, {}_0\mathbf{A}\, {}_\phi\mathbf{f} \qquad (\phi = \theta, t) \tag{1.6.2}$$

and then applying corrections in year t proportional to the errors observed in year θ:

$$ {}_t\mathbf{u}^* = {}_\theta\hat{\mathbf{u}}({}_\theta\hat{\mathbf{u}}^x)^{-1}\, {}_t\mathbf{u}^x \tag{1.6.3}$$

Tilanus' version of RAS was found to perform no better than SCM and hence hardly seems justified in view of the extra labour involved.

No information was provided on the accuracy of the RAS update, and there are several reasons for thinking it may not have been high.[18] However Tilanus also applied the observed **A** matrix in place of the RAS update in equation (1.6.1) and still found no improvement over SCM.

It will be noticed that, while all his three projections make allowance, explicit or implicit, for coefficient changes between year 0 and year θ, none allow for further changes between year θ and projection year t, and no attempt is made to modify RAS or SCM to take such further changes into account. However Tilanus does investigate the effect of fitting linear trends to individual coefficients and extrapolating; the results are not encouraging. For example, projections of intermediate demands for the four years 1958–1961, using an extrapolated matrix based on annual data for the period 1949–57, compares unfavourably with alternative projections using the 1957 matrix without adjustment. Similarly, Barker (Chapter 2 below) finds U.K. input–output projections for 1968 based on a 1963 coefficient matrix to be more accurate than alternative projections derived from linear extrapolation of coefficients observed in 1954 and 1963. Neither author attempts to explain this apparently surprising result.

Suppose a coefficient a_{ij} to be given by random normal fluctuations about a linear trend:

$$a_{ij} = \alpha\phi + e \tag{1.6.4}$$

where e is $N(0, \sigma^2)$. Suppose observations to be available at $\phi = 0, 1$. Three simple projection methods are compared in Table 1.6, opposite.

Method (*3*) involves a lower mean square error than method (*1*) if the stochastic element dominates the time trend, specifically if $2\sigma^2 > \alpha^2$. This provides a possible basis for Barker's results. However in the presence of a time trend, method (*3*), even if superior in terms of mean square error, will provide a biased estimate of ${}_2a$.[19]

It will be appreciated that the high variance associated with method (*1*) results from the danger of interpreting random fluctuations as a time trend. In Tilanus' case, with a series of observations, this danger is somewhat reduced. Moreover it is possible from the data to estimate σ^2 and it is unfortunate that Tilanus provides no information of this kind. More

Table 1.6 Simple Projection Methods

Projection method (see Note below)	$E({}_2a^*)$	*var* $({}_2a^*)$	*Mean Square Error*
(*1*) linear trend ${}_2a = 2\,{}_1a - {}_0a$	2α	$3\sigma^2$	$3\sigma^2$
(*2*) mean ${}_2a = \frac{1}{2}({}_0a + {}_1a)$	$\alpha/2$	σ^2	$\sigma^2 + \frac{9}{4}\alpha^2$
(*3*) no change ${}_2a = {}_1a$	α	σ^2	$\sigma^2 + \alpha^2$

Note: The *ij* suffixes are omitted throughout.

important, the possibility that some trends in coefficients are non-linear could be explored. Figure 1.1 (on page 20) illustrates how fitting a linear trend may yield *predictably* poor, biased projections. Crude as 'no change' projections are, one can quite easily do worse![20]

Tilanus' reliance on mechanical methods is quite deliberate: "In our opinion," he writes "it is inadmissible to take extraneous information into account because our input–output predictions are all made *ex post*. We cannot perform input–output experiments in an objective way, if we do not have an objective forecasting mechanism. . . . In case of real life forecasting, however, such extraneous information as is available can always be integrated within one of the pure input–output models. In fact no input–output research workers will advocate a strict, mechanical, application of input–output forecasting models." ([42] p. 53). But what is the purpose of testing the relative merits of these 'objective' methods, if no-one is going to use them? For the tests to have any relevance it must be presumed that the rankings obtained are unaffected when the methods are somehow modified to incorporate additional information. This seems unlikely to be the case. Much information, aiding the interpretation of past trends and the prediction of further developments, is likely to be of a detailed kind. This will be difficult to incorporate unless detailed projections are made.

A further consideration is that planners may well want to do more than make a single projection; they may wish to explore the effects of alternative policies and alternative developments. Such explorations cannot be carried out satisfactorily without projecting the whole input–output matrix.[21] In certain circumstances, a method such as SCM may be an adequate short cut. But it is only by projecting the complete matrix that full account can be taken of valuable additional information.

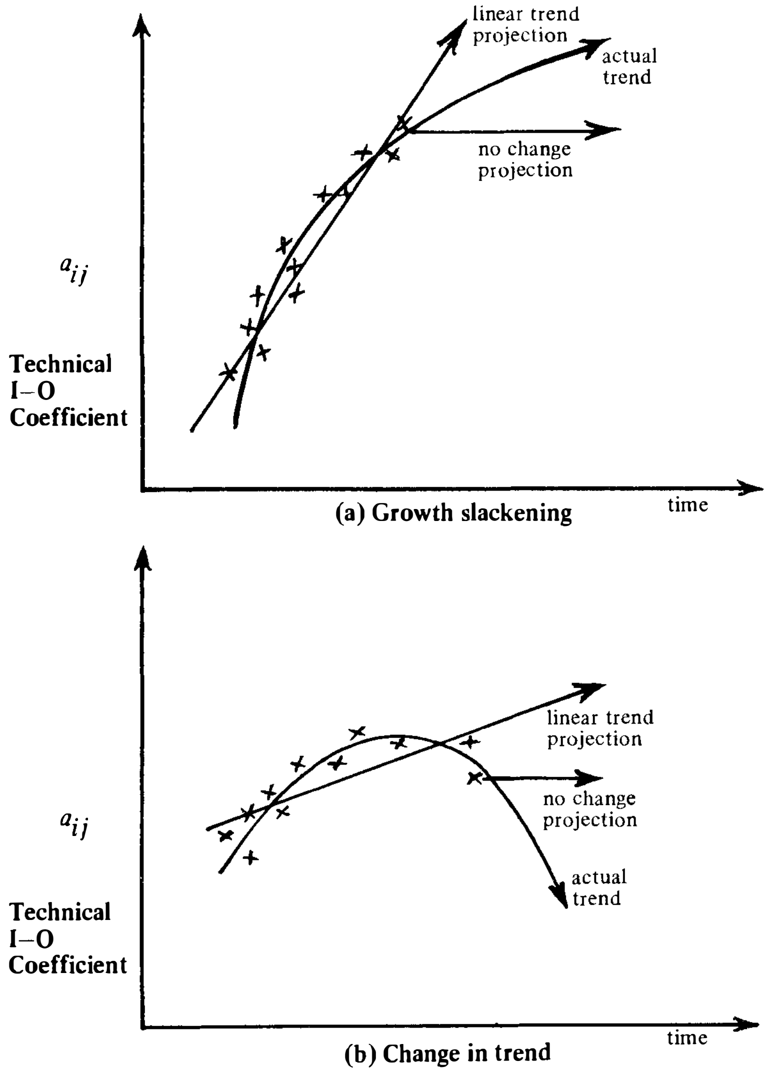

Figure 1.6 Examples of Trend Projections[1]

(a) Growth slackening (b) Change in trend

Note: 1. The diagrams show the effect of making crude (linear trend or no change) projections when the actual trend is curvilinear. In case (b) the no change projection is less inaccurate than the linear trend projection. In both cases (a) and (b) inspection of the time series of residuals from the linear trend would be revealing.

1.7 CONCLUSIONS

The principal conclusions are listed briefly below:

A. *Minimum information problems:*

1. Tests (Bacharach [5], Omar [33], Schneider [35]) have shown various methods of adjusting and updating matrices to produce broadly similar results, no method being clearly superior. However, the linear and quadratic programming methods (Matuszewski *et. al.* [32], Friedlander [17]) can, and in practice do, yield negative elements (even though the initial data is non-negative) unless additional constraints are explicitly introduced. This greatly complicates the solution. RAS is computationally considerably simpler than any other method preserving signs.
2. RAS projection is equivalent to exponential projection of individual elements and hence involves unacceptable upward bias. Bacharach's modification [5] is not wholly satisfactory and an alternative scheme was suggested.
3. Empirical tests show RAS updates to be reliable only under favourable circumstances. Projections using RAS updates are extremely unreliable and have been found to perform no better than projections based on the original matrix plus a statistical correction (Tilanus [42]). It must not be concluded, however, that other methods of *solving the same problem* would do markedly better. Failures of RAS must be attributed to the slender informational base.

B. *Incorporating additional information*

4. RAS updates are substantially improved by exogenous estimation of certain elements (Paelinck and Waelbroeck [34], Bacharach [5]). It is often possible to determine *ex ante* (or from previous experience) which elements are likely to move differently from other elements in their row and column and thus merit special attention. Allen [1] has demonstrated that reliable extimation of total intermediate outputs (**Xi**) depends on the accuracy of a relatively small number of major coefficients; accordingly attention should be directed to obtaining accurate estimates of these; see Chapter 3.
5. Information on individual coefficients is generally abundant, but much of it is not very reliable. Moreover, contrary to assumption, the control estimates (**u** and **v**) are frequently unreliable: this was, for example, a major source of error in the Cambridge updating exercise [11]. A scheme was developed by Lecomber [28] for incorporating information of varying reliability into input–output models and this approach is tested in Chapter 3 below.

6. Projections may be substantially improved by taking into account row and column balances for a series of years (Omar [33]).
7. If complete matrices for a series of years are available, RAS is an inefficient method of projection. Alternative methods, involving an explicit stochastic model incorporating row and column effects are proposed.
8. An explanation is offered of Tilanus' [42] finding that, in predicting intermediate outputs, a recent historical matrix performs better than one obtained by linear extrapolation. It is suggested that a more flexible approach to projecting the matrix, using different forms of trend and incorporating additional information, might have yielded better results.
9. Tilanus' [42] observation that intermediate output projections based on an RAS-updated matrix are no more accurate than projections based on his 'statistical correction method' (using the base year matrix) indicates the limitations of minimum-information RAS; however it is only if the whole matrix is updated and projected that additional information can readily be incorporated. It has already been noted that this leads to a substantial improvement over minimum-information RAS and hence over SCM.

Central to all these suggestions is the idea that, as far as possible all available information should be used, even that which is not fully reliable or appropriate and that, where possible, a subjective assessment of the reliability of information should be taken into account in a more or less formal way in the estimation procedure. If *all* information has been taken into account, then there will be no occasion for rejecting the results as implausible, for such a judgement must depend on further information so far ignored. This is a counsel of perfection certainly: there will be costs in utilising more information—both in the collation of the data and in the estimation routines (for example the modified RAS procedures here suggested are more complicated than their minimum-information counterparts). Other information may not occur to the research worker until presented with conflicting results. So it will still be necessary to judge the sophisticated (estimation routines) by the simple (plausibility tests). But the role of posterior *ad hoc* judgement will be greatly reduced if, as far as possible, information is incorporated and judgement exercised at an earlier stage.

FOOTNOTES

[1] Some confusion surrounds the paper by Deming and Stephan [14], since the verbal and algebraic accounts conflict. The RAS method is described in the text, while the algebra

relates to the variant referred to later in this paper as the Friedlander method. It appears that it was the latter method that was actually used in the calculations (Stephan [36]).

[2] This work also includes a rigorous investigation of the mathematical properties of RAS and related methods.

[3] Throughout, left subscripts are used to label matrices and vectors, right subscripts to identify particular elements within matrices or vectors. Thus ${}_{0}\mathbf{X}$. (Exception: Table 1.2).

[4] Omar used general quadratic programming routines and found that, for large matrices, the problem was beyond the scope of most computers. Matuszewski, Pitts and Sawyer [32] in their linear programming adjustment, successfully adapted a procedure whereby bounds are not imposed until violated. But the textual comment stands.

[5] Henry [24] emphasises that some matrices may legitimately contain negative elements (for example bi-products are sometimes treated as negative inputs in input-output tables). If so, the RAS procedure does not necessarily converge, and if it does the minimand is clearly not that given in Table 1.1 (logarithms of negative numbers being undefined). The Friedlander minimand given by Henry (his equation 3) also breaks down, but a simple modification—replacing ${}_{\bullet}x_{ij}$ (ξ_{ij} in his notation) in the denominator by $|{}_{\bullet}x_{ij}|$—gives satisfactory answers. Some insight on the problem of adjusting matrices including negative elements is given by the vector analogue. For example consider adjusting $(-1, 2)$ to sum to 4. Pro-rata adjustment (analogous to RAS) gives $(-4, 8)$, while the modified Friedlander adjustment gives $(1, 3)$. It is the perverse movement of the negative elements that can lead to nonconvergence of the RAS process in the matrix case. Two more general methods of adjusting negative matrices are given on pp. 5–6.

[6] $\mathbf{X}^*$ is unaffected by multiplying $\mathbf{E}$ by a scalar.

[7] An alternative and seemingly simpler analogue, $\mathbf{X}^* = {}_{\bullet}\mathbf{X} + \hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}$, must be rejected because the control totals $(\mathbf{u} - {}_{0}\mathbf{X}\mathbf{i})$ and $(\mathbf{v} - {}_{\bullet}\mathbf{X}'\mathbf{i})$ are liable to include negative elements.

[8] Just as the 'index number problem' is no argument for preferring an unweighted index.

[9] But see Barker's critical appraisal of the Paelinck and Waelbroeck tests (Chapter 4 below).

[10] These surmises are confirmed by a comparison of the official United Kingdom input–output tables for 1954 and 1963 (C.S.O. [45][46]).

[11] Both excluded diagonal elements, often a source of difficulty, as the aircraft example illustrates. Exclusion of diagonal elements, straightforward enough in *ex post* tests, is frequently impracticable in genuine updating exercises, as in the Cambridge case.

[12] Except in the trivial case where each variable grows at the same rate.

[13] This matrix originally derived for 1966 was used without further modification (apart from the elimination of some major implausibilities) for projections for 1970.

[14] As obtained from the 1954 matrix. Elements estimated exogenously are excluded from these tests.

[15] Or that the matrix for year θ was obtained by the more elaborate updating methods of section 1.3 so that ${}_{0}\mathbf{A}$ and ${}_{\theta}\mathbf{A}$ are not connected by any simple functional relationship.

[16] This minimand gives excessive weight to small elements. Accordingly Johansen prefers to minimise Σe_{ij}^2; a computationally simpler alternative is to minimise $\Sigma c_{ij}(\log e_{ij})^2$, the weights being chosen in relation to individual elements (Lecomber [29]).

[17] It may be shown (Lecomber [29]) that, in contrast to Stone and Leicester's procedure, ${}_{0\theta}\hat{\mathbf{r}} = {}_{\theta 0}\hat{\mathbf{r}}^{-1}$, ${}_{\bullet\theta}\hat{\mathbf{s}} = {}_{\theta 0}\hat{\mathbf{s}}^{-1}$. There is nevertheless still an arbitrary choice, since the multipliers may be applied to either ${}_{0}\mathbf{A}$ or ${}_{\theta}\mathbf{A}$.

[18] Firstly, the updating procedure is entirely mechanical, no attempt being made to allow even for elements likely to misbehave, such as primary fuel inputs. Secondly, Tilanus' A

matrix excludes imported inputs, with a consequent reduction in its stability. This latter feature almost certainly accounts for Tilanus' finding that the matrix shows greater inter-temporal stability at current prices than constant prices. For further evidence on this issue see Barker in Chapter 2 below.

[19] If there is thought to be no time-trend then method (2) is the obvious one to use. It is difficult to find a justification for (3).

[20] Tilanus also experiments with linear trends fitted to the last three points, a method which should perform moderately well for the series plotted in the diagrams. However many coefficients exhibit strong cyclical movements mainly associated with cyclical variations in the product mix of user industries. This would be sufficient to account for the poor performance of this method in Tilanus' experiments. This source of error could have been reduced by explicit analysis of cyclical effects.

[21] *One* could of course be constructed by RAS or other *updating* methods to fit the marginal constraints obtained by SCM. Such filling-in procedures are indeed employed by the Batelle Institute in Geneva [16] but this additional step seriously reduces the computational advantages of SCM and the accuracy of a matrix obtained in this way has yet to be established.

APPENDIX TO CHAPTER 1

EXPLAINING THE TRENDS

The projection procedures discussed in the text (pp. 9–17) are based on an analysis of the movement of matrices through time. Time is the only explanatory variable. It is of course always more satisfactory to seek the underlying causes of intertemporal movements. Of the large number of conceivable models, only those based on the row-and-column hypothesis (namely that elements of the matrix are influenced by variables associated with particular rows and columns) will be considered: for example, international trade flows may be explained in terms of the gross domestic product of the exporter and that of the importer.

Explanatory variables may be brought into the analysis in one of two ways. The first is to relate these to **r** and **s** multipliers obtained in one of the methods described above. Stone and Leicester [39], in their analysis of manpower matrices, find a close correlation between the column multipliers representing, broadly, changes in manpower input coefficients[1] by industry, and the growth in industry outputs. Accordingly, in making projections, the multipliers ${}_{\theta t}\mathbf{s}$ are found as functions of growth in industry outputs rather than the time span, $t - \theta$. The author has used the same technique in the analysis of U.K. exports (1961–1966) cross-classified by commodity (rows) and destination (columns).

Algebraically

$$r_i = \alpha\, {}_{\theta}y_i^{\beta}, \qquad s_j = \gamma\, {}_{\theta}z_j^{\delta} \qquad \text{(A1.1)}$$

${}_{\theta}\mathbf{y}$ and ${}_{\theta}\mathbf{z}$ are indices (year $\theta = 1$) of world trade by commodity and destination.

[1] It is of course common practice to remove the most obvious row and column effects before estimating the multipliers. In input–output work, for example, it is usual to assume elements of intermediate demand proportional to industry outputs and to find multipliers connecting matrices of input–output *coefficients*.

These expressions may be incorporated into equation (1.5.8).

$$ {}_{\phi}x_{ij} = (\alpha\gamma)\, {}_{\phi}y_i^{\beta}\, x_{ij}\, {}_{\phi}z_j^{\delta}\, {}_{\phi}e_{ij} \qquad (\phi = 0, \theta) \qquad \text{(A1.2)}$$

and $(\alpha\gamma)$, β, δ and x_{ij} may be found by applying least squares to the log transform. It should be noticed that, in contrast to the network models of Linnemann [31] and others, no attempt is made to explain the *levels* of elements (represented by x_{ij}), only their trends. By dropping x_{ij} from equation (A1.2) both levels and changes are explained in terms of the row and column parameters. The approach may clearly be generalised to the analysis of a series of matrices. The greater number of degrees of freedom then permit elaborations such as:

$$ {}_{\phi}x_{ij} = a_i y_i^{b_i}\, x_{ij}\, c_j\, z_j^{d_j}\, {}_{\phi}e_{ij} \qquad (\phi = 0 \ldots n) \qquad \text{(A1.3)}$$

We may compare the equations (A1.2) and (A1.3) with

$$ {}_{\phi}x_{ij} = x_{ij}\, {}_{\phi}w_{ij}^{b_{ij}}\, {}_{\phi}e_{ij} \qquad (\phi = 0 \ldots n) \qquad \text{(A1.4)}$$

where w_{ij} is some variable specific to the ijth cell (e.g. world trade cross-classified by commodity and destination). Equation (A1.4) breaks down into sub-sets of relationships relating to particular cells, each with its own explanatory variables and its own parameters; each therefore becomes a separate regression problem.

The distinctive feature of equations (A1.2) and (A1.3) is that common variables and/or parameters are used across whole rows and down whole columns. Fewer variables are involved, simplifying data collection and derivation of independent variables in a projection exercise. Fewer parameters are involved, increasing the degrees of freedom; this will increase the accuracy of the estimates provided the underlying hypotheses (e.g. about uniformity of row and column effects) are a sufficiently close approximation to reality; this may be tested by examining the pattern of residuals. For large matrices, the computational demands are however rather heavy.

CHAPTER 2

Some Experiments in Projecting Intermediate Demand

T. S. BARKER

University of Cambridge, Department of Applied Economics

2.1 INTRODUCTION

In a disaggregated medium term model of an economy, a crucial link between estimates of demand for commodities and output of industries is the projected input–output matrix. Various methods for projecting this matrix from one or more observations have been proposed and investigated, the most comprehensive studies being those by Ghosh on U.K. data [19] and Tilanus on Dutch data [42]. A usual problem is the absence of comparable observations for a time series of industrial demands: in the post-war period, full U.K. census of production data, including information in input and output, is only available for the years 1948, 1954, 1963 and 1968 and the 1948 results are distorted by the return to a peacetime economy. (For 1970 the C.S.O. has recently published a table 'updated' from the 1968 census-based table). In addition the results are given in current prices and with changing classifications as the Standard Industrial Classification is revised. In consequence model builders are usually faced with a situation of having to project an input–output matrix four to five years ahead with full information from a census of production up to nine years old, but with more recent, partial information on row and column sums of the matrix and various cells within it.

This paper investigates various methods of projecting intermediate demand for the U.K. in 1963, using the 1954 census results and partial information for 1960. The analysis is at a 45-commodity level of disaggregation in both 1963 prices and current prices. Projections are made of the matrix of industrial demands but the main basis for comparison is the vector of row sums of this matrix, the sales of each commodity for intermediate use.

2.2 INPUT–OUTPUT TABLES FOR 1954, 1960 AND 1963

Since the publication of the 1963 Census of Production it has been possible to construct both in current and constant prices input–output tables on a comparable basis for 1954 and 1963. At Cambridge this was

done by first estimating flows for commodities and industries within the Growth Project's Social Accounting Matrix (SAM) framework. The make matrix **M**, showing the production of commodities by industries, and the absorption matrix, **X** showing the current consumption of commodities by industries were combined through the 'industry technology' assumption[1], whereby both principal and non-principal products of any industry are assumed to be produced by the same input structure. This gives

$$\mathbf{A} = \mathbf{X}\hat{\mathbf{g}}^{-1}\mathbf{M}\hat{\mathbf{q}}^{-1} \qquad (2.2.1)$$

where **g** and **q** are vectors of industry and commodity outputs respectively. The method is discussed by Armstrong in Chapter 5 below.

An input–output table for 1960 may be constructed by the same process, with the exception of the inter-industry flows for which only partial information is available. The matrices were constrained to add to estimated row and column totals by the RAS method. However, the results of the 1963 census were used in estimating some of the 1960 inter-industry flows so that, in order to project 1963 without the benefit of hindsight, a 1960 table using only 1954 and 1960 information has been adopted in most of the forecasts given in this chapter.

2.3 THE EXPLANATION OF INTERMEDIATE DEMAND

Intermediate demand for products can be derived from cost–minimising behaviour with a given technique of production. If the technique is such that no substitution between inputs is possible in the production of an output then, provided no inputs are wasted, their demand is in fixed proportion to output. This is the basis of the Leontief input–output system, which can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{q} \qquad (2.3.1)$$

where **x** is the vector of intermediate demands, **q** the vector of outputs and **A** a matrix of fixed coefficients.

Outputs, intermediate demands and net final demands are related by the identity

$$\mathbf{q} = \mathbf{x} + \mathbf{f} \qquad (2.3.2)$$

where **f** is the vector of net final demands (i.e. net of imports). If we assume final demands to be given, then **q** and **x** can be derived as

$$\begin{aligned}\mathbf{q} &= \mathbf{A}\mathbf{q} + \mathbf{f}\\ &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{f} \qquad (2.3.3)\end{aligned}$$

and

$$\begin{aligned} \mathbf{x} &= \mathbf{Aq} \\ &= \mathbf{A}(\mathbf{I} - \mathbf{A})^{-1}\mathbf{f} \\ &= [(\mathbf{I} - \mathbf{A})^{-1} - \mathbf{I}]\mathbf{f} \end{aligned} \tag{2.3.4}$$

However, there are several reasons for expecting change in the **A** matrix. When several techniques are available, choice of technique, and therefore input structure, will partly depend on the cost of the inputs and substitution becomes possible as relative prices of inputs change. If there has been a persistent shift in the relative prices of substitutable inputs then, as equipment embodying the uneconomical technique is replaced, the input structure will change. This type of coefficient change was investigated by Wigley [57] for fuel inputs in the U.K. between 1948 and 1964 when the price of oil fell relative to the price of coal.

There are other reasons for changing coefficients. The introduction of new techniques, independent of changing relative prices, may also change the input structure. Furthermore, if there are economies of scale in the use of some inputs so that some marginal input coefficients are smaller than the corresponding average ones, average coefficients may fall as output levels grow. Finally, coefficient values may change if output is not homogeneous and the mix of products in that output alters. These last two factors making for change, economies of scale and product mix, might very well lead to the cyclical discrepancies between observed intermediate demand totals and totals from final demands and an input–output matrix which were found by Arrow and Hoffenberg [4].

These factors influencing the change in input–output coefficients through time have long been recognised and there is considerable literature on the topic. The approaches which have been adopted to account for the change can be grouped into four categories:

(i) Trends in coefficients;
(ii) Updating of the base matrix;
(iii) Restricted price substitution;
(iv) Non-homogeneous production functions.

2.4 TRENDS IN COEFFICIENTS

This is the simplest approach, relying on the factors influencing change to operate systematically through time. At least two input–output tables are required to estimate the trend in each coefficient. We have

$$\mathbf{x} = \mathbf{Aq} = (\mathbf{A}_0 + \mathbf{A}_1\tau)\mathbf{q} \tag{2.4.1}$$

where $\mathbf{A}_0$ and $\mathbf{A}_1$ are matrices of parameters and τ is the year. With only two base matrices $\mathbf{A}_0$ and $\mathbf{A}_1$ are exactly determined, but the trends which are estimated must be considered very uncertain. Tilanus [42] found that with ten input–output tables, measured from current price data, linear trends gave worse results than using coefficients of the most recent table. If we take the input–output tables estimated for 1954 and 1960, but from *constant price* data, and extrapolate to 1963 his findings are confirmed. Table 2.1 shows the observed intermediate demand for commodities in 1963 and the deviations (predicted less observed demands) expressed as percentage of the observed demands for three projections, each assuming as given the vector of net final demand in 1963. The first and second assume that coefficients remain at their 1954 and 1960 levels respectively, and the third extrapolates the linear trend between coefficients forward to 1963. Any coefficient which becomes negative in the extrapolation is assumed to be zero in 1963. The last two rows of the table show the square roots of the mean squared error for elements of the vector of intermediate demands and for those of the matrix defined as

$$\mathbf{X}^* = \mathbf{A}\hat{\mathbf{q}} \tag{2.4.2}$$

where $\hat{\mathbf{q}}$ denotes a diagonal matrix with the elements of vector $\mathbf{q}$ along the leading diagonal. The row sums of $\mathbf{X}^*$ are equal to those of $\mathbf{X}$ and hence

$$\mathbf{X}^*\mathbf{i} = \mathbf{A}\hat{\mathbf{q}}\mathbf{i} \tag{2.4.3}$$

Combining this with equation (2.2.1) gives:

$$\begin{aligned} \mathbf{X}^*\mathbf{i} &= \mathbf{X}\hat{\mathbf{g}}^{-1}\mathbf{M}\hat{\mathbf{q}}^{-1}\hat{\mathbf{q}}\mathbf{i} \\ &= \mathbf{X}\hat{\mathbf{g}}^{-1}\mathbf{Mi} \\ &= \mathbf{X}\hat{\mathbf{g}}^{-1}\hat{\mathbf{g}}\mathbf{i} \\ &= \mathbf{Xi} \end{aligned} \tag{2.4.4}$$

where $\mathbf{i}$ is the summation vector $(1, 1, \ldots, 1)$.

The 1954 and the extrapolated input–output matrices both give worse predictions than the one estimated for 1960. As a comparison with these projections a 'naive' method has been used to give a fourth set of results. These are shown in the last column of Table 2.1, which is calculated as

$$\mathbf{x} = \hat{\mathbf{b}}_{60}\mathbf{f} \tag{2.4.5}$$

where $\mathbf{b}_{60} = \hat{\mathbf{f}}_{60}^{-1}\mathbf{x}_{60}$, that is the levels of intermediate demand in 1960 as proportions of net final demand in that year. On the root mean squared error criterion this gives similar results to the use of the 1954 matrix but worse results than using the 1960 matrix or extrapolating. A slightly less naive forecast, taking as vector $\mathbf{b}$ in equation (2.4.5) the extrapolation of $\mathbf{b}_{54}$ and $\mathbf{b}_{60}$, gave even worse results.

One characteristic of the naive forecasts is the excellent prediction of intermediate demand for the last seven commodities which are included

Table 2.1 The Projection of Intermediate Demand from Input–Output Tables for 1954 and 1960, U.K. 1963

Social Accounting Matrix Commodity[1]	Inter-mediate Demand 1963 £mn. (1)	Percentage errors for the projection using			
		$\mathbf{A}_{54}$ (2)	$\mathbf{A}_{60}$ (3)	$\mathbf{A}_0 + \mathbf{A}_1\tau$ (4)	The 'naive' method (5)
1.1 Raw meat	157	9	−19	−31	−17
1.2 Cereals	395	−19	−21	−22	−22
1.3 Agric. prod. n.e.s.[2]	418	27	5	− 4	7
2 Coal	682	48	15	− 3	12
3 Mining prod. n.e.s.[2]	140	8	− 1	− 3	−32
4 Cereal products	553	8	—	− 5	1
5.1 Meat and fish prod.	10	−12	−16	−16	2
5.2 Processed food n.e.s.[2]	311	−15	−15	−14	−19
6 Drink	112	−10	−12	−11	−13
7 Tobacco manufactures	4	8	7	6	8
8 Coke	181	49	41	31	188
9 Refined mineral oil	462	−33	−19	−10	−38
10 Chemicals n.e.s.[2]	1556	− 9	− 4	—	− 1
11 Iron and steel	1660	12	16	14	−27
12 Non-ferrous metals	838	—	9	14	15
13 Engineering prod.	1906	−18	− 4	5	− 1
14 Ships etc.	145	18	23	24	13
15 Motor vehicles	521	− 1	14	22	21
16 Aircraft	157	− 2	66	109	69
17 Vehicles n.e.s.[2]	79	93	24	−14	−11
18 Metal goods n.e.s.[2]	1059	17	12	12	3
19 Textile fibres	409	− 8	2	− 5	− 5
20 Textiles	1198	24	12	− 2	13
21 Leather, clothing, ftw.	149	36	13	− 1	12
22 Building materials	490	1	− 5	− 9	−42
23 Pottery and glass	176	−14	− 9	− 5	−27
24 Timber etc.	614	12	8	2	−24
25 Paper and board	423	−11	4	19	23
26 Paper n.e.s.[2]	851	−18	− 2	8	− 7
27 Rubber products	255	− 1	− 2	− 4	−27
28 Manufactures n.e.s.[2]	272	−39	−11	4	−24
29 Construction	830	− 1	17	26	23
30 Gas	165	—	− 1	− 6	7
31 Electricity	583	−36	−15	− 5	− 2
32 Water	54	7	31	44	71
33 Transport and comns.	1459	− 1	− 1	− 1	− 2
34 Distribution	874	4	4	4	−10
35 Services n.e.s.[2]	2165	− 5	3	7	9

Table 2.1—continued

Social Accounting Matrix Commodity[1]	Inter-mediate Demand 1963 £mn. (1)	Percentage errors for the projection using			
		$\mathbf{A}_{54}$ (2)	$\mathbf{A}_{60}$ (3)	$\mathbf{A}_0 + \mathbf{A}_1\tau$ (4)	The 'naive' method (5)
36 Tobacco	105	−10	− 8	− 8	—
37 Crude mineral oil	376	3	−15	− 8	—
38 Iron ore and scrap	78	− 6	23	31	− 3
39 Non-ferrous ores	77	25	25	21	—
40 Woodpulp	115	− 5	9	27	—
41 Butter	4	−71	6	43	—
42 Tea and coffee	10	−24	−47	−57	—
Total	23078	0·2	3·3	4·7	0·4
Mean squared error for vector of intermediate demand, £ million		110	69	73	116
Mean squared error for matrix of intermediate demand, £ million		15·1	9·4	10·3	..

Notes: 1. The definition of Social Accounting Matrix (SAM) commodities in terms of the Standard Industrial Classification, 1958, is to be found in Volume 3 of *A Programme for Growth* [11].
2. n.e.s. equals 'not elsewhere specified'.

to allow special treatment of complementary imports. This is to be expected since domestic production of these commodities is very small so that net final demand is almost exactly identified with intermediate demand disregarding sign. The other projections of this intermediate demand are very poor forecasts since they relate the demand to the output of the sectors which use the inputs. This output in turn is often badly forecast.

2.5 UPDATING OF THE BASE MATRIX

The results in Table 2.1 point clearly to the need for having up-to-date input–output tables in order to improve on a naive forecast. This agrees with the conclusions of Tilanus [42] that in forecasting intermediate demand it is not so much the number of input–output tables that are available that counts as how up-to-date these tables are. One method of constructing more recent tables in years where no census of production has been made is the RAS method developed in the Cambridge Growth

Table 2.2 The Projection of Intermediate Demand by Updating Methods, U.K. 1963

	Social Accounting Matrix Commodity	Intermediate demand 1963 £mn. (1)	Percentage errors for the forecast using: $\hat{\mathbf{r}}_{60}\mathbf{A}_{54}\hat{\mathbf{s}}_{60}$ (2)	$\hat{\mathbf{r}}_{63}\mathbf{A}_{54}\hat{\mathbf{s}}_{63}$ (3)	$\hat{\mathbf{r}}_{63}\mathbf{A}_{60}\hat{\mathbf{s}}_{63}$ (4)
1.1	Raw meat	157	−18	−29	−29
1.2	Cereals	395	−21	−20	−20
1.3	Agric. prod. n.e.s.	418	5	− 2	− 2
2	Coal	682	14	− 2	—
3	Mining prod. n.e.s.	140	− 1	− 3	− 3
4	Cereal products	553	—	—	—
5.1	Meat and fish prod.	10	−16	−17	−16
5.2	Processed food n.e.s.	311	−15	−14	−14
6	Drink	112	−13	− 9	− 8
7	Tobacco manufactures	4	7	5	5
8	Coke	181	41	26	26
9	Refined mineral oil	462	−20	−14	−12
10	Chemicals n.e.s.	1556	− 4	2	1
11	Iron and steel	1660	15	8	9
12	Non-ferrous metals	838	10	13	13
13	Engineering prod.	1906	− 4	3	3
14	Ships etc.	145	24	19	18
15	Motor vehicles	521	15	21	20
16	Aircraft	157	67	110	112
17	Vehicles n.e.s.	79	26	− 5	− 7
18	Metal goods n.e.s.	1059	13	10	9
19	Textile fibres	409	2	− 6	− 6
20	Textiles	1198	12	− 1	− 1
21	Leather, clothing, ftw.	149	13	3	3
22	Building materials	490	− 5	−10	−10
23	Pottery and glass	176	− 9	− 6	− 7
24	Timber etc.	614	7	− 2	− 1
25	Paper and board	423	4	10	10
26	Paper n.e.s.	851	− 2	2	1
27	Rubber products	255	− 2	− 5	− 6
28	Manufactures n.e.s.	272	−11	6	6
29	Construction	830	16	21	23
30	Gas	165	− 1	−10	−11
31	Electricity	583	−15	− 4	− 5
32	Water	54	31	41	42
33	Transport and comns.	1459	− 1	− 3	− 4
34	Distribution	874	4	3	3
35	Services n.e.s.	2165	3	5	5
36	Tobacco	105	− 8	− 9	− 9

Table 2.2—continued

	Social Accounting Matrix Commodity	Intermediate demand 1963 £mn. (1)	Percentage errors for the forecast using: $\hat{\mathbf{r}}_{60}\mathbf{A}_{54}\hat{\mathbf{s}}_{60}$ (2)	$\hat{\mathbf{r}}_{63}\mathbf{A}_{54}\hat{\mathbf{s}}_{63}$ (3)	$\hat{\mathbf{r}}_{63}\mathbf{A}_{60}\hat{\mathbf{s}}_{63}$ (4)
37	Crude mineral oil	376	−16	−10	− 8
38	Iron ore and scrap	78	23	23	24
39	Non-ferrous ores	77	25	21	21
40	Woodpulp	115	9	17	17
41	Butter	4	8	103	97
42	Tea and coffee	10	−47	−55	−55
	Total	23078	3·3	3·2	3·2
	Mean squared error for vector of intermediate demand, £million		68	60	60
	Mean squared error for matrix of intermediate demand, £million		11·3	11·5	9·3

Notes: See Table 2.1.

Project [11] which combined the updating of the 1954 table to 1960 with projections to 1966 using the **r** and **s** multipliers of the updating method. These were calculated as

$$\mathbf{A}_2 = \hat{\mathbf{r}}^{\theta}\mathbf{A}_0\hat{\mathbf{s}}^{\theta} \tag{2.5.1}$$

where $\mathbf{A}_0$ is the base matrix, $\mathbf{A}_2$ the projected matrix, **r** and **s** the multipliers which give the estimated matrix with control totals, $\mathbf{A}_1$, and $\theta > 0$ taking a value depending on the intervals between $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{A}_2$. Several adjustments were made to the projected matrix; in particular, many fuel coefficients were individually estimated for 1966 and the column totals of $\mathbf{A}_{1966}$ were constrained to be equal to those of $\mathbf{A}_{1960}$ (otherwise the assumption of exponential trends imparts an upward bias to the projection).

Tilanus [42] experimented with RAS and a similar correction method and compares projections based on an updated table (without allowing for trends in multipliers) and the table directly estimated from a census of production. His base, updated and projected tables were all estimated from currently priced data. He found that the updated table is just as good or even better as a predictor compared with the observed table.

Table 2.2 sets out the results of similar experiments on U.K. data. It shows the effects on projections of using a mechanically updated 1960 matrix, instead of one with as much partial information incorporated in it as possible, and the effects of introducing trends in mutlipliers.

Column 2 of Table 2.2 shows the percentage deviations for intermediate demand in 1963, using as the input–output matrix

$$\mathbf{A} = \hat{\mathbf{r}}_x \mathbf{X}_{54} \hat{\mathbf{s}}_x \hat{\mathbf{g}}_{60}^{-1} \hat{\mathbf{r}}_m \mathbf{M}_{54} \hat{\mathbf{s}}_m \hat{\mathbf{q}}_{60}^{-1} \tag{2.5.2}$$

where $\mathbf{r}_x$, $\mathbf{s}_x$, $\mathbf{s}_m$ and $\mathbf{r}_m$ are vectors of multipliers which correct the rows and columns of the absorption and make matrices for 1954, $\mathbf{X}_{54}$ and $\mathbf{M}_{54}$, so that they add to the estimated totals for 1960. The $\mathbf{X}$ and $\mathbf{M}$ matrices are adjusted independently so as to make the most of the information available for 1960. If the results in column 2 of Table 2.2 are compared with those in column 3 of the Table 2.1 we find only one substantial difference; the forecasts of the elements of the matrix of intermediate demands are improved when the adjustments were made to the mechanically updated matrix. Otherwise the row sums of this matrix are projected equally well with both tables estimated for 1960.

Columns 3 and 4 show the percentage deviations after allowing for trends in the multipliers, but constraining the column totals of the projected input–output matrix to those estimated for 1960. Column 3 is based on the 1960 matrix estimated as equation (2.5.1) above whilst column 4 incorporates other information on the 1960 coefficients. Again we find that this extra information improves the estimates of the elements of the intermediate demand matrix, but not the elements of the vector of its row totals. We note that the inclusion of trends in the multipliers reduces the root mean squared error by approximately 12 per cent.

2.6 RESTRICTED PRICE SUBSTITUTION

Price sensitivity can be introduced in a very restricted way by measuring and projecting input–output coefficients which are weighted averages of volume and value coefficients. The relationship

$$\mathbf{X} = \mathbf{A}\hat{\mathbf{q}} \tag{2.6.1}$$

is combined with

$$\hat{\mathbf{p}}_x \mathbf{X} = \mathbf{A}^* \hat{\mathbf{p}}_q \hat{\mathbf{q}} \tag{2.6.2}$$

to give

$$\mathbf{X} = \mathbf{A}^{(1-\theta)} \hat{\mathbf{p}}_x^{-\theta} \mathbf{A}^{*\theta} \; \hat{\mathbf{p}}_q^{\theta} \; \hat{\mathbf{q}} \tag{2.6.3}$$

where $\mathbf{p}_x$ and $\mathbf{p}_q$ are vectors of price indices relating to $\mathbf{x}$ and $\mathbf{q}$, $\mathbf{A}^*$ is a matrix of value coefficients, θ is a constant and matrices to the power θ denote that each element is raised to that power. It is assumed that the price index of an input is the same whichever industry or commodity uses it. Taking logarithms and differentiating we obtain

$$\mathrm{d}\log(\mathbf{X}\hat{\mathbf{q}}^{-1}) = -\theta \, \mathrm{d}\log(\mathbf{p}_x \mathbf{i}' \hat{\mathbf{p}}_q^{-1}) \tag{2.6.4}$$

Thus the own price elasticity of any input into any output is equal to a constant, $-\theta$.

This form for the demand function for industrial inputs can be derived

from constant elasticity of substitution (C.E.S.) production functions. The general form of the set of C.E.S. functions is

$$\mathbf{q} = (\mathbf{B}\mathbf{x}^{-\rho})^{-1/\rho} \text{ where } \mathbf{B} > \mathbf{0}, \mathbf{B}\mathbf{i} = \mathbf{i} \tag{2.6.5}$$

Under the assumption of cost minimisation subject to these technical relationships it has been shown by Theil and Tilanus [41] that all the cross-price elasticities of substitution are zero, and the own-price elasticities are equal to a constant $\left(-\dfrac{1}{\rho+1}\right)$ provided the prices are deflated by an index of the prices of all inputs into a particular commodity. Hence we can identify θ in equation (2.6.3) with $\dfrac{1}{\rho+1}$, if the output price level moves with a weighted average of the input prices.

We must note just how restrictive these assumptions are: every input is substitutable for any other input in exactly the same way in each industry. This is only slightly more general than the conventional input–output assumption that there is no possibility of input substitution at all, i.e. that $\theta = 0$, and that $\rho = \infty$.

Alternative assumptions about the value of θ will yield new forecasts of intermediate demands, which (if $\theta > 0$) will depend on movements of relative input prices as well as outputs. Two special cases are $\theta = 0$, the assumption implicit in the forecasts in sections 2.4 and 2.5 above, and $\theta = 1$ which corresponds to the estimation of coefficients from value data alone. Table 2.3 gives four sets of forecasts for different values of θ allowing for trends in the row and column multipliers and constraining the column sums of the projected matrix. Figure 2.1 shows how the total errors change with different values of θ.

It is clear that using the mean squared error criterion the optimal value of θ lies between 0 and -0.1, indicating a rather low degree of substitutability. This is perhaps not surprising since no allowance is made for the fact that some inputs (e.g. fuels) are much more substitutable one for the other than are inputs in general. This low value for θ compares with a value between -0.75 and -0.5 found by Tilanus [42] for predictions of Dutch intermediate demand. He uses a different criterion for optimality[2], which involves a lower penalty for errors with high absolute magnitudes but low relative ones, but this makes no difference to the conclusion that U.K. intermediate demand appears much less substitutable, in the restricted way we are measuring it. One reason for the discrepancy may be the treatment of imports: Tilanus includes all imports in primary inputs whereas we have included them with other commodity flows as intermediate inputs. Since competitive imports and domestic substitutes are likely to have high elasticities of substitution, Tilanus' results may well be reflecting the substitution of imports for domestically produced inputs when these inputs rise in price and vice versa.

Table 2.3 The Projection of Intermediate Demand from Mixed Volume and Value Input–Output Coefficients, U.K. 1963

Social Accounting Matrix Commodity	Inter-mediate Demand 1963 £mn. (1)	Percentage errors for a projection using mixed coefficients			
		$\theta = 0$ (2)	$\theta = -0{\cdot}1$ (3)	$\theta = -0{\cdot}5$ (4)	$\theta = -1$ (5)
1.1 Raw meat	157	−29	−25	−32	−30
1.2 Cereals	395	−20	−25	−25	−28
1.3 Agric. prod. n.e.s.	418	− 2	3	− 3	—
2 Coal	682	− 2	9	—	2
3 Mining prod. n.e.s.	140	− 3	− 4	− 1	1
4 Cereal products	553	—	− 6	− 9	−16
5.1 Meat and fish prod.	10	−17	−14	−39	−55
5.2 Processed food n.e.s.	311	−14	−14	−22	−24
6 Drink	112	− 9	−33	−50	−38
7 Tobacco manufactures	4	5	3	6	8
8 Coke	181	26	29	22	19
9 Refined mineral oil	462	−14	−12	− 8	− 2
10 Chemicals n.e.s.	1556	2	1	5	9
11 Iron and steel	1660	8	11	16	25
12 Non-ferrous metals	838	13	6	12	12
13 Engineering prod.	1906	3	—	2	2
14 Ships etc.	145	19	23	20	20
15 Motor vehicles	521	21	19	29	37
16 Aircraft	157	110	45	56	15
17 Vehicles n.e.s.	79	− 5	6	23	60
18 Metal goods n.e.s.	1059	10	8	13	16
19 Textile fibres	409	− 6	10	5	− 4
20 Textiles	1198	− 1	11	2	6
21 Leather, clothing, ftw.	149	3	17	6	10
22 Building materials	490	−10	−10	− 7	− 4
23 Pottery and glass	176	− 6	− 7	− 6	− 4
24 Timber etc.	614	− 2	2	1	5
25 Paper and board	423	10	1	10	10
26 Paper n.e.s.	851	2	− 3	3	5
27 Rubber products	255	− 5	− 5	− 5	− 3
28 Manufactures n.e.s.	272	6	4	2	− 2
29 Construction	830	21	21	24	28
30 Gas	165	−10	− 6	− 5	1
31 Electricity	583	− 4	− 5	− 5	− 6
32 Water	54	41	30	7	−18
33 Transport and comns.	1459	− 3	− 1	3	14
34 Distribution	874	3	− 2	−15	−25
35 Services n.e.s.	2165	5	1	− 9	−19

Table 2.3—continued

Social Accounting Matrix Commodity	Inter-mediate Demand 1963 £mn. (1)	Percentage errors for a projection using mixed coefficients			
		$\theta = 0$ (2)	$\theta = -0{\cdot}1$ (3)	$\theta = -0{\cdot}5$ (4)	$\theta = -1$ (5)
36 Tobacco	105	− 9	−10	− 7	− 6
37 Crude mineral oil	376	−10	−12	− 7	− 3
38 Iron ore and scrap	78	23	29	25	26
39 Non-ferrous ores	77	21	16	23	25
40 Woodpulp	115	17	13	16	15
41 Butter	4	103	61	48	39
42 Tea and coffee	10	−55	−50	−53	−48
Total	23078	3·2	3·5	2·4	3·4
Mean squared error for vector of intermediate demand, £million		60	56	77	119
Mean squared error for matrix of intermediate demand, £million		11·5	10·8	12·1	13·7

The introduction of weak overall substitutability reduces the mean error of the best forecast made so far by just over 7 per cent, although it remains high at £56 million.

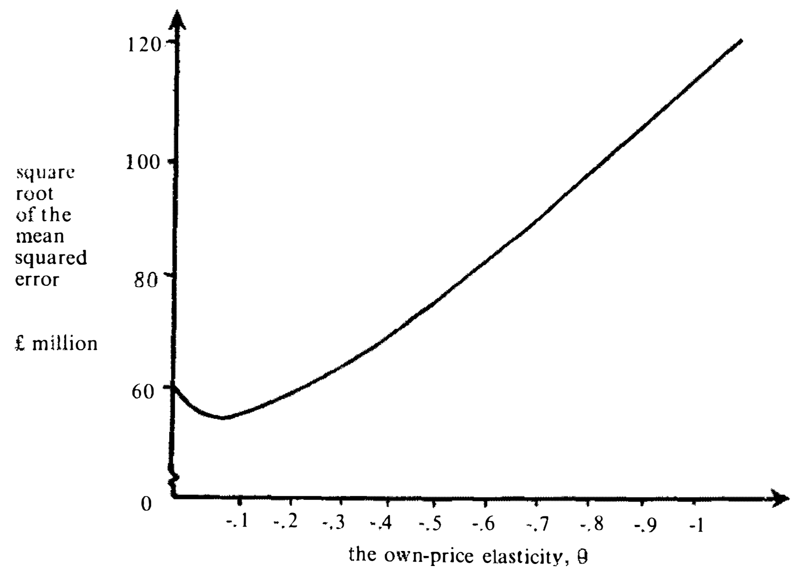

Figure 2.1 The Relation between the Own-Price Elasticity and the Mean Error

2.7 NON-HOMOGENEOUS PRODUCTION FUNCTIONS

Another suggestion for improving projections in input–output models has been the introduction of non-proportionality in the relation between input and output. Thus we have

$$\mathbf{x} = \mathbf{A}_0\mathbf{i} + \mathbf{A}_1\mathbf{q} \tag{2.7.1}$$

where $\mathbf{i}$ is the unit vector. This form allows for non-constant returns to scale for different inputs although, as in the analysis of trends in coefficients, at least two input–output tables are required for the estimation of $\mathbf{A}_0$ and $\mathbf{A}_1$. Ghosh [19] reports the results of experiments with this model on fairly aggregated tables for the U.K. and U.S. concluding that it might well yield better forecasts than the assumption of proportional input coefficients.

When this model was estimated on the 1954 and 1960 intermediate flows and used to project 1963 it gave disappointing results and the errors were generally larger than those in the other projections, with the exception of the ones using the 1954 matrix and values of the price substitution coefficient θ less than -0.5. The square root of the mean squared error of the elements of the vector of forecast final demand turned out to be £78 million and that of the elements of the matrix £11.8 million.

2.8 A COMPARISON OF THE PROJECTIONS OF INTERMEDIATE DEMAND

Inspection of the various projections of intermediate demands for 1963 which have been given in this paper reveals that there are several groups of commodities which behave according to type. Six such groups have been selected and summary statistics for the various projections have been calculated. These are given in Tables 2.4 and 2.5; Table 2.4 shows the square roots of the mean squared errors of the projections in each group of commodities and Table 2.5 shows the percentage error of the forecasts in each group defined as above.

Food, drink and tobacco

The projections of this group are all rather poor, with above average errors, and generally underestimate the observed 1963 values. There is no obvious reason why this should happen. An examination of the individual coefficients involved merely reveals that the 1963 coefficients are out of line with the 1954 and 1960 ones, the cereal inputs in 1963 being higher, cereal product inputs lower whilst inputs of other agricultural products are lower and processed food inputs higher than would be expected from the earlier observations.

Fuels

This is a particularly interesting group since the projections clearly demonstrate that the introduction of a price elasticity of substitution can

Table 2.4 Mean Errors for Intermediate Demand in Characteristic Groups, U.K. 1963

		Mean errors in projections of intermediate demand for 1963 in £mn.						
Category	SAM Commodity	A_{54} (1)	A_{60} (2)	$A_0 + A_1\tau$ (3)	$\theta = 0$ (4)	$\theta = -0.1$ (5)	$\theta = -0.5$ (6)	$\theta = -1$ (7)
1. Food, drink and tobacco	1, 4–7	53	36	40	36	44	53	61
2. Fuels	2, 8, 9, 30, 31	191	79	36	38	46	28	23
3. Products with low own-price elasticities	10, 24, 26, 28, 33–35	97	41	67	47	16	91	188
4. Metal products	11–13, 18	218	157	152	105	104	159	230
5. Products with high intermediate demand, 1960	15, 16, 22, 29	6	96	151	137	109	134	152
6. Textiles and clothing	19–21	170	84	18	16	81	19	43
Total		110	69	73	60	56	77	119

Table 2.5 Percentage Errors for Intermediate Demand in Characteristic Groups, U.K. 1963

Category	SAM commodity	Percentage mean errors in projections of intermediate demand 1963 using						
					$\hat{r}_{63}A_{54}\hat{s}_{63}$			
		A_{54} (1)	A_{60} (2)	$A_0 + A_1\tau$ (3)	$\theta = 0$ (4)	$\theta = -0.1$ (5)	$\theta = -0.5$ (6)	$\theta = -1$ (7)
1. Food, drink and tobacco	1, 4–7	1.9	−7.8	−12.1	−9.6	−12.3	−17.3	−18.8
2. Fuels	2, 8, 9, 30, 31	2.6	—	−2.4	−3.4	0.9	−1.6	0.3
3. Products with low own-price elasticities	10, 24, 26, 28, 33–35	−5.6	0.5	0.4	2.3	0.1	−1.5	−2.1
4. Metal products	11–13, 18	0.7	7.2	10.5	7.4	5.8	9.9	13.2
5. Products with high intermediate demand, 1960	15, 16, 22, 29	−0.6	14.7	22.9	20.4	14.8	20.2	21.5
6. Textiles and clothing	19–21	17.6	9.8	−2.6	−1.8	11.3	3.0	4.0
Total		0.2	3.3	4.7	3.2	3.5	2.4	3.4

markedly improve forecasts. The projection using an extrapolated input–output matrix and the one where θ, the own-price elasticity is less than -0.5 are both superior to the remainder: the time trend is proxy for a movement in coefficients caused by increasing cheapness of oil relative to coal in the period 1954–63 and substitution of one for the other as basic industrial fuel.

Products with low own-price elasticities

Chemicals, timber, paper, manufactures n.e.s., and the three services, transport, distribution and services n.e.s., are all products where allowance of a small degree of price substitution (θ between -0.05 and -0.1) improves the forecast of intermediate demand. Not too much credence can be put on some of these results since the price series for Chemicals and Services in particular are of poor quality; however it does make economic sense that the timber, paper and plastics bought by industry should show evidence of some substitution possibilities.

Metal products

The basic metal commodities, iron and steel and non-ferrous metals, together with the principal metal using commodities engineering products and metal goods are all closely linked so it is not surprising that projections of their intermediate demand move in unison. The forecasts are generally higher than the observed outcome and again they appear to be improved when some allowance is made for substitution.

Products with high intermediate demand in 1960

Motor vehicles, aircraft, construction and building materials are all projected very well when the 1954 input–output matrix is used and very badly when the coefficient levels of 1960 are allowed to affect the outcome. The demand for building materials is of course closely related to construction output, so it is to be expected that this demand will deviate with construction demand. The other products are capital goods and although the intermediate inputs are current inputs it appears that the high cyclical levels of gross investment in 1960 are affecting the values of the input–output coefficients for these inputs. Each of these products has an exceptionally high ratio of intra-industry demand to total industrial demand: this coefficient, the input of the commodity as a proportion of the output of the same commodity, is particularly sensitive to cyclical variation in output.

Textiles and clothing

Intermediate demand projections for these products are improved when the coefficients are allowed to change on trend, either individually or as the row and column multipliers vary through time.

2.9 CONCLUSIONS

The experiments in projecting intermediate demand reported in this paper confirm the need to use the most recent input–output table it is possible to construct, using if necessary an updating procedure like the RAS method. They also reveal the weakness of making one assumption about the behaviour of input–output coefficients, for example that they remain constant or that the own-price elasticity of industrial demand is -1, and invoking it for every coefficient. Certain groups of coefficients behave in a characteristic manner and this information should be used in formulating an input–output table for a future year. To the extent that industrial expertise is used in projecting the coefficients these characteristics will probably be incorporated in the projection but such expertise is best channelled by first estimating a future input–output table, then revising it in the light of known or expected changes. If this table already contains as much information about changes as can feasibly be gleaned from past tables, the industrial expert can concentrate on other changes which by their nature do not show up in the past statistics.

One final comment on the nature of the experiments is worth making. All the models tested are deterministic: comparisons between models can only be made by forecasting output or intermediate demand for one or more years in which these are known. The danger is that we make no allowance for the errors, in measurement and specification, that we know to exist in the data underlying the models. With only one input–output table little can be done about this if we retain the basic Leontief model; we must estimate as many coefficients as we have observations. With at least two independent tables the position changes and stochastic models can be constructed, allowing perhaps for the estimation of price elasticities. As we have seen, this form can be used to improve the forecasting power of input–output relationships. It would be a considerable advance if these effects could be introduced in a more general way and if comparisons between models could be based more firmly on their statistical framework.

[1] Alternative assumptions would make a slight difference to the results, but in view of the magnitude of the projection errors this can probably be disregarded.

[2] If z is the realised value and $\tilde{z}$ the forecast value we have measured mean squared errors as $\frac{\sum_i^n (\tilde{z}_i - z_i)^2}{n}$ whereas Tilanus used $\frac{\sum_i^n (\log \tilde{z}_i - \log z_i)^2}{n}$.

CHAPTER 3

Some Tests on a Generalised Version of RAS

R. I. G. ALLEN, *National Institute of Economic and Social Research* and J. R. C. LECOMBER, *University of Bristol*

3.1 INTRODUCTION

In the United Kingdom, the majority of input–output coefficients, particularly those relating to manufacturing industries, can be accurately determined only on the basis of the full quinquennial Census of Production which provides complete and detailed information about the purchases and sales of individual census trades. Since input–output relationships are now an integral part of national accounts statistics, this deficiency of up-to-date information has encouraged the development of various approximation techniques for revising or projecting a base-year set of input–output relationships given only a bare minimum of data for the year in which an approximated table is required. Of these techniques the RAS, or biproportional, method developed at Cambridge during the early 1960's [11] [37] is perhaps the most practicable and has become widely adopted. For example, the Central Statistical Office has used the technique as a basis for estimating the provisional United Kingdom input–output tables for 1963 [49] [53], for 1968 [48] and for 1970 [50].

It is surprising that so little attention has been paid to evaluating the performance of RAS and, until recently, no direct tests had been made using United Kingdom data. However, Allen [1] has demonstrated the dependence of intermediate output estimates on a relatively small number of major coefficients, and suggests that in non-census years great efforts should be made to obtain good data for these while mechanical methods such as RAS are confined to minor coefficients. In fact, a good deal of non-census information relating to particular input–output cells is available from sources such as the *Digest of Energy Statistics*, the *Annual Abstract* and various nationalised industry reports; data of this kind has been incorporated into the C.S.O.'s updating exercises. But such information is of very varying suitability and quality and while the incorporation of correct exogenous information must be a gain (Allen's experiments suggest a substantial gain), the value of less accurate information is less clear, especially if no account is taken of its possible inaccuracies.

This paper is divided into the following sections: first, the results and principal conclusions of Allen's paper are briefly summarised; second, the serious effect of incorporating inaccurate exogenous estimates of individual coefficients or of row and column totals on RAS updates are considered, using as an illustration the C.S.O's provisional version of the 1963 input–output table; third, the algebra of a generalised method of RAS, originated by Lecomber [28], which explicitly allows for the inclusion of a wide variety of information, and takes account of its varying reliability, is set out. Finally, this generalised model is tested on British data.

3.2 THE PERFORMANCE OF THE RAS METHOD

The basis of the RAS method is the hypothesis originated by Stone [37] that the various determinants of change in input–output coefficients (economies of scale, technological evolution, variations in relative prices, and so on) may be summarised by biproportional relationships in which each industry is characterised by a pair of 'substitution' and 'fabrication' multipliers (r_i and s_j respectively) which are assumed to operate uniformly over the rows and columns of the input–output matrix. In its simplest form, the RAS procedure involves the determination of the (unique) set of values for r_i and s_j which, when applied to an observed base year coefficient matrix **A**, generates a second matrix $\mathbf{A}^*$ whose elements are consistent with a pair of vectors $\mathbf{u}^*$ and $\mathbf{v}^*$ representing the observed values of total intermediate output and input by industry in the update years. In mathematical terms the problem is therefore to find

$$\mathbf{A}^* = \hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}} \tag{3.2.1}$$

such that

$$(\mathbf{A}^*\hat{\mathbf{x}}^*)\,\mathbf{i} = \mathbf{X}^*\mathbf{i} = \mathbf{u}^* \tag{3.2.2}$$

and

$$(\mathbf{A}^*\hat{\mathbf{x}}^*)'\,\mathbf{i} = \mathbf{X}^{*\prime}\mathbf{i} = \mathbf{v}^* \tag{3.2.3}$$

where $\mathbf{x}^*$ is the vector of gross industrial output in the update year, where $\mathbf{A}^*\mathbf{x}^* = \mathbf{X}^*$ is the estimated updated inter-industry flow matrix for the update year and where $\mathbf{i}$ is the unit summation vector (1, 1, ..., 1). The symbol (ˆ) placed above a vector indicates the formation of a diagonal matrix in which the elements of the vector are placed in the leading diagonal, with zeros elsewhere.

While the assumptions underlying equations (3.2.1), (3.2.2) and (3.2.3) are not entirely implausible, they have no particular economic justification, and their validity rests entirely on the empirical evidence. Experiments by Paelinck and Waelbroeck [34] on Belgian data (1953–59) and by Schneider [35] on United States data (1947–58) involved the comparison of cell values of the RAS update with those of the outturn matrix $\mathbf{A}^*$ and indicated that, on average, the RAS update was generally superior to the unadjusted matrix **A** as an estimate of $\mathbf{A}^*$. Paelinck and Waelbroeck also showed that RAS updates could be greatly improved by direct exogenous estimation of coefficients which had proved to be particularly unreliable or unstable.

Some tests by Allen [1] suggest that, in practice, intermediate demands are heavily dependent on a small number of key coefficients which tend to form part of a fairly stable hierarchical arrangement. This suggests a modification of the simple RAS procedure in which major coefficients for the forecast year are first identified and then estimated from exogenous information, the RAS adjustment being applied to the residual coefficients. Thus, if c_{ij} are the elements of the forecast matrix to be estimated exogenously, and if **C** is the matrix comprising these elements (with zeros elsewhere), the revised problem is to estimate

$$\begin{aligned}\mathbf{A}^* &= \mathbf{C} + \hat{\mathbf{r}}(\mathbf{A} - \mathbf{C})\hat{\mathbf{s}} \\ &= \mathbf{C} + \hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}}\end{aligned} \tag{3.2.4}$$

subject to the constraints (3.2.2) and (3.2.3) above. It should be noted that equation (3.2.4) is formally equivalent to Paelinck and Waelbroeck's modification in which, however, the c_{ij} entries are the values of those coefficients thought likely to be the most unstable between the base and update years.

In testing this model, Allen made projections of intermediate demands for 1968 using as his base matrix (i) the initial base-year matrix for 1954; (ii) the simple RAS update of (i) as given by equation (3.2.1); (iii) modified RAS updates of (i) as given by equation (3.2.4), assuming various levels of exogenous information about major coefficients, and (iv) the outturn matrix for 1963.

To select his major coefficients, Allen proposed a criterion (denoted by *I* in Table 3.1) in which the sensitivity of intermediate demand projections to changes in the values of particular coefficients was measured. To check the performance of this *ex ante* criterion a standard (*II*) was established in which coefficients were ranked according to the observed pattern of errors resulting from the estimation of the 1963 matrix by RAS.

Table 3.1 summarises the results by listing the mean projection errors arising from each of these experiments expressed as percentages of observed intermediate demands in 1968. It will be seen that, on average, the mean projection error falls from about 29 per cent of the control totals when the unadjusted 1954 matrix is employed to slightly over 14 per cent when the simple RAS update is used and to approximately 7 per cent with the observed 1963 matrix. The substantial reduction in the mean

Table 3.1 Percentage Errors in Estimating Intermediate Demands for 1968 from Various Updated Versions of the 1954 Input–Output Table

Industry	*Actual*	*RAS update of 1954 matrix to 1963 with percentage (θ) of exogeneously determined major coefficients for 1963*										*Actual*
		$\theta = 0$	$\theta = 0.05$		$\theta = 0.10$		$\theta = 0.20$		$\theta = 0.50$			$(\theta = 1)$
			I	*II*	*I*	*II*	*I*	*II*	*I*	*II*		
1	12.8	3.6	1.9	1.9	2.0	2.1	2.0	1.9	1.9	1.7		1.7
2	13.3	11.5	7.5	8.1	3.5	3.6	3.4	3.4	2.7	2.6		2.5
3	7.4	6.1	5.3	4.7	4.5	4.4	5.0	4.6	4.6	5.0		4.7
4	38.0	16.3	15.4	15.2	15.1	14.8	11.7	11.2	11.0	10.7		10.5
5	15.4	8.6	5.5	5.3	5.1	5.0	4.7	5.0	5.1	5.2		5.0
6	46.3	16.5	17.3	14.5	14.6	13.3	13.9	13.1	12.5	12.6		12.1
7	66.2	29.9	18.7	18.2	16.0	15.5	16.2	15.5	15.4	14.9		14.2
8	24.2	21.9	19.9	18.0	12.4	12.3	7.2	7.0	4.1	4.0		3.9
9	58.1	32.7	26.9	26.9	14.9	14.2	11.7	11.4	10.7	10.5		10.0
10	30.9	15.6	11.1	10.7	11.3	11.4	11.2	11.3	10.7	10.2		10.1
11	26.3	12.2	9.3	9.1	9.2	8.9	8.8	8.8	8.6	8.3		8.0
12	23.9	14.8	12.9	13.3	8.6	8.6	8.1	8.0	7.5	8.2		8.0
13	43.8	6.1	5.5	4.8	5.2	5.3	5.6	5.3	4.6	4.6		4.3
14	16.1	15.3	9.1	9.6	9.0	8.9	9.1	8.6	8.9	8.9		8.5
15	12.7	3.5	4.4	3.6	4.5	4.3	2.5	2.6	1.5	1.4		1.2
All industries	29.0	14.3	11.4	10.9	9.0	8.8	8.1	7.8	7.3	7.2		7.0

Source: Allen [1], Table 3. For industry definitions, see Appendix.

error by about 50 per cent following application of the simple RAS method is still further improved by comparatively small injections of exogenous information. For example, at the 10 per cent level of information the mean error (relative to projections based on the actual 1963 coefficient matrix) is negligible for the large majority of industries—in nine cases less than $1\frac{1}{2}$ per cent and greater than 5 per cent for only three industries. Furthermore, Criterion *I* generally works extremely well as a proxy for Criterion *II*.

3.3 THE EFFECT OF INACCURATE INFORMATION

A possible weakness of the above tests is that they do not take into account the serious anomalies which may arise if exogenous estimates of individual coefficients or of row and column constraints are inaccurate. This problem may be illustrated by reference to the Central Statistical Office's preliminary version of the 1963 input–output table published in 1966 [49]. This matrix was largely derived by RAS, the row and column totals being estimated exogenously from 1963 data (including the preliminary results of the 1963 census).

Table 3.2 Differences between Provisional and Final Estimates of 1963 Input–Output Table, Larger Elements[1, 2]

Error range per cent[3]	*Row totals*[4]	*Column totals*[4]	*Elements estimated exogenously*	*Elements estimated by RAS*	*All large elements*
under 1	1	1	2	1	3
1–	3	9	7	6	13
5–	7	5	1	7	8
10–	5	4	15	10	25
20–	4	3	16	22	38
50–	1	0	3	10	13
100 and over	1	0	8	5	13
Total	22	22	52	61	113

Sources: provisional estimates: Upton [53].
final estimates: C.S.O. [46]

Notes: 1. The transport, distribution and services industries have been aggregated for the purposes of this comparison.
2. Small elements (both provisional and the final estimate under £10 million) were omitted from the comparison; of 307 such elements, 141 were estimated at zero on both occasions.
3. The errors were defined as ((provisional estimate/final estimate) − 1).
4. Excluding diagonal elements.

In Table 3.2, the provisional estimates have been compared with the corresponding figures as derived from the final input–output accounts, with smaller elements being excluded from the comparison. From the extreme right-hand column of the table it will be seen that, out of 113 larger elements, 26 (23 per cent) were estimated with errors of 50 per cent or more and 79 per cent with errors of 10 per cent or more. This demonstrates the potential inaccuracy of an *operational* updating exercise and suggests that *ex post* tests of the kind discussed above must be approached with considerable caution.

The exogenous estimates are a major source of inaccuracy: 32 per cent of column totals, 50 per cent of row totals and 52 per cent of exogenous elements are estimated with errors of ten per cent or more; it was hardly appropriate to treat such estimates as reliable in filling in the remaining elements by RAS. The RAS-estimated elements are even more inaccurate, 61 per cent being erroneous by at least ten per cent. To quote a single example, the input of chemicals into textiles was initially estimated as 21 as compared with a final estimate of 78.3 (an error of 73 per cent). The under-estimate may be in part attributed to under-estimates of the row total for chemicals (by 10 per cent) and the column total for textiles (by 48 per cent). But a further factor is a failure in biproportionality due to the increasing share of man-made fibres in textile output. A rather better estimate of this cell could have been made using available information on the product mix within the textiles industry.

3.4 THE ALGEBRA OF A MODIFIED RAS METHOD

The assumptions underlying the standard RAS procedure may fairly easily be relaxed so that a greater variety of information of varying reliability may be incorporated into the method. Such a scheme was originally devised by Lecomber [28] and was used in the derivation of the base matrix for the Labour Government's National Plan of 1965.

Equation (3.2.4) may be given a wider interpretation by removing the condition that for all i and j, either c_{ij} or $e_{ij} = 0$. Specifically, let an initial estimate of $\mathbf{X}^*$ be made (denote by $\mathbf{Z}$), derived partly from $\mathbf{X}$, partly from exogenous data for the update year, but ignoring row and column constraints, and let the matrix $\mathbf{E}$ embody any views on the relative accuracy of cells of $\mathbf{Z}$.[1] $\mathbf{C}$ is then set equal to $\mathbf{Z} - \mathbf{E}$ so that (3.2.4) becomes

$$\mathbf{X}^* = (\mathbf{Z} - \mathbf{E}) + \hat{\mathbf{r}}\mathbf{E}\hat{\mathbf{s}} \tag{3.4.5}$$

subject as before to (3.2.2) and (3.2.3).

The procedure may be further generalised to allow for inaccuracy in the estimates of the row and column sums. Attach error estimates, $\mathbf{e}_u$ and $\mathbf{e}_v$, to $\mathbf{u}$ and $\mathbf{v}$ and estimate (3.4.5) together with

$$\mathbf{u}^* = \mathbf{u} + \mathbf{e}_u - \hat{\mathbf{r}}\mathbf{e}_u \tag{3.4.6}$$

$$\mathbf{v}^* = \mathbf{v} + \mathbf{e}_v - \hat{\mathbf{s}}\mathbf{e}_v \tag{3.4.7}$$

subject to (3.2.2) and (3.2.3).

Alternatively, $\mathbf{E}$, $\mathbf{e}_u$ and $\mathbf{E}_v$ may be adjusted in a fully biproportional scheme. Define $\mathbf{E}_A$ as the augmented matrix:

$$\mathbf{E}_A = \left[\begin{array}{c|c} \mathbf{E} & \mathbf{e}_u \\ \hline \mathbf{e}_v & 0 \end{array}\right] \tag{3.4.8}$$

and find the unique matrix $\mathbf{F}_A$ related to $\mathbf{E}_A$ by the biproportional relationship

$$\left[\begin{array}{c|c} \mathbf{F} & \mathbf{f}_u \\ \hline \mathbf{f}_v & 0 \end{array}\right] = \mathbf{F}_A = \hat{\mathbf{r}}\mathbf{E}_A\hat{\mathbf{s}} \tag{3.4.9}$$

and satisfying the constraints[2]

$$\mathbf{F}_A\mathbf{i} = \begin{pmatrix} \mathbf{u} + \mathbf{e}_u \\ \mathbf{e}_v\mathbf{i} \end{pmatrix} - \begin{pmatrix} (\mathbf{Z} - \mathbf{E})\,\mathbf{i} \\ 0 \end{pmatrix} \tag{3.4.10}$$

$$\mathbf{F}_A'\mathbf{i} = \begin{pmatrix} \mathbf{v} + \mathbf{e}_v \\ \mathbf{e}_u'\mathbf{i} \end{pmatrix} - \begin{pmatrix} (\mathbf{Z}' - \mathbf{E}')\,\mathbf{i} \\ 0 \end{pmatrix}. \tag{3.4.11}$$

$\mathbf{X}^*$, $\mathbf{u}^*$, $\mathbf{v}^*$ are then derived as:

$$\mathbf{X}^* = (\mathbf{Z} - \mathbf{E}) + \mathbf{F} \tag{3.4.12}$$

$$\mathbf{u}^* = (\mathbf{u} + \mathbf{e}_u) - \mathbf{f}_u \tag{3.4.13}$$

$$\mathbf{v}^* = (\mathbf{v} + \mathbf{e}_v) - \mathbf{f}_v \tag{3.4.14}$$

and it is easily checked that the constraints (3.4.13) and (3.4.14) imply

$$\mathbf{X}^*\mathbf{i} = \mathbf{u}^*, \qquad \mathbf{X}^{*\prime}\mathbf{i} = \mathbf{v}^*.$$

3.5 SOME TESTS ON THE MODIFIED RAS METHOD

To examine the performance of this generalised model the following five tests were performed, the results of which are summarised in Table 3.3:

(i) Standard RAS using ${}_{54}\mathbf{X}$ with $\mathbf{u}$ and $\mathbf{v}$ from the 1963 tables;
(ii) As (i) but inserting exogenous estimates for 'key coefficients' as taken from the 1963 tables;
(iii) As (ii) but using *pre-Census* estimates of the key coefficients;
(iv) As (iii) but using also 'provisional' estimates of the row and column totals;

(v) RAS modified according to the augmented matrix technique, incorporating 'provisional' estimates both for the elements of **X** and for the row and column totals, together with an augmented matrix reflecting subjective *ex ante* estimates of the relative reliability of this information.[3]

Table 3.3 Cumulative Error Distribution of Estimates of Larger Interindustry Flows[1]

Errors over specified per cent (cumulative basis)	Number of test (i)	(ii)	(iii)	(iv)	(v)
100 and over	6	7	10	18	1
50 and over	33	22	34	49	20
20 and over	72	66	82	89	71
10 and over	107	85	118	119	104
5 and over	122	92	129	128	122
1 and over	137	106	138	138	137
All	140	140	140	140	140
Median error	20.6	18.3	26.8	30.7	20.1

Note: 1. Actual flow over £10 million in 1963.

A comparison of the first two columns shows the favourable effect of inserting the correct values for 34 key coefficients.[4] Column (iii) shows the effect of using 'pre-Census' information of these cells. These estimates compare unfavourably with those not only of column (ii) but also of column (i), reflecting the poor quality of some of these exogenous estimates (cf. Table 3.2). Column (iv) shows the effect of using provisional estimates of the row and column totals as well. Some of these estimates are very poor, and the effect on the estimation of this matrix is serious. Finally comparison of column (v) with column (iv) (the only other set of estimates not making illegitimate use of Census information) indicates the strength of the modified RAS method. It is striking that column (v) also compares favourably with columns (i) and (iii) and even with column (ii) especially in respect of large errors, even though these three use Census information.

3.6 CONCLUSIONS

Provided the row totals are correct, errors in the updated flow matrix, and hence in the associated coefficient matrix, tend to be mutually offsetting, leading to offsetting errors in projection work and other (primal) applications. Errors in the estimated row totals are however

likely to be particularly damaging. An important feature of the modified RAS method is that it improves estimates of the row totals. It also improves estimates of the column totals, which is important in dual applications. The extent of the improvement, in this exercise, is shown in table 3.4.

Table 3.4 Cumulative Error Distribution of Estimates of Row and Column Totals

Errors over specified per cent (cumulative basis)	Errors in row totals: Initial estimates	Errors in row totals: RAS-adjusted estimates	Errors in column totals: Initial estimates	Errors in column totals: RAS-adjusted estimates
50 and over	0	0	0	0
20 and over	4	1	1	1
10 and over	7	3	4	4
5 and over	11	9	8	7
1 and over	14	12	14	13
All	15	15	15	15
Median error	9.7	6.8	5.5	4.0

As a brief postscript, it should be noted that in chapter 4 below Barker casts considerable doubt on the usual interpretation of the Paelinck and Waelbroeck tests. The control totals for the update year were, he points out, themselves derived by updating methods though making substantial use of industrial information. He concludes that the proper moral of these tests is that 'the RAS method, given extra information about a few special coefficients, does as well as industrial expertise in projecting the Belgian table' (p. 66). Likewise, he concludes from his own tests on the provisional 1963 tables for the U.K. that 'RAS as a method of updating appears to do equally well (or equally badly) as extraneous estimates provided by government departments' (p. 66).

The evidence presented in this Chapter fully supports these conclusions *so long as* extraneous information is treated, as it was in both these exercises, as accurate. However, as soon as allowance is made for its lack of reliability, the incorporation of such information leads to a very substantial improvement over RAS, as the tests described in this Chapter amply demonstrate.

[1] Thus, where an element is known with certainty, the corresponding e_{ij} is set equal to zero, so that the element is not modified in the RAS process; the more uncertain the element the higher e_{ij} is set. But note that there is a certain arbitrariness in the process in that the resultant estimates are affected by multiplying **E** by a scalar; purists should use an analogous Friedlander adjustment procedure which is not subject to this defect, at the same time

imposing side-constraints to prevent negative elements, and solving by general quadratic programming methods. The virtue of the messier RAS routines is their property of automatically preserving signs. Also note that there are some limitations on the permissible values of the e_{ij}.

[2] The final row constraint in equation (3.4.10) ensures that the sum of the row sums (and hence the sum of the column sums) is unchanged in the adjustment. The procedure may be further generalised by setting the final element of $\mathbf{E}_A$ equal to a positive number (depending on the error assigned to the sum) and suitably adjusting the constraints (3.4.10) and (3.4.11).

[3] The reader may wonder how suitable reliability estimates can be derived. For example, when a similar scheme was outlined many years ago Bacharach [5] complained that it made "demands on a delicate 'feeling for numbers' which not all possess". In the present exercise the elements were divided into six reliability grades on the basis of knowledge of how the various elements had been derived. Values of α = 0, 0.2, 0.4, 0.6, 0.8 and 1.0 were assigned to these grades and **E**-values were then set equal to α times the corresponding element value.
The tests were in fact carried out after the publication of the 1963 Census. It was crucial to their interpretation that additional census information was not introduced inadvertently. It was to avoid this danger that no attempt was made to derive independent estimates of **Z** and that the estimates used were those made by the C.S.O. and at the Department of Applied Economics, Cambridge before the Census was published. No such genuinely *ex ante* estimates of the **E**-values were available, but it should be noted that (i) we had not worked with the 1963 Census and were generally unfamiliar with it, and (ii) the **E**-values were written down quickly without much thought or any research.

[4] The improvement is almost entirely attributable to the 34 correct cells. There is virtually no improvement in the estimation of the remaining cells.

APPENDIX TO CHAPTER 3

PREPARING THE 1954 AND 1963 INPUT–OUTPUT TABLES

The basic raw material for the experiments were 15-industry, constant price input–output tables derived from the official U.K. tables for 1954 and 1963 (commonly called the Yellow Book [45] and the Purple Book [46] respectively) and a control vector of intermediate output for each of the 15 industries for 1968. The two sets of official tables are closely in agreement on most important points of methodology, but the following notes summarise some of the main problems of estimation which remained. The 15-industry matrices are shown in Tables A3.2 and A3.3.

(*i*) *Classification and aggregation.* The 1954 official tables are based on the 1948 Standard Industrial Classification (SIC) whereas the 1963 tables adopt the often quite different 1958 SIC. However, the summary 11-industry flow matrix given in Table 1 of the Yellow Book was adjusted by the C.S.O. to conform, as far as possible, to the 1958 SIC, and this table forms the basis for our 1954 matrix. We were able to supplement this information by including some additional sectors for which the effect of the change in industrial classification had either

been insignificant or could be readily accounted for, and we were consequently able to enlarge the list of industries under review to 15 (see Table A3.1 below). Where the main effect of the change in classification was to shift one or more complete census trades from one SIC Order to another, the input structure of the relevant trade was reconstructed from the appropriate census report, was consolidated in accordance with the 1958 SIC and was finally added to or subtracted from the relevant column or columns of the unadjusted 1954 table. These adjustments inevitably led to certain small inconsistencies between the total inputs and total outputs of some industries and the 15-industry table was

Table A3.1 Classification of Industries

Industry number	*Industry name*	*1958 SIC Order*	*Purple Book* industry number*
1	Agriculture, forestry and fishing	I	1, 2
2	Mining and quarrying	II	3, 4
3	Food, drink and tobacco	III	5–11
4	Chemicals and allied industries	IV	12–18
5	Metal manufacture	V	19–21
6	Engineering and electrical goods	VI	22–36
7	Vehicles	VII, VIII	39–42
8	Miscellaneous metal goods	IX	37, 38
9	Textiles, leather and clothing	X–XII	43–51
10	Building materials and allied industries	XIII	52–54
11	Timber, furniture, paper and allied industries	XIV, XV	55–59
12	Other manufacturing industries	XVI	61
13	Construction	XVII	62
14	Gas, electricity and water	XVIII	63–65
15	Services	XIX–XXIV	66–70

* U.K. Central Statistical Office, *Input–Output Tables for the United Kingdom, 1963*, [46], Appendix C.

balanced by a process of successively pro-rating rows and column cells until the marginal totals equalised. A final set of adjustments was required to account for changes in the C.S.O's treatment of customs and excise duties between 1954 and 1963.

(*ii*) *Valuation.* Each element in the 1963 flow table was deflated to the equivalent flow at 1954 prices. The ideal solution would have been to deflate each transaction recorded in the census returns for 1963 by an individual price index relating to the flow itself and preferably imputed from a comparison of the corresponding quantity and value estimates, but lack of detail in the Census reports ruled out such an approach. The problem is further complicated by the ambiguity of the 'net selling value' concept employed in the census: intermediate purchases may be valued on either an 'ex works' or a 'delivered' basis, but the method of valuation employed in any given case is not recorded in the Census reports. The

Table A3.2 Industry × Industry Coefficients for 1954 at 1954 Prices

Industry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	—	—	0.180	—	—	—	—	—	0.008	—	0.009	0.001	—	—	—
2	0.007	—	0.004	0.082	0.027	0.005	0.003	0.004	0.011	0.143	0.013	0.036	0.018	0.243	0.014
3	0.173	—	—	0.013	—	—	—	—	0.001	—	0.002	—	—	—	0.006
4	0.082	0.021	0.032	—	0.076	0.024	0.021	0.017	0.030	0.026	0.033	0.095	0.042	0.022	0.016
5	0.001	0.031	0.002	0.017	—	0.159	0.153	0.233	0.002	0.010	0.007	0.011	0.073	0.023	0.003
6	0.019	0.044	0.009	0.020	0.042	—	0.081	0.034	0.023	0.032	0.018	0.025	0.035	0.049	0.009
7	0.010	0.003	0.002	0.001	0.007	0.008	—	0.008	0.002	0.004	0.002	0.005	0.005	0.002	0.034
8	0.016	0.007	0.010	0.019	0.027	0.045	0.062	—	0.009	0.008	0.013	0.020	0.020	0.004	0.002
9	0.006	0.008	0.004	0.008	0.002	0.010	0.012	0.007	—	0.006	0.036	0.095	0.004	0.001	0.007
10	0.002	0.003	—	0.003	0.009	0.002	0.001	—	—	—	—	0.006	0.114	0.008	0.001
11	0.003	0.019	0.024	0.019	0.004	0.021	0.021	0.021	0.011	0.052	—	0.031	0.030	0.003	0.041
12	0.005	0.013	0.006	0.010	0.002	0.019	0.030	0.018	0.010	0.002	0.010	—	0.017	0.006	0.009
13	0.026	0.031	0.003	0.005	0.005	0.004	0.004	0.004	0.007	0.004	0.004	0.006	—	0.003	0.014
14	0.010	0.020	0.005	0.021	0.023	0.012	0.010	0.016	0.005	0.030	0.009	0.022	0.003	—	0.021
15	0.114	0.098	0.089	0.124	0.169	0.090	0.082	0.095	0.123	0.177	0.146	0.103	0.082	0.118	—

Source: See text.

Table A3.3 Industry × Industry Coefficients for 1963 at 1954 Prices

Industry	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	—	—	0.138	0.002	—	—	—	—	0.015	0.001	0.006	—	—	—	0.001
2	0.003	—	0.006	0.072	0.021	0.002	0.003	0.002	0.009	0.080	0.012	0.006	0.020	0.218	0.003
3	0.194	0.001	—	0.011	—	—	—	—	0.001	0.002	—	—	—	—	0.003
4	0.074	0.013	0.033	—	0.077	0.024	0.016	0.025	0.048	0.058	0.037	0.149	0.023	0.031	0.016
5	0.003	0.019	0.007	0.013	—	0.125	0.137	0.274	0.002	0.011	0.011	0.010	0.045	0.018	0.002
6	0.006	0.033	0.009	0.016	0.051	—	0.094	0.024	0.016	0.026	0.015	0.015	0.036	0.051	0.015
7	0.006	0.003	0.001	0.002	0.016	0.009	—	0.004	0.002	0.005	0.001	0.003	0.003	0.003	0.017
8	0.016	0.004	0.022	0.018	0.020	0.046	0.092	—	0.010	0.015	0.012	0.092	0.019	0.004	0.009
9	0.006	0.008	0.002	0.003	0.001	0.005	0.010	0.004	—	0.006	0.020	0.064	0.002	0.001	0.009
10	0.004	0.005	0.002	0.008	0.018	0.012	0.006	0.002	0.001	—	0.005	0.003	0.123	0.006	—
11	0.004	0.010	0.043	0.022	0.005	0.021	0.018	0.012	0.016	0.034	—	0.044	0.060	0.005	0.041
12	0.005	0.011	0.006	0.009	0.005	0.016	0.032	0.006	0.016	0.007	0.012	—	0.013	0.002	0.010
13	0.018	0.013	0.004	0.010	0.004	0.004	0.004	0.003	0.003	0.003	0.003	0.002	—	0.003	0.009
14	0.012	0.020	0.010	0.029	0.045	0.013	0.012	0.020	0.015	0.041	0.013	0.017	0.004	—	0.017
15	0.142	0.051	0.131	0.165	0.126	0.101	0.075	0.112	0.128	0.135	0.152	0.125	0.031	0.082	—

Source: See text.

simpler, though less satisfactory, technique adopted here was to apply a single value deflator to all the outputs, intermediate or final, of each industry using the Department of Trade and Industry's indices of mean wholesale prices for materials as published in the *Annual Abstract of Statistics* which comply with the 1958 SIC and are also grounded on 1954 prices. For industries where it was not possible to obtain direct estimates of price deflators which corresponded exactly with our own industrial classification, it was necessary either to use an index number of materials purchased, or to build up a weighted price deflator from index numbers relating to individual commodities, where the weights were the relative value of each commodity output in the industry's product-mix.

(*iii*) *Intermediate outputs in 1968.* To act as control totals when making the projections of intermediate demand to 1968 we used a vector of intermediate output derived from the provisional set of 35 industry input–output accounts for 1968 prepared by the C.S.O. [48]. These estimates are based partly on published annual production series and on annual data for certain important industrial inputs (e.g. iron and steel), partly on the preliminary returns to the 1968 Census of Production and partly on approximations of coefficients derived from RAS updates of 1963 input–output coefficients. Despite the fact that these tables are adjusted to conform with other independently estimated national income magnitudes, this procedure may impart serious degrees of bias to some of the estimates. In addition, there is a further awkward change in the classification of industry groups onto the revised 1968 SIC. In most instances, however, the important changes in definition involved transfers of trades or sub-trades within an SIC Order. For example, although the production of paint and printing ink formed one census trade in the 1958 SIC and were separated in the 1968 SIC, they both form part of Order IV Chemicals and allied industries. In the few cases when a change in classification ran across one of our industry boundaries, complete trades were involved. The principal examples of such trades were Vegetable and animal oils and fats and Engineers' small tools and gauges. Having obtained estimates of the values of gross and final output for these activities in 1968 from census and national income data, it was then possible to adjust the C.S.O's output totals to correspond broadly with the 1958 SIC. These estimates were modified by a set of 1968 wholesale price deflators which also had to be reconciled with the 1958 SIC. Individual price indices for product categories which 'switched' industries between 1963 and 1968 were therefore estimated as accurately as possible and, by applying appropriate gross output weights, the price indices by broad industrial categories were adjusted.

CHAPTER 4

An Analysis of the Updated 1963 Input–Output Transactions Table

T. S. BARKER

University of Cambridge, Department of Applied Economics

4.1 Introduction

The experience of the U.K. Central Statistical Office in updating the 1954 industry × industry transactions table to provide provisional estimates of the 1963 table was highly relevant to the work of the Cambridge Growth Project in updating input–output tables annually between 1954 and 1968. Previous experience in updating—notably the widely quoted study by Paelinck and Waelbroeck [34]—appears to be highly favourable to RAS as an updating method. Thus their study showed that in estimating the 1959 input–output table for Belgium using the 1953 table and control totals for 1959, the simple RAS method produced only 9 elements with an error greater than 1 per cent (in absolute value) out of a total of 270 non-zero elements.

The C.S.O. produced two provisional tables for 1963, in 1966 [49] and 1968 [53] before the final table was published in 1970 [46]. The differences between the provisional tables and the final one are striking and appear to contradict the Belgian experience in updating input–output coefficients. An analysis of the discrepancies is provided by Allen and Lecomber in Chapter 3 above. They show that there were substantial errors in estimating the row and column totals of the provisional table in addition to those in the independently estimated elements of the absorption matrix. However, it is not clear from the analysis what proportion of the total errors is due to errors in the row and column constraints as distinct from the application of RAS to these constraints. The Belgian tests concerned RAS only, the row and column constraints being the ones in the final tables.

This chapter presents an attempt to distinguish the errors in the 1963 provisional table that are directly attributable to RAS. The 1966 version of this table has been taken and aggregated to 12 industrial sectors, for comparison with the final table. The errors are analysed as four components: those due to exogenous estimates of the elements of the table;

Table 4.1 The Final Summary Input–Output Transactions Table for 1963

Industry	1 Agri-culture	2 Mining	3 Food	4 Mineral oil	5 Chemi-cals	6 Metals	7 Engi-neering	8 Textiles	9 Other manu-factur-ing	10 Con-struc-tion	11 Gas	12 Services n.e.s.
	£mn.											
1. Agriculture, forestry, fishing	—	—	398	—	3	—	—	31	11	—	1	14
2. Mining and quarrying	5	—	16	3	166	39	17	18	84	65	346	41
3. Food, drink and tobacco	324	1	—	—	25	—	—	1	1	—	—	32
4. Mineral oil refining	29	8	12	—	30	18	27	7	26	21	40	106
5. Other chemicals	95	13	110	21	—	125	141	92	198	53	8	99
6. Metal manufacture	6	41	21	1	29	—	1197	3	36	147	28	29
7. Engineering	47	65	94	12	72	159	—	58	120	186	87	505
8. Textiles, leather and clothing	10	16	5	—	7	1	50	—	90	7	1	109
9. Other manufacturing	21	44	167	2	91	55	378	70	—	635	21	642
10. Construction	30	21	13	—	22	7	30	8	10	—	4	112
11. Gas, electricity and water	20	34	28	6	62	83	110	31	66	11	—	209
12. Services n.e.s.	236	86	380	75	313	231	741	264	477	218	130	—

Source: U.K., Central Statistical Office, [46].

those due to RAS; those resulting from a change in the composition of final demand (except stockbuilding) between the provisional and final tables; and lastly those due to errors in estimating the remaining columns of final demand and the rows of taxes, imports and other primary inputs.

4.2 THE 1963 INPUT–OUTPUT TRANSACTIONS MATRIX

Table 4.1 shows the final estimates of purchases by industry of domestic industrial production, as given in the final tables for 1963 [46]. The differences between the provisional and final estimates are given as absolute values in Table 4.2 and as percentages of the final estimates in Table 4.3. The *1966* version of the provisional table was chosen as it was likely to show the greater error. However, a comparison between the 1966 and 1968 provisional tables shows that the row and column totals were not for the most part revised. Table 4.4 shows a comparison of the errors in the 1966 and 1968 provisional tables, where only those elements greater than £10 million in 1963 are chosen. These results would indicate that if anything the 1968 version was more inaccurate than the 1966 one, although the probable explanation is the greater disaggregation in the 1968 table. (Allen and Lecomber estimate the errors on a 22 industry table, taking transport, distribution and other services as one sector.) Certainly the pattern of errors for both provisional tables is very similar.

These errors are disturbingly high: nearly all (80 per cent) of the larger elements contain errors greater than 10 per cent if we assume that the final tables are correct. About a quarter show errors greater than 50 per cent.

4.3 AN ANALYSIS OF THE REVISIONS TO THE PROVISIONAL TABLE

The differences between the 1966 and 1970 version of the input–output table shown in Tables 4.2, 4.3 and 4.4 above can be divided into four components. First there are the revisions to the exogenous elements in the table—those elements which can be directly estimated from industry data without the final Census of Production Reports being available. Information was given by government departments for the inputs and outputs of Agriculture, Forestry and fishing, Coal mining, Mineral oil refining, and Gas, electricity and water. This list covers the major fuel industries so that those input–output flows most likely to be affected by changes in fuel prices and technology are not derived by RAS adjustment. This is important since in the Belgian study [34] it was observed that the fuel coefficients were particularly badly explained by RAS and the adjustment method performed appreciably better when these were estimated exogenously.

A second component of the difference is due to the RAS adjustment

Table 4.2 Revisions to the 1966 Version of the 1963 Table (Actual)

Industry	1 Agri-culture	2 Mining	3 Food	4 Mineral oil	5 Chemi-cals	6 Metals	7 Engi-neering	8 Textiles	9 Other manu-factur-ing	10 Con-struc-tion	11 Gas	12 Services n.e.s.
						£mn.						
1. Agriculture, forestry, fishing	—	−5	−173	—	3	—	—	3	2	—	1	14
2. Mining and quarrying	4	—	−1	1	1	13	−12	2	−11	18	1	−3
3. Food, drink and tobacco	9	1	—	—	−10	—	—	−1	−2	—	—	−69
4. Mineral oil refining	9	6	2	—	−50	−3	6	—	7	—	8	54
5. Other chemicals	4	−3	−15	−27	—	40	−97	66	53	−58	−14	−62
6. Metal manufacture	6	−7	16	1	3	—	54	1	21	23	−1	1
7. Engineering	11	21	27	10	−11	83	—	33	20	4	17	−178
8. Textiles, leather and clothing	6	8	−8	—	−7	−1	−36	—	−14	−1	1	22
9. Other manufacturing	−40	−5	76	−10	11	40	−60	54	—	218	−4	−190
10. Construction	—	3	—	—	11	−1	−3	1	−4	—	−4	−5
11. Gas, electricity and water	4	6	−14	1	2	23	−23	−4	−17	2	—	−27
12. Services n.e.s.	−14	−10	−60	8	−18	−24	−75	105	−15	−29	−31	—

Sources: U.K., Central Statistical Office [46] [49].

Table 4.3 Revisions to the 1966 Version of the 1963 Table (Percent)

Industry	1 Agri-culture	2 Mining	3 Food	4 Mineral oil	5 Chemi-cals	6 Metals	7 Engi-neering	8 Textiles	9 Other manu-factur-ing	10 Con-struc-tion	11 Gas	12 Services n.e.s.
						Percent of Final Estimate						
1. Agriculture, forestry, fishing	—	—	−44	—	100	—	—	10	18	—	100	100
2. Mining and quarrying	80	—	−6	33	1	33	−71	11	−13	28	—	−7
3. Food, drink and tobacco	3	100	—	—	−40	—	—	−100	−200	—	—	−216
4. Mineral oil refining	31	75	17	—	−167	−17	22	—	27	—	20	51
5. Other chemicals	4	−23	−14	−129	—	32	−69	72	27	−109	−175	−63
6. Metal manufacturing	100	−17	76	100	10	—	5	33	58	16	−4	3
7. Engineering	23	32	29	83	−15	52	—	57	17	2	19	−35
8. Textiles, leather and clothing	60	50	−160	—	−100	−100	−72	—	−16	−14	100	20
9. Other manufacturing	−190	−11	45	−500	12	73	−16	77	—	34	−19	−30
10. Construction	—	14	—	—	50	−14	−10	12	−40	—	−100	−5
11. Gas, electricity and water	20	18	−50	17	3	30	−21	−13	−26	18	—	−13

Table 4.4 Errors in the Provisional Input–Output Transactions Matrix for 1963: the 1966 and 1968 Versions

Error range per cent	Number		Per Cent	
	1968 Version	1966 Version	1968 Version	1966 Version
under 1	3	5	3	5
1–	13	9	11	9
5–	8	5	7	5
10–	25	29	22	30
20–	38	26	34	27
50–	13	17	12	17
100 and over	13	6	12	6
Total	113	75	101	99

Sources: Chapter 3 for the 1968 version of the provisional tables and Tables 4.1 and 4.2 above for the 1966 version.

of the 1954 transactions to the 1963 control totals, with all the exogenously determined rows and columns removed from the table. The control totals in the provisional 1963 table are estimated from the first results of the 1963 Census of Production together with National Income Accounting data for the categories of final expenditure. However, since these control totals were substantially revised between 1966 and 1970, the revisions in the elements of the transactions table are partly due to the RAS adjustment and partly due to the control total revision. To isolate the RAS contribution, the provisional table was re-estimated using the final control totals. The differences between these estimates of the elements of the table and the final estimates were due to the RAS adjustment.

The remaining differences between the provisional and final elements are due to revision in the row and column totals of industrial demands. These can be divided into those which arise because Blue Book [47] figures are revised, and those which result from a different division of Blue Book expenditures between the products of the different industries. This division has been made because the estimation of the reclassified expenditures is more uncertain than the expenditures themselves. The errors due to changes in the composition of final expenditures have been estimated by recalculating consumers', government and investment expenditures for the final 1963 table using the composition given in the provisional (1966) table. Expenditures on those commodities for which independent information was available were excluded, and the treatment of excise duties on drink and tobacco was made consistent as between the

provisional and final tables. This new pattern of final expenditure, adding to the same totals as the final table, were used to provide new totals in a second RAS adjustment. The differences between the two adjusted tables gives the errors due to revisions in the composition of final domestic expenditures excluding stockbuilding. The remaining differences are the result of changes in the export and stockbuilding columns, the Blue Book figures for expenditures by consumers, government and investors, the import and indirect tax rows and the estimates of value added by labour and capital.

The errors attributable to these four factors are shown in Table 4.5 classified by the absolute size of the percentage error. Only the errors in the larger elements of the table are shown here, 'larger' being defined as entries exceeding £10 million in the provisional or final transactions table.

Table 4.5 An Analysis of the Errors in the Provisional (1966) Input–Output Transactions Matrix for 1963

Error range (positive or negative) per cent	Number of errors in:				Total	
	Exogenous elements (1)	RAS elements (2)	Composition of final demand (3)	Exogenous totals (4)	Gross (5)	Net (6)
under 1	4	3	5	2	14	5
1–	4	2	11	5	22	9
5–	4	4	9	2	19	5
10–	16	14	10	12	52	29
20–	13	16	7	17	53	26
50–	7	5	2	5	19	17
100 and over	4	—	—	1	5	6
Total	52	44	44	44	184	97
	Percentage of total number of errors					
under 1	8	7	11	5	8	5
1–	8	5	25	11	12	9
5–	8	9	20	5	10	5
10–	31	32	23	27	28	30
20–	25	36	16	39	29	27
50–	13	11	5	11	10	17
100 and over	8	—	—	2	3	6
Total	101	100	100	100	100	99

Since the errors for each factor tend to cancel out, the overall errors given in the sixth column are less than the sum of the errors given in the fifth column. This is demonstrated in Table 4.6 which gives the breakdown of the larger errors (those greater than ±£50 million). The table also shows that since few very large flows in absolute values were estimated exogenously there are only three such flows estimated with errors greater than £50 million.

A comparison of the sources of error show that in general the exogenously estimated elements have proven the most unreliable, followed by the effects attributable to the exogenous totals which in turn are hardly bettered by those due to RAS adjustments with the composition of demand having smaller although noticeable effects.

Table 4.6 An Analysis of Coefficient Errors in Updating the 1963 Table

		£ million					
				Number of errors in:			
Sales by	*Purchases by*	*Final 1963 value*	*Total error*	*Exogenous elements*	*RAS elements*	*Composition of final demand*	*Exogenous totals*
9. Other manuf.	10. Construction	635	218	—	86	1	130
9. Other manuf.	12. Services	642	−190	—	−76	4	−118
7. Engineering	12. Services	505	−178	—	−2	61	−237
1. Agriculture	3. Food	398	−173	−173	—	—	—
12. Services	8. Textiles	264	105	—	−45	7	142
5. Chemicals	7. Engineering	141	−97	—	−38	−29	−30
7. Engineering	6. Metals	159	83	—	41	14	28
9. Other manuf.	3. Food	167	76	—	38	3	35
12. Services	7. Engineering	741	−75	—	155	43	−273
3. Food	12. Services	32	−69	—	−4	8	−73
5. Chemicals	8. Textiles	92	66		−39	−12	39
5. Chemicals	12. Services	99	−62	—	1	−20	−43
9. Other manuf.	7. Engineering	378	−60	—	−89	21	8
12. Services	3. Food	380	−60	—	−39	21	−43
5. Chemicals	10. Construction	53	−58	—	−50	−22	14
4. Mineral oil	12. Services	106	54	54	—	—	—
9. Other manuf.	8. Textiles	70	54	—	24	—	30
6. Metals	7. Engineering	1197	54	—	−11	15	50
5. Chemicals	9. Other manuf.	198	53	—	45	−16	24
4. Mineral oil	5. Chemicals	30	−50	−50	—	—	—

4.4 A COMPARISON WITH THE BELGIAN TESTS

The results of the RAS adjustment on the 1954 table given above appear to contradict those of the Paelinck and Waelbroeck study [34] on Belgian data. In the U.K. test nearly all the errors are greater than 1 per cent; in the Belgian test only 9 out of 270 potential errors fell into this category. It does not seem likely that the two most obvious differences between the studies (level of aggregation and time span) can explain these strikingly different results. The Belgian table covered 21 industries whilst the one updated in this paper covers 12 and there are six years between the base and updated Belgian tables and nine years between the British tables. These differences would tend to lead to worse results in the British test, but they should not completely upset the Belgian findings.

Table 4.6 cont'd

		As per cent of final 1963 values				
			Number of errors in:			
Sales by	*Purchases by*	*Total error*	*Exogenous elements*	*RAS elements*	*Composition of final demand*	*Exogenous totals*
9. Other manuf.	10. Construction	34	—	14	—	21
9. Other manuf.	12. Services	−30	—	−12	1	−18
7. Engineering	12. Services	−35	—	—	12	−47
1. Agriculture	3. Food	−44	−44	—	—	—
12. Services	8. Textiles	40	—	−17	3	54
5. Chemicals	7. Engineering	−70	—	−27	−20	−22
7. Engineering	6. Metals	52	—	26	9	17
9. Other manuf.	3. Food	45	—	23	2	21
12. Services	7. Engineering	−10	—	21	6	−37
3. Food	12. Services	−216	—	−13	26	−229
5. Chemicals	8. Textiles	72	—	43	−13	42
5. Chemicals	12. Servicing	−63	—	1	−21	−43
9. Other manuf.	7. Engineering	−16	—	−23	5	2
12. Services	3. Food	−16	—	−10	6	−11
5. Chemicals	10. Construction	−109	—	−94	−42	26
4. Mineral oil	12. Services	51	51	—	—	—
9. Other manuf.	8. Textiles	77	—	34	—	43
6. Metals	7. Engineering	5	—	−1	1	4
5. Chemicals	9. Other manuf.	27	—	23	−8	12
4. Mineral oil	5. Chemicals	−167	−167	—	—	—

The explanation probably lies in the fact that the 1959 Belgian table was not directly estimated, but relied heavily on the 1953 table. Paelinck and Waelbroeck make it fairly clear that the 1959 table they use is provisional, estimated by extrapolating the 1953 tables using available statistics for 1959 flows as well as industrial expertise on the likely changes between 1953 and 1959. Although a final table for 1959 was available, the authors stress that 'le tableau definitif [for 1959] ... n'est en fait plus comparable du tout au tableau de 1953.' The reason is presumably the revisions to classifications of establishments to industries and those of industries themselves comparable to revisions of the Standard Industrial Classifications between the British censuses of production for 1954 and 1963.

Paelinck and Waelbroeck compared the table for 1959 produced by updating the 1953 table using partial information and industrial expertise with the one produced by RAS. The conclusion to be drawn from their results should be that the RAS method, given extra information about a few special coefficients, does as well as industrial expertise in projecting the Belgian table. It is misleading to imply that RAS as a method can do almost as well in providing an up-to-date table as taking a census of production. Rather it does almost as badly, judging from British evidence, as using partial information on particular flows. It seems equally misleading to conclude from the Belgian study that input–output coefficients are relatively stable. The inherent conservatism of forecasters of technical change is well known and understandable: therefore it is not too surprising that in the updated 1959 Belgian table, 132 out of 270 non-zero elements showed no change from the 1953 values.

An assumption underlying this interpretation is that the results of applying RAS to the British table for 1954 to provide a provisional 1963 table are not materially affected by errors in reclassifying and updating the 1954 table so as to conform to conventions and definitions of the 1963 one—that is from a 1948 SIC to a 1958 one. Woodward [58] describes some of the problems in estimating a 70 × 70 sectoral table for 1954 on the same basis as the 1963 table but he was able to use data in the 1963 Census on earlier flows in the same SIC categories. This was presumably not available when the provisional 1963 tables were constructed. However, it is doubtful that minor changes in classification, which could not be taken into account in preparing the 1954 table to be used as a base in the updating exercise, could result in errors of over 20 per cent.

4.5 CONCLUSIONS

RAS as a method of updating appears to do equally well (or equally badly) as extraneous estimates provided by government departments

for the provisional 1963 table. In fact the revisions to the so-called 'hard' data are so great that it would possibly have been more accurate to estimate all the cells of the table by RAS. However it is only fair to point out that if the purpose of updating is to provide estimates of input–output coefficients rather than transactions then the difference in coefficients will be much smaller than those in transactions when the tables are revised since both an industry's inputs and outputs tend to be revised together.

Other tests of RAS as a method of updating the 1954 table, this time constructed after the 1963 Census results became available, confirm the general magnitude of the errors involved in the estimates of the larger cells (see Ch. 2, Table 2.2). It remains to be demonstrated whether a more complicated updating procedure involving the extrapolation of trends in coefficients would produce better estimates than the simple RAS projection.

What clearly emerges from this study is the importance of providing the correct row and column constraints in the provisional table. Revisions to these totals account for the major part of revisions to the elements of the table which were not estimated exogenously. An examination of the revisions reveals that for the most part they were in the row totals, the result of revising the classification converters which produce the final demands for industrial products. It is reassuring that the C.S.O. have recently been concentrating on improving this area of their updating methods.

CHAPTER 5

Technology Assumptions in the Construction of U.K. Input–Output Tables

A. G. ARMSTRONG
University of Bristol

5.1 INTRODUCTION

In many economies the basic data for the construction of input–output tables is collected in the form of purchases of commodities by industries. Various methods can be used to convert this data into the standard form of either commodity × commodity or industry × industry tables; these have been discussed by Stone [37], Stone, Bacharach and Bates [11], United Nations System of National Accounts [52] and Gigantes [20] [21]. The two relatively straightforward methods which involve the assumption of either a commodity technology or an industry technology can be developed into various forms of mixed or hybrid technology assumptions. One form of hybrid technology assumption was used in calculating the commodity × commodity tables for 1963 in the U.K. (C.S.O. [46]). This chapter aims to examine this use of a hybrid technology assumption in practical work, to evaluate the problems involved and to compare the results with those of the simple technology assumptions. It seems desirable to do this before the theoretical treatment of this subject is further developed.

The derivation of commodity × commodity tables using both a commodity and an industry technology assumption is illustrated in section 5.3 where it is shown that the industry technology solution requires an assumption to be made not only about input structures but also about the output structures of industries if commodity × commodity tables are being used in projection work.

The use of the hybrid technology assumption is examined in section 5.4 and it is shown that care must be taken to distinguish between the use of input assumptions and output assumptions. The methods used to mix the technology assumptions and to remove negative coefficients in the derivation of the U.K. tables for 1963 are described in section 5.5.

The final section of the chapter suggests how the various forms of tables discussed here can be regarded as forming a triangle, the corners

of which are given by the simple assumptions. Commodity × commodity tables calculated on the commodity, the industry, and the hybrid technology assumptions are compared and the amount of variation within this triangle is examined. The conclusion is reached that apart from a few particular instances the proportion of subsidiary production is low and that there are not marked differences between tables calculated on the different technology assumptions. The smaller cells of off-diagonal production can probably be treated fairly arbitrarily; the use of a hybrid technology approach is recommended for the larger cells as giving flexibility and it is shown in the paper that even more flexibility can be achieved in the adjustment process to remove negative coefficients.

5.2 NOTATION

The notation used in this paper follows very closely that of the U.N. System of National Accounts (S.N.A.) (U.N. [52]) and the essential parts of the input–output accounting framework are reproduced below.

Matrices are shown as capital letters and a prime (′) superscript is used to indicate transposition. Vectors are written as column vectors and are shown by small letters; row vectors are shown as transposed column vectors. The symbol (ˆ) above a vector is used to indicate a diagonal matrix with the elements of the vector in the diagonal.

Table 5.1 Notation for Flows

	Commodities	*Industries*	*Final demand*	*Total*
Commodities		$\mathbf{X}$	$\mathbf{f}$	$\mathbf{q}$
Industries	$\mathbf{M}$			$\mathbf{g}$
Primary Inputs		$\mathbf{y}'$		
Total	$\mathbf{q}'$	$\mathbf{g}'$		

From this accounting framework various other matrices can be calculated as shown below. The derivation of some elements is not shown because various methods exist which yield different solutions.

The matrices and vectors are defined as follows:

A commodity × commodity coefficient matrix
B coefficients relating to purchases of commodities by industries
C product–mix matrix, the columns of which show the proportions in which a particular industry produces various commodities

Table 5.2 Notation for Coefficients

	Commodities	*Industries*	*Final demand*
Commodities	$\mathbf{A}$ $\mathbf{W} = \mathbf{A}\hat{\mathbf{q}}$	$\mathbf{B} = \mathbf{X}\hat{\mathbf{g}}^{-1}$	
Industries	$\mathbf{C} = \mathbf{M}'\hat{\mathbf{g}}^{-1}$ $\mathbf{D} = \mathbf{M}\hat{\mathbf{q}}^{-1}$	$\mathbf{E}$ $\mathbf{Z} = \mathbf{E}\hat{\mathbf{g}}$	$\mathbf{e}$
Primary Inputs	$\mathbf{z}'$		

D market share matrix, the columns of which show the proportions in which various industries produce the total output of a particular commodity
E industry × industry coefficient matrix
M make matrix showing the values of commodities produced by industries
X the values of purchases of commodities by industries—the absorption matrix
W the values of purchases of commodities by commodities (the commodity × commodity flow matrix)
Z the value of purchases of the industrial outputs by industries (the industry × industry flow matrix)
e final demands for the output of industries
f final demands for commodities
g industry outputs
q commodity outputs
y primary inputs into industries
z primary inputs into commodities.

5.3 COMMODITY × COMMODITY TABLES

Many establishments (or producing units) produce only one commodity or range of commodities which are the characteristic product(s) of the industry to which they are classified. Some establishments produce other commodities which are not among the characteristic products of the industry to which they are classified. As a result, in the make matrix, **M**, industries are often recorded as producing several commodities. The amount of this subsidiary production varies between industries and is often not large. However, it is this lack of complete correspondence between industries and commodities which presents problems in deriving 'pure' input–output tables (either commodity × commodity or industry × industry).

The absorption matrix records the inputs of commodities into industries. Most of these inputs are required to produce the characteristic product of the industry but some are required to produce its subsidiary products. In order to estimate the input structure of commodity j from the known input structure of industry j (i.e. in order to estimate a column of **A** from a known column of **B**) it is necessary to deduct from the inputs of industry j those inputs which are required for the production by industry j of commodities other than its principal or characteristic product, commodity j. It is also necessary to add in the inputs required for the production of commodity j in other industries. Typically, no information is available as to the allocation of inputs in an establishment or industry between the various commodities produced and it is, therefore, necessary to make some assumption about these input structures in order to derive a commodity × commodity table from the information in the absorption and the make matrices.

Two basic assumptions are possible and these are generally referred to as the commodity technology and industry technology assumptions. The former assumes that a commodity has the same input structure in whichever industry it is produced. The industry technology assumption on the other hand, assumes that all commodities produced by an industry are produced with the same input structure and thus commodities will have different input structures depending on the industry in which they are produced.

If we use the commodity technology assumption then the inputs into industry j comprise the weighted average of the inputs into each of the commodities which it produces, the weights being the proportions in which industry j produces the various commodities. We can thus write:

$$\mathbf{B} = \mathbf{A}_C\mathbf{C}$$

i.e.
$$\mathbf{A}_C = \mathbf{B}\mathbf{C}^{-1}. \tag{5.3.1}$$

This is the standard commodity technology solution illustrated by Example 1 in the Appendix where the inverse $\mathbf{C}^{-1}$ which has dimensions industry × commodity serves as a matrix of weights to convert matrix **B** which has dimensions commodity × industry into matrix **A**—the commodity × commodity matrix.

If, on the other hand, we use an industry technology assumption the inputs into commodity j will be the weighted average of the inputs into each of the industries which produces commodity j and the weights will be the market shares of each industry in the production of commodity j. We thus obtain:

$$\mathbf{A}_I = \mathbf{B}\mathbf{D}. \tag{5.3.2}$$

Matrix **D** here has the same dimensions and plays the same role as

$\mathbf{C}^{-1}$ in (5.3.2) above). This is illustrated by Example 2 of the Appendix to this Chapter.

It is not intended at this stage to discuss which of the two technology assumptions is likely to be the more appropriate to use in particular circumstances. It will, however, be useful to point out an important difference between the two solutions if matrix **A** is to be used in projection work. If we assume that there is no change in techniques of production over time a commodity technology assumption implies that matrix **A** will be stable and there are no complications. If, however, we are using an industry technology assumption, the lack of technical change will then imply that inputs into industries are stable. Matrix **B** will then be stable and in order to use matrix **A** in projection work it is necessary to know the details of matrix **D**, the market share matrix. If the market shares of industries change this will alter the weights in (5.3.2) and will thus alter matrix **A** even though the input structures of industries do not change.

Thus, when the industry technology assumption is used matrix **A** will be stable over time only if both **B** and **D** are stable. This can be put more generally be saying that in order to make an estimate of **A** for a future year it is necessary to project both **B** and **D** because the inputs into commodities are determined by the inputs into industries which produce them. The need to forecast output structures (i.e. where commodities are produced) as well as input structures will arise again in the section below when hybrid technology solutions are examined.

5.4 HYBRID TECHNOLOGY SOLUTIONS

The solutions suggested in section 5.3 above require an unnecessarily rigid approach; it is necessary to assume either that all subsidiary production has a commodity technology or that all subsidiary production has an industry technology. It seems reasonable to expect that some subsidiary production might fit a commodity technology assumption whilst for other elements of subsidiary production an industry technology assumption may be more appropriate. In general one would expect that most commodities have the same input structure wherever they are produced but, particularly where subsidiary production is made of by-products of an industrial process, the assumption of an industry technology may be more appropriate. In so far as an industry technology assumption is used it will be necessary to specify in which industries commodities are being produced before solving the model.

This mixture of technology assumptions was suggested by Gigantes and Matuszweski [21] and is incorporated in the U.N. System of National Accounts (U.N. [52]). In order that various elements of production can be treated on different assumptions it is necessary to split the make

matrix into two matrices, $\mathbf{M}_1$ and $\mathbf{M}_2$, where $\mathbf{M}_1$ includes those elements of production for which a commodity technology assumption is deemed appropriate and $\mathbf{M}_2$ includes those elements which are to be treated on an industry technology assumption. It will simplify the explanation if the latter are referred to as by-products although as we shall see later in section 5.5 it may well be desirable to include in $\mathbf{M}_2$ elements of subsidiary production other than by-products.

The basic accounting equation derived from the accounting framework in section 5.2 is:

$$\mathbf{q} = \mathbf{A}\mathbf{q} + \mathbf{f} \tag{5.4.1}$$

In order to allow for the hybrid technology assumption this must be written in expanded form:

$$\mathbf{q} = \mathbf{A}_1\mathbf{q}_1 + \mathbf{A}_2\mathbf{q}_2 + \mathbf{f} \tag{5.4.2}$$

where $\mathbf{q}_1$ and $\mathbf{q}_2$ are the commodity outputs produced respectively on the commodity and industry technology assumptions.

The derivation of $\mathbf{A}_1$ and $\mathbf{A}_2$ from **B** requires two sets of weights as was shown in equations (5.3.1) and (5.3.2). Given our technology assumptions we define our weights as:

$$\mathbf{C}_1 = \mathbf{M}_1\hat{\mathbf{g}}_1^{-1} \tag{5.4.3}$$

$$\mathbf{D}_2^* = \mathbf{M}_2\hat{\mathbf{q}}_2^{-1}. \tag{5.4.4}$$

We shall assume first that the production of by-products follows fixed market shares and we write:

$$\mathbf{g}_2 = \mathbf{M}_2\mathbf{i} = \mathbf{D}_2\mathbf{q} \tag{5.4.5}$$

where column i in $\mathbf{D}_2$ specifies the market share of each industry's by-products in the total production of commodity i. Equation (5.4.2) then becomes:

$$\mathbf{q} = \mathbf{B}\mathbf{C}_1^{-1}\mathbf{q}_1 + \mathbf{B}\mathbf{D}_2^*\mathbf{q}_2 + \mathbf{f}. \tag{5.4.6}$$

From (5.4.5) we can write:

$$\begin{aligned}\mathbf{q}_2 &= \mathbf{M}_2'\mathbf{i} = \hat{\mathbf{q}}\mathbf{D}_2'\mathbf{i}\\ &= \widehat{\mathbf{D}_2'\mathbf{i}}\mathbf{q}\end{aligned} \tag{5.4.7}$$

and combining (5.4.4) and (5.4.5) we have:

$$\begin{aligned}\mathbf{D}_2^*\mathbf{q}_2 &= \mathbf{M}_2\hat{\mathbf{q}}_2^{-1}\mathbf{q}_2\\ &= \mathbf{D}_2\hat{\mathbf{q}}\hat{\mathbf{q}}_2^{-1}\mathbf{q}_2\\ &= \mathbf{D}_2\mathbf{q}.\end{aligned} \tag{5.4.8}$$

If we now substitute (5.4.7) and (5.4.8) into (5.4.6) we have:

$$\begin{aligned}\mathbf{q} &= \mathbf{BC}_1^{-1}(\mathbf{q} - \mathbf{q}_2) + \mathbf{BD}_2\mathbf{q} + \mathbf{f} \\ &= \mathbf{BC}_1^{-1}(\mathbf{q} - \widehat{\mathbf{D}'_2\mathbf{iq}}) + \mathbf{BD}_2\mathbf{q} + \mathbf{f} \\ &= \mathbf{B}[\mathbf{C}_1^{-1}(\mathbf{I} - \widehat{\mathbf{D}'_2\mathbf{i}}) + \mathbf{D}_2]\mathbf{q} + \mathbf{f} \qquad (5.4.9)\end{aligned}$$

and comparing this with our basic accounting equation (5.4.1) we have:

$$\mathbf{A} = \mathbf{B}[\mathbf{C}_1^{-1}(\mathbf{I} - \widehat{\mathbf{D}'_2\mathbf{i}}) + \mathbf{D}_2]. \qquad (5.4.10)$$

It can be seen that if $\mathbf{M}_2 = \mathbf{0}$ then $\mathbf{D}_2 = \mathbf{0}$; also $\mathbf{M}_1 = \mathbf{M}$ and $\mathbf{C}_1 = \mathbf{C}$ and this solution becomes $\mathbf{A} = \mathbf{BC}^{-1}$ which is the simple commodity technology assumption. On the other hand if $\mathbf{M}_1 = \mathbf{0}$, then $\mathbf{C}_1 = \mathbf{0}$ and $\mathbf{D}_2 = \mathbf{D}$ and we have the industry technology solution $\mathbf{A} = \mathbf{BD}$. This result is the same as equation 3.16 in the S.N.A. (U.N. [52]) but its method of derivation is different. The solution in the S.N.A. makes an assumption about the relation between $\mathbf{g}_1$ and $\mathbf{q}_1$ ($\mathbf{g}_1 = \mathbf{C}_1^{-1}\mathbf{q}_1$) but no such assumption was necessary for the solution here. An output structure assumption such as that relating $\mathbf{g}_1$ to $\mathbf{q}_1$ is necessary if an industry technology assumption is used or if one is solving for **g** but is not required when solving for **q** on a commodity technology assumption.

A different solution is obtained if we assume that the output of by-products in $\mathbf{M}_2$ is linked to the outputs of the producing industries. In this case in place of (5.4.5) we have:

$$\mathbf{q}_2 = \mathbf{M}'_2\mathbf{i} = \mathbf{C}_2\mathbf{g} \qquad (5.4.11)$$

and from this we can obtain:

$$\begin{aligned}\mathbf{g}_2 &= \mathbf{M}_2\mathbf{i} = \hat{\mathbf{g}}\mathbf{C}'_2\mathbf{i} \\ &= \widehat{\mathbf{C}'_2\mathbf{i}}\mathbf{g}. \qquad (5.4.12)\end{aligned}$$

To find this solution we return to our accounting equation at the first row of (5.4.9)

$$\mathbf{q} = \mathbf{BC}_1^{-1}(\mathbf{q} - \mathbf{q}_2) + \mathbf{BD}_2\mathbf{q} + \mathbf{f}$$

but now with a different ouput structures assumption we use (5.4.11) and obtain:

$$\mathbf{q} = \mathbf{BC}_1^{-1}(\mathbf{q} - \mathbf{C}_2\mathbf{g}) + \mathbf{BD}_2\mathbf{q} + \mathbf{f}. \qquad (5.4.13)$$

To complete this solution we need to give the relation between **g** and **q**.

If we combine the relation $\mathbf{g}_1 = \mathbf{D}_1\mathbf{q}_1$ with (5.4.11) & (5.4.12) we have:

$$\begin{aligned}\mathbf{g} &= \mathbf{g}_1 + \mathbf{g}_2 \\ &= \mathbf{D}_1\mathbf{q}_1 + \widehat{\mathbf{C}'_2\mathbf{i}}\mathbf{g} \\ &= \mathbf{D}_1(\mathbf{q} - \mathbf{C}_2\mathbf{g}) + \widehat{\mathbf{C}'_2\mathbf{i}}\mathbf{g}\end{aligned}$$

which solving for **g** gives:

$$\mathbf{g} = (\mathbf{I} - \widehat{\mathbf{C}'_2\mathbf{i}} + \mathbf{D}_1\mathbf{C}_2)^{-1}\mathbf{D}_1\mathbf{q}. \qquad (5.4.14)$$

Following Gigantes we can write this as:

$$\mathbf{g} = \mathbf{Hq}$$

where **H** is the matrix in (5.4.14) relating **g** and **q**.

In this case (5.4.13) becomes:

$$\mathbf{q} = [\mathbf{BC}_1^{-1}(\mathbf{I} - \mathbf{C}_2\mathbf{H}) + \mathbf{BD}_2]\mathbf{q} + \mathbf{f}$$

and thus

$$\mathbf{A} = \mathbf{B}[\mathbf{C}_1^{-1}(\mathbf{I} - \mathbf{C}_2\mathbf{H}) + \mathbf{D}_2]. \qquad (5.4.15)$$

In this case, if $\mathbf{M}_2 = \mathbf{0}$, then $\mathbf{D}_2 = \mathbf{0}$ and $\mathbf{C}_2 = \mathbf{0}$; also $\mathbf{C}_1 = \mathbf{C}$ and we have the simple commodity technology solution $\mathbf{A} = \mathbf{BC}^{-1}$. On the other hand if $\mathbf{M}_1 = \mathbf{0}$, $\mathbf{D}_1 = \mathbf{0}$ and thus $\mathbf{H} = \mathbf{0}$, also $\mathbf{C}_1 = \mathbf{0}$ and $\mathbf{D}_2 = \mathbf{D}$ and we have the simple industry technology solution $\mathbf{A} = \mathbf{BD}$.

This result is somewhat different from that in the S.N.A. paragraphs 3.90 and 3.91. This second S.N.A. model seems, in fact, to be a simple industry technology solution applied to both $\mathbf{M}_1$ and $\mathbf{M}_2$. Different assumptions about output structures are used for $\mathbf{M}_1$ and $\mathbf{M}_2$ but by introducing a market share assumption for $\mathbf{M}_1$ an undesirable result has been obtained. This matrix, $\mathbf{M}_1$, is solved on a commodity technology assumption and no assumption about the relation between $\mathbf{q}_1$ and $\mathbf{g}_1$ is necessary to solve for **q** (see the equation (5.4.9) above). It would seem that the use of matrix $\mathbf{D}_1$ here has, in effect, produced an industry technology solution for $\mathbf{M}_1$. If $\mathbf{M}_2 = \mathbf{0}$, then $\mathbf{C}_2 = \mathbf{0}$; also $\mathbf{M}_1 = \mathbf{M}$ and $\mathbf{D}_1 = \mathbf{D}$ and S.N.A. equation 3.22 becomes $\mathbf{A} = \mathbf{BD}$ the simple industry technology assumption. On the other hand if $\mathbf{M}_1 = \mathbf{0}$, then $\mathbf{D}_1 = \mathbf{0}$ and the complex matrix **S** cannot be formed and thus no solution emerges. Gigantes [20] in fact does solve a simple industry technology model with a market share assumption for $\mathbf{M}_1$ and a product mix assumption for $\mathbf{M}_2$ and arrives at the same solution for **A** as in S.N.A. equation 3.22.[1]

Gigantes [20] also offers a solution of a mixed technology model but it seems possible that here also the matrices **C** and **D** have not played their role as intended. In Gigantes' model if $\mathbf{M}_2 = \mathbf{0}$ his solution reduces correctly to $\mathbf{A} = \mathbf{BC}^{-1}$—the commodity technology solution.[2] On the other hand if $\mathbf{M}_1 = \mathbf{0}$, then $\mathbf{M}_2 = \mathbf{M}$ and his solution should reduce to the industry technology solution $\mathbf{A} = \mathbf{BD}_2$; but it seems that no solution would be possible because his $\mathbf{A}$ will be zero.

It emerges from this discussion that the role of the matrices $\mathbf{C}$ and $\mathbf{D}$ must be clearly specified. There seems to be a danger of them being applied in such a way that they act as output structure assumptions when it is intended that they should be acting as weights on input structures. The test of an acceptable hybrid solution would seem to be that if $\mathbf{M}_1 = \mathbf{0}$ (i.e. no subsidiary production is being treated on a commodity technology assumption) then an industry technology solution should result, whereas if $\mathbf{M}_2 = \mathbf{0}$ a commodity technology solution should result. It is not clear that either the U.N. or Gigantes' solution meet this criterion, although my solutions (5.4.10) and (5.4.15) seem satisfactory.

5.5 CONSTRUCTION OF THE 1963 U.K. TABLES

When the data for the U.K. input–output tables for 1963 were being processed various methods of constructing the commodity × commodity tables from the make and absorption matrices were discussed.[3] It was decided that a hybrid technology assumption should be used. This approach is outlined in Ch. 3 of C.S.O. [46] and will now be described in somewhat more detail.

The solution which we followed was that given in the S.N.A. equation 3.16 and summarized in equation (5.4.10) above. Principal products and some subsidiary production were thus treated on a commodity technology assumption whilst the remainder of subsidiary production was treated on an industry technology assumption. The market share assumption about the output structures of this second group of commodities was used. It is possible to argue that the product-mix assumption might be more appropriate here. This would certainly be the case if this subsidiary production, which is treated on an industry technology, consists primarily of by-products whose outputs are tied to the outputs of the industries in which they are produced. However, in this application of this method (as indicated below) much of the production which was treated on an industry technology assumption was not simple by-products but what may be regarded as normal subsidiary production. Here a market share assumption seemed reasonable and, in fact, yielded a somewhat simpler solution than the alternative as given above in equation (5.4.15).

The technology assumption

The first stage in the application of this hybrid technology method to the 1963 U.K. tables was to split the make matrix into its two components

$\mathbf{M}_1$ and $\mathbf{M}_2$. In order to do this all the off-diagonal cells in $\mathbf{M}$ were examined and wherever necessary reference was made to individual Census reports to find out precisely which products were represented by various cells. There were some obvious cases of by-products. For instance, the production of Gas by the Coke ovens industry and the production of Miscellaneous chemicals by the Gas industry. In such cases the requirement of an industry technology was very clear but there were many cells where the choice was difficult, particularly with the limited technical knowledge at our disposal.

It must be remembered, however, that the seventy commodities classified in the tables are, in fact, not homogeneous commodities but are themselves collections of commodities, and in some cases quite diverse commodities may be classified together. This lack of complete homogeneity does in general present problems in input–output work but in this context it made our task a little simpler. There were a number of instances where the subsidiary production was a product whose inputs were not typical of the commodity group to which it is classified by the Standard Industrial Classification; in these cases the inputs structure of the particular product was often felt to be closer to that of the industry where it was produced and an industry technology assumption was used. A leading example of this is construction activity by the industry Industrial plant and steel work where it seemed likely that the 'product' would be closer to the industry in terms of its input structure than to the whole mass of construction with its large inputs of cement, building materials, glass and timber. Two other illustrations of our choice of industry technology are the production by the industry, Soap, oils and fats of the commodities, Other cereal foodstuffs and Other food. In the first case the product involved was animal feed and in the second, margarine. In both cases we felt that the input structure of these products would be close to that of the industry than the commodity. It is true that margarine is one of the principal products of the Other food industry but so are many other commodities with quite different input structures. As a result the input structure of margarine is not very similar to that of the Other food industry, whereas the principal products of the Soap, oil and fats industry will have an input structure fairly similar to that of margarine.

It was for reasons such as these that we allocated the various cells in the make matrix to either $\mathbf{M}_1$ or $\mathbf{M}_2$. There was, however, as is suggested in C.S.O. [46], a 'no-man's land' where the choice of technology was difficult. A particular example was the construction work of the Electricity industry which is the largest single off-diagonal cell in the make matrix. This work, largely work on 'mains and services', has an input structure rather different from the 'average' Construction commodity but, unlike the case of industrial plant mentioned above, the input structure is not like that of the producing industry. As it was necessary to choose either an industry or commodity technology assumption we chose the latter, as the lesser of the two evils, although when we came to carry out our final adjustments to the **A** matrix described below we effectively gave this element of production an input structure of its own which was like neither Construction nor Electricity. Many of the cells which it proved difficult to allocate were the subsidiary production of the engineering and metal goods industries. It was apparent that for many of these there was a broad similarity between their input structures and that whether we allocated them to $\mathbf{M}_1$ or $\mathbf{M}_2$ would make relatively little difference to the resulting **A** matrix. In many of these cases we selected an industry technology in order to avoid the possibility of a commodity technology approach producing a negative coefficient(s) in **A**. It is worth noting in passing that there is no need for the whole of a cell in the make matrix to be allocated to either $\mathbf{M}_1$ or $\mathbf{M}_2$. It is quite possible that an off-diagonal cell in **M** may be made up of more than one product and it may then be desirable to allocate part of the cell to $\mathbf{M}_1$ and part to $\mathbf{M}_2$.

It is not possible to present in this paper the details of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ in this application of the hybrid technology model to the U.K. data. All that can be done is to summarise the allocation; 55 per cent of subsidiary production was treated on an industry technology assumption and 45 per cent on a commodity technology.

The appearance of negative entries

When the first results of our calculations of the **A** matrix based on equation (5.4.10) were obtained there were some negative entries. This is a well-known feature of solutions of the **A** matrix based on the commodity technology solution and arises in the following way. As was stated in section 5.3 column j of **A** is made of (a) column j of **B** *plus* (b) the inputs into those units of commodity j made in other industries *minus* (c) the inputs into other commodities produced by industry j. In most cases the amounts to be deducted from cells in **B**, b_{ij}, are insignificant. However, if there is a sizeable input of commodity i into industry k there will be a similarly large input of i into commodity k. Now, if industry j produces commodity k as a significant proportion of its output the amount to be deducted, as at (c) above, from the cell b_{ij} in **B** may not be insignificant and it is quite possible for this deduction (c) to be greater than the value b_{ij} especially as the latter may be zero or very small. In such cases a_{ij} will be negative.

Before discussing how to remove these negative coefficients it will be useful to look at some of the reasons why negative cells appeared in our hybrid **A** matrix for 1963. The simplest reason is that our decision of the technology assumption was incorrect and in some cases the appearance

of negative coefficients led us to re-examine and sometimes to revise our allocation of cells between $\mathbf{M}_1$ and $\mathbf{M}_2$.

A second reason which was fairly common was that the particular piece of subsidiary production fitted neither an industry nor a commodity technology assumption. In our calculations the input of Printing (commodity 59) into Other food (commodity 9), $a_{59,9}$, was first estimated as -0.3 (per 1,000). This negative entry in A can be traced to the use of the commodity technology assumption for the production of commodity 69 (Distribution) by industry 9. There is quite a large input of commodity 59 into industry 69 but none into industry 9. Consequently when the calculation attempts to remove from $b_{59,9}$ an amount relevant to the production of commodity 59 by industry 9 a negative value, $-0{\cdot}3$, appears as the estimate for $a_{59,9}$. The reason for this is very probably that the commodity 69 produced by industry 9 (which is largely merchanting activities) does not require inputs of commodity 59 (Printing), although the typical commodity 69 does have significant inputs of printing. Despite this it seemed desirable to use a commodity technology here; the use of an industry technology would have resulted in significant inputs of agricultural produce into distribution which could not be accepted.

In a number of cases negative elements arose from the above reason because there was a large diagonal element in the **B** matrix. Thus, in the example quoted earlier, $i = k$; if industry j produces some commodity i and b_{ii} is large, b_{ij} may well turn to a negative a_{ij}. In many columns of **B** the largest entry is the diagonal entry and this gave rise to a number of negative values. Here again this could not be avoided as a commodity technology assumption was necessary. The largest negative cell which arose in **A** was -16 (per 1,000) for the input of Construction (commodity 62) into Electricity (commodity 64). As was mentioned earlier £105 mn. of Construction is produced by the Electricity industry. The value of the coefficient $b_{62,62}$ is 167 per 1,000 and thus the amount being deducted from cell 62, 64 in the absorption matrix, ($x_{62,64}$) to form $w_{62,64}$, the value in the commodity–commodity flow matrix, **W**, is $167/1{,}000 \times 105 =$ £18 mn. However the value of $x_{62,64}$ is only £2 mn. and hence the value of $w_{62,64}$ is estimated at £-16 mn. and $a_{62,64}$ is -16 per 1,000.[4]

A final possible reason for the occurrence of negative elements in **A** should be mentioned. It seems possible that in some cases firms when completing their returns may record only those inputs relevant to their principal production. In these circumstances negative cells may well appear when the inputs of subsidiary production are being transferred out of an industry if they were not recorded in the first place.

The removal of negative entries

Various techniques have been suggested for the removal of negative coefficients. The Cambridge Growth Project [11] adopted the working assumption that all negative entries be set to zero. Almon [2] has suggested a more sophisticated approach which would carry out the calculation of $\mathbf{A} = \mathbf{B}\mathbf{C}^{-1}$ by an iterative process; if negative entries arise at any stage of the calculation they are set to zero and the matrix is progressively balanced. This latter method is interesting and has been shown to converge fairly quickly but we chose what seemed to us to be the more practical method and one which, as we indicate below, gave us flexibility in our adjustment process. We followed basically Stone's suggestion [11] but with two differences; firstly, there were cells where a small positive entry in **B** became a negative entry in **A** and although we set the majority of negative coefficients to zero in some of these cases we replaced a negative entry by a small positive entry. Secondly where we removed a negative entry we made compensating adjustments in other entries in the matrix so that the overall row and column accounting constraints were still met.

In carrying through the adjustment process to remove the negative entries we found it convenient to work with the entries in the **W** matrix of flows rather than the coefficient matrix **A**. Our process of adjustment is best described by reference to the two negative entries mentioned in the previous sub-section: $w_{59,9}$ and $w_{62,64}$. In the first case, our first estimate of the cell $w_{59,9}$ was £-0.3 mn. and we adjusted this by setting it to zero. In deciding which other column should be involved in the adjustment in order to keep the row total of **W** unchanged we referred to the cell in the make matrix which had been treated on the commodity technology assumption and had given rise to the negative element. We then used the column pertaining to the commodity in that cell for the compensating adjustment. In this case the offending cell was $m_{9,69}$ and we reduced $w_{59,69}$ by £0.3 mn.

It then remained to balance these two columns. In general we looked for another commodity which appeared as an input into both the commodities (9 and 69). If possible, we tried to find an input commodity which our technology assumption had not caused to be sufficiently transferred from 69 to 9 in view of our impression of the peculiar input structure of this off-diagonal production. (The treatment described below of cell $w_{62,64}$ is a good example of this.) Alternatively we made our compensating adjustments to a row which had large inputs into both commodities 9 and 69 and thus one where we felt this adjustment could be absorbed, particularly if there was already a change between the cells in **B** and those in **A**.

The other type of adjustment process which we carried out is illustrated by the case of Construction (62) and Electricity (64) mentioned above. Here the commodity technology applied to the production of Construction by Electricity coupled with a large input of Construction into Construction produces a negative entry. The cell $x_{62,64}$ is £2 mn. and the first estimate of $w_{62,64}$ was £-16 mn. This was due to the cell of subsidiary production having an input structure unlike either industry or commodity. As a result too much Construction was transferred from column 64 to column 62 and in our adjustment process we transferred £17 mn. back from $w_{62,62}$ to $w_{62,64}$ thus setting $w_{62,64}$ = £1 mn. The exact size of the value chosen for this cell was an arbitrary compromise between 0 and the value of $x_{62,64}$. The former would have suggested that electricity can be produced without any inputs of construction (repair and maintenance) whereas the latter would have suggested that the production transferred out (construction work by the Electricity industry) has no input of Construction (sub-contracting). The size of this adjustment, £17 mn. positive to column 64 and negative to column 62 necessitated a similarly large adjustment to some other row(s) into these columns. In this case most of the adjustment was carried out in row 34, Insulated wires and cables. There is a significant input of these wires into the Electricity industry but a relatively small input into Construction. The use of commodity technology transfers only a small amount of wires but, in fact, much of the input in $x_{34,64}$ is related to the construction activity by the Electricity industry and not to its principal product, Electricity. We therefore made a large transfer from $w_{34,64}$ to $w_{34,62}$ which balanced the columns of our matrix and improved our estimates of the input structures of columns 62 and 64.

An alternative to this approach in general, would have been to allow the compensating adjustments in the rows to take place in the row of primary inputs. Whereas this would have been simpler, we would not have been able to allow for the peculiar input structures of some of the cells of subsidiary production such as construction work by the Electricity industry.

5.6 RESULTS OF VARIOUS TECHNOLOGY ASSUMPTIONS

Previously, in the above sections various technology assumptions were discussed as methods of calculating commodity × commodity tables. The aim of this section is to examine how great (or small) are the differences between tables calculated on different assumptions, where the main differences tend to be located and whether the extent of differences is dependent on the level of aggregation.

We start with matrix **B** and making the two extreme assumptions we obtain $\mathbf{A}_C$ and $\mathbf{A}_I$. These three matrices can be regarded as forming the corners of a triangle. Hybrid technology solutions, $\mathbf{A}_H$, will be compromises between the extremes $\mathbf{A}_C$ and $\mathbf{A}_I$ and can thus be regarded as lying within the triangle. The 'dimensions' of this triangle and also the location of $\mathbf{A}_H$ can be found by measuring the differences between the various matrices. This will show the overall extent to which different technology assumptions result in different commodity × commodity tables. It is desirable also to compare the matrices in value terms as well as in coefficient terms. There are thus 12 comparisons which can be made (assuming only one version of $\mathbf{A}_H$ is calculated).

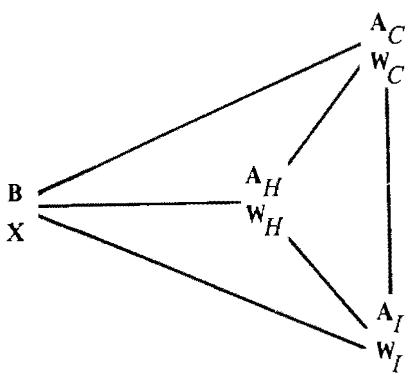

Figure 5.1 Technology Triangle

Table 5.3 Comparison of A_C and A_I: 70 × 70 table

Size of difference between the cells of the matrices (per thousand)	$\mathbf{A}_C$ *and* $\mathbf{A}_I$		$\mathbf{W}_C$ and $\mathbf{W}_I$	$\mathbf{A}_C$ *and* $\mathbf{A}_I$ *excluding small differences in* $\mathbf{A}$ *and in* $\mathbf{W}$[3]	
	(1)[1]	*(2)*[2]	*(3)*	*(4)*[1]	*(5)*[2]
Under 0.1	805	(797)	1473	—	—
0.2–	354	(353)	310	—	—
0.3–	548	(540)	460	—	—
0.6–	577	(569)	321	19	(18)
1.5–	216	(211)	129	44	(44)
2.5–	182	(175)	103	99	(96)
5.5–	94	(82)	36	72	(61)
10.5–	49	(33)	12	46	(30)
20.5–	24	(19)	6	23	(18)
50 and over	1	(1)	—	1	(1)
Total	2850	(2780)	2850	2850	(2780)
Mean	1.5	(1.3)	0.6	0.9	(0.7)

Notes: 1. All entries.
2. Excluding diagonal entries.
3. Any differences between $\mathbf{A}_C$ and $\mathbf{A}_I$ are ignored where the difference in the value of any cell between $\mathbf{W}_C$ and $\mathbf{W}_I$ is less than 1.

In order to explore the dimensions of this triangle the six matrices involved ($\mathbf{A}_C$, $\mathbf{A}_I$, $\mathbf{A}_H$, $\mathbf{W}_C$, $\mathbf{W}_I$ and $\mathbf{W}_H$) were calculated from the basic absorption and make matrices. The inputs of domestically produced inputs only were considered and imports were omitted. (Matrix $\mathbf{A}_H$ appears as the *d* columns in Table I in C.S.O. [46].) The basic matrices were aggregated from 70 × 70 to 35 × 35 and to 13 × 13[5] and the three **A** and **W** matrices were again calculated.

Table 5.4 Comparison of A_C and A_I: 35 × 35 table

Size of difference between the cells of the matrices (per thousand)	$\mathbf{A}_C$ *and* $\mathbf{A}_I$		$\mathbf{W}_C$ *and* $\mathbf{W}_I$	$\mathbf{A}_C$ *and* $\mathbf{A}_I$ *excluding small differences in* $\mathbf{A}$ *and in* $\mathbf{W}$[3]	
	(1)[1]	*(2)*[2]	*(3)*	*(4)*[1]	*(5)*[2]
Under 0.5	410	(407)	424	—	—
0.5–	200	(195)	197	61	(59)
1.5–	66	(65)	66	43	(42)
2.5–	70	(61)	66	50	(43)
5.5–	32	(24)	26	29	(22)
10.5–	17	(9)	15	17	(9)
20.5–	6	(5)	8	6	(5)
50 and over	1	(1)	—	1	(1)
Total	802	(767)	802	802	(767)
Mean	1.7	(1.5)	1.8	1.3	(1.1)

Notes: As in Table 5.3 above.

The results of comparing $\mathbf{A}_C$ and $\mathbf{A}_I$ are shown in Tables 5.3, 5.4 and 5.5 for the different levels of aggregation. These were obtained by subtracting one matrix from the other and then, ignoring the signs of these differences, a frequency distribution of differences as in Table 5.3 column 1 was obtained. These coefficient matrices are calculated as input per £1,000 of output and it can be seen that out of a total of 2,850 non-zero cells in the matrices in 2,284 cells (i.e. 80 per cent of the non-zero cells) the difference between $\mathbf{A}_I$ and $\mathbf{A}_C$ was no greater than 1. On the other hand there are 74 cells where the differences between $\mathbf{A}_I$ and $\mathbf{A}_C$ are greater than 10 and the mean difference is 1.5. In column 3 of Table 5.3 the differences between $\mathbf{W}_I$ and $\mathbf{W}_C$ are shown and here only 10 per cent of the non-zero cells show differences greater than 1, and the mean difference is 0.6. Before passing judgment on these differences they should be converted from absolute to relative differences. The average size of the non-zero cells in **X** is 7·0 so that mean relative difference between $\mathbf{W}_I$ and $\mathbf{W}_C$ is about 8 per cent[6] (see Table 5.6).

Table 5.5 Comparison of A_C and A_I: 13 × 13 table

Size of differences between the cells of the matrices (per thousand)	$\mathbf{A}_C$ *and* $\mathbf{A}_I$		$\mathbf{W}_C$ *and* $\mathbf{W}_I$	$\mathbf{A}_C$ *and* $\mathbf{A}_I$ *excluding small differences in* **A** *and in* **W**[3]	
	(1)[1]	*(2)*[2]	*(3)*	*(4)*[1]	*(5)*[2]
Under 0.5	68	(65)	41	—	—
0.5–	52	(52)	37	48	(48)
1.5–	13	(13)	18	13	(13)
2.5–	11	(8)	25	11	(8)
5.5–	9	(6)	20	9	(6)
10.5–	6	(2)	10	6	(2)
20.5–	1	(1)	12	1	(1)
50 and over	—	(—)	—	—	(—)
Total	160	(147)	163	160	(147)
Mean	1·9	(1.5)	5.3	1.9	(1.5)

Notes: As in Table 5.3 above.

It is obviously not necessary to pay much attention to differences of less than 1 (per 1,000) in the **A** matrices. Further some of the larger differences may represent cells where the differences in the respective **W** matrices are small. This will depend on the value of total commodity output in the column. For instance, where the commodity output is £200 mn., a difference of 5 in a cell in the **A** matrices will be a difference of only 1 in the **W** matrices. Such small differences in flows can reasonably

Table 5.6 Percentage Differences Between Various Flow Matrices

Matrices	*Size of matrices*		
	70 × 70	*35 × 35*	*13 × 13*
X and $\mathbf{W}_H$	4.5	3.4	2.2
X and $\mathbf{W}_C$	5·9	3.9	2.8
X and $\mathbf{W}_I$	6.1	4.3	3.2
$\mathbf{W}_C$ and $\mathbf{W}_I$	7.8	7.1	4.2
$\mathbf{W}_H$ and $\mathbf{W}_C$	4.7	2.9	1.8
$\mathbf{W}_I$ and $\mathbf{W}_H$	4.5	4.5	2.6
Mean cell value in **X**	7	25	126

Note: This table shows the mean difference between cells in the two matrices expressed as a percentage of the mean cell value (ignoring zero cells) of the base matrix. Thus the entry for $\mathbf{W}_C$ and $\mathbf{W}_I$ (70 × 70) is the mean difference, 0.6 rounded from 0.56, as shown in Table 5.3 as a percentage of the mean cell value in **X**, 7.2 (unrounded).

be overlooked as being within the margin of error in estimating many of the basic flows in **X**. Thus column 4 of Table 5.3 includes only those differences between $\mathbf{A}_C$ and $\mathbf{A}_I$ which are matched by differences greater than 1 between $\mathbf{W}_C$ and $\mathbf{W}_I$. The result of this is to omit many recorded differences in column 1 in the **A** matrices between 1 and 10 and to reduce the mean difference from 1.5 to 0.9.

One feature of the matrix of differences is that many of the larger differences occur in the diagonal entries. In the absorption matrix these entries are often among the largest in the columns and will be included in the inputs into subsidiary production when a commodity technology is used but not when an industry technology assumption is used. Some of these flows are, in fact, intra-firm flows and present valuation problems and a case can be made for omitting them. In columns 2 and 5 of Table 5.3 the differences between $\mathbf{A}_C$ and $\mathbf{A}_I$ are shown excluding the diagonal entries. Of these, 33 are greater than 5 and the mean difference falls to 1.3 if they are omitted.

The comparison between the commodity and industry technology calculations is made for more aggregated matrices in Tables 5.4 and 5.5. There is a slight increase in the mean difference between $\mathbf{A}_C$ and $\mathbf{A}_I$ in the smaller matrices. The number of non-zero cells in the 35 × 35 table is 800 and in the 13 × 13 table it is 160. The typical size of the coefficients in the smaller matrices will be much greater than in the 70 × 70 matrix so that the relative differences between $\mathbf{A}_C$ and $\mathbf{A}_I$ are much less in more aggregated matrices. The mean difference between $\mathbf{W}_C$ and $\mathbf{W}_I$ increases considerably in the smaller tables but so too does the average cell size in the base matrix.

Similar frequency distributions can be drawn up for the differences between all the other pairings of matrices and the results are summarised in Figures 5.2 and 5.3 and in Table 5.6. The triangles in Figure 5.2 show the mean difference between the coefficient matrices as calculated in Tables 5.3 to 5.5 columns 1 and 4. Figure 5.3 presents the same information for the differences between the flow matrices (as column 3 of Tables 5.3 to 5.5). In Table 5.6 the mean difference between the flow matrices has been expressed as a percentage of the mean cell size in the basic matrix, **X**.

The main points which emerge from a study of these results will now be summarised:

(i) The largest difference between coefficient matrices is between $\mathbf{A}_C$ and $\mathbf{A}_I$. The differences here are greater than the differences between **B** and either $\mathbf{A}_C$ and $\mathbf{A}_I$. Even though the dimensions of the table are being altered from commodity × industry to commodity × commodity this has less effect on the coefficients than does the choice of technology assumption.

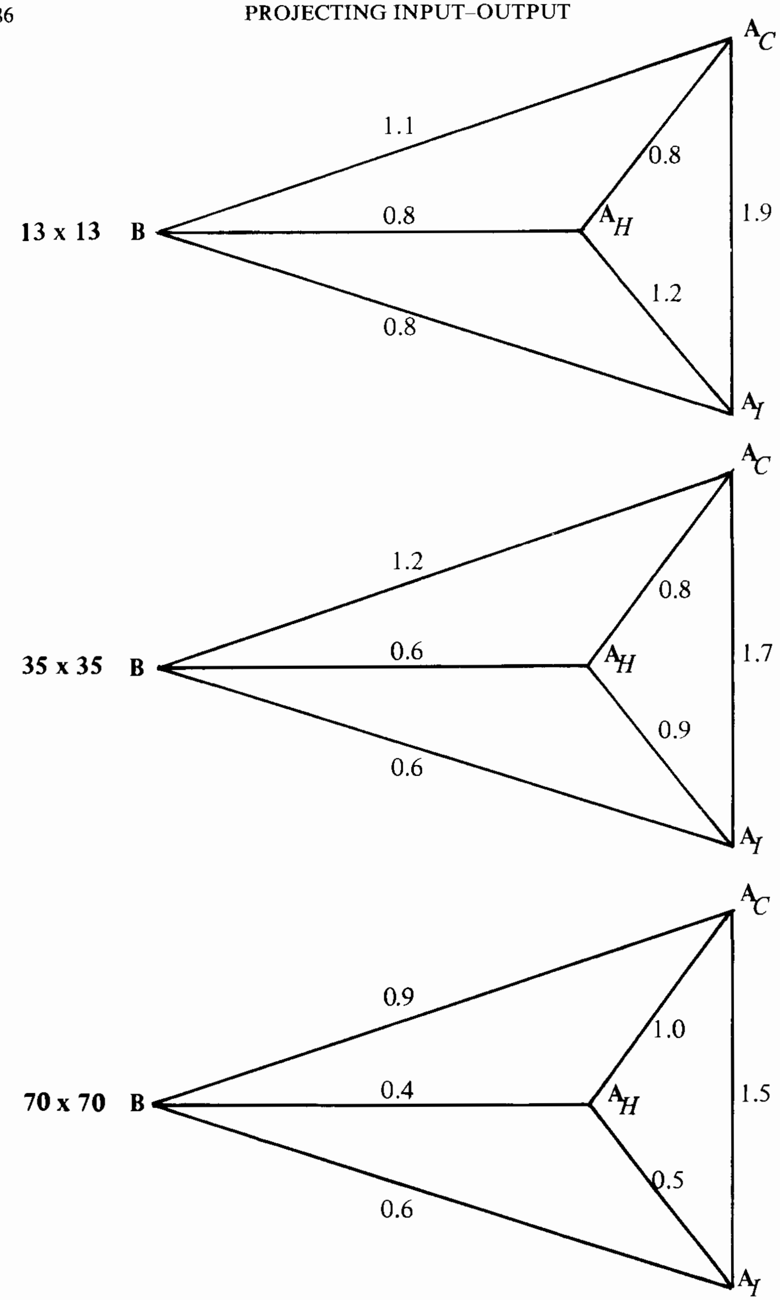

Figure 5.2 Mean Differences between Cells in Coefficient Matrices

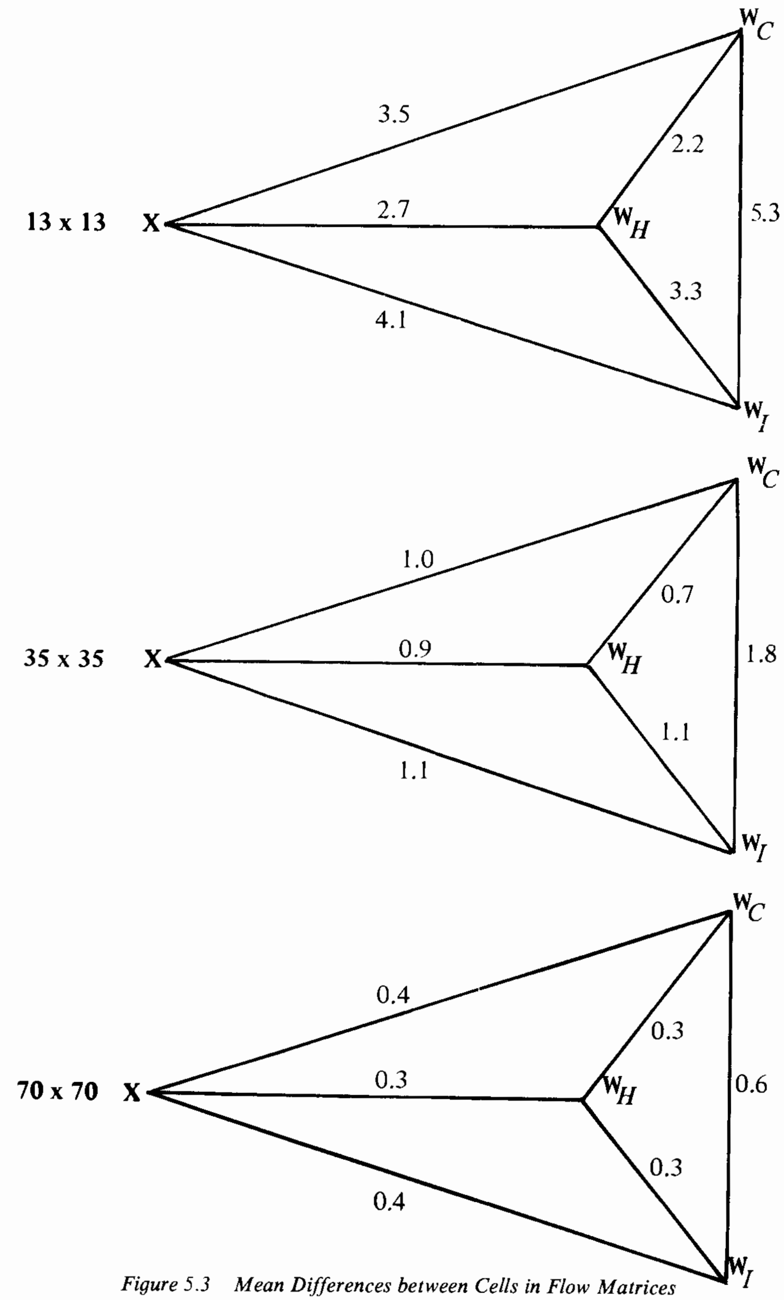

Figure 5.3 Mean Differences between Cells in Flow Matrices

(ii) The hybrid matrix, $\mathbf{A}_H$, in general lies somewhat nearer to $\mathbf{A}_C$ than to $\mathbf{A}_I$ although subsidiary production was divided roughly equally between the two technology assumptions in the 70 × 70 and in the smaller tables rather more was treated on the industry technology assumption. In so far as $\mathbf{A}_H$ is a weighted average of $\mathbf{A}_I$ and $\mathbf{A}_C$ the difference between $\mathbf{A}_I$ and $\mathbf{A}_C$ is roughly equal to the sum of the differences between $\mathbf{A}_H$ and each of the two extremes.

(iii) Of the three matrices, $\mathbf{A}_H$, $\mathbf{A}_I$ and $\mathbf{A}_C$, the hybrid matrix appears to lie closest to the base matrix **B**. Also $\mathbf{A}_H$ is closer to **B** than to $\mathbf{A}_I$ or $\mathbf{A}_C$.

(iv) The differences between the **A** matrices increase slightly as the matrices become more aggregated. The average size of the entries in the **A** matrix is much larger in the more aggregated tables than in the 70 × 70 tables and thus the relative differences between the **A** matrices are much less in the aggregated matrices.

(v) The mean difference between the flow matrices (**X** and **W**) is greater in the more aggregated tables (Figure 5.3) but the relative differences are much less when the tables are aggregated (Table 5.6). This is particularly noticeable when the tables are aggregated from 35 × 35 to 13 × 13.

The remainder of this section is devoted to a brief investigation into the reasons for the differences observed above. As was observed at the beginning of the Chapter differences only arise to the extent that industries record subsidiary production. If there was complete correspondence between industries and commodities the make matrix would be a diagonal matrix, $\mathbf{A} \equiv \mathbf{B}$ and $\mathbf{W} \equiv \mathbf{X}$ and the triangle would disappear. If the industry technology assumption is used for all subsidiary production then $\mathbf{A}_I$ will differ from **B** depending on the extent to which each commodity is produced in other than its own industry. On the other hand the difference between $\mathbf{A}_C$ and **B** will depend on the extent to which each industry produces other than its own commodity. Consequently the differences between $\mathbf{A}_C$ and $\mathbf{A}_I$ will depend on both the extent to which an industry produces other than its principal product and the extent to which a commodity is produced elsewhere. Further, it would be expected that the differences between **B** and $\mathbf{A}_H$ would be less than the differences between **B** and either $\mathbf{A}_C$ or $\mathbf{A}_I$. Some of the latter differences will arise where certain cells of subsidiary production are treated on the 'wrong' assumption; in $\mathbf{A}_H$, since it is possible to mix assumptions, these errors can be eradicated.

No reference has been made so far to the extent to which the columns of **B** differ from each other. The differences between the results obtained with the two technology assumptions depend partly on the amount of subsidiary production and partly on differences in the relevant columns

of **B**. If there is little difference between the input structures of industries in **B** then the choice of technology assumptions becomes much less important even if there is a considerable amount of subsidiary production.

In order to form an overall impression of the extent to which these two factors were important, the amount of subsidiary production was measured as an index. An index of non-exclusiveness was defined as the percentage of the total output of a commodity produced in other than its own industry, and was measured by the column sums of matrix **D** (excluding the diagonal entries). An index of non-specialisation was defined as the percentage of an industry's output accounted for by its subsidiary products and was measured as the column sums of the **C** matrix (excluding the diagonal entries). These two indices were added to form indices of subsidiary production for each of the 70 commodities. These indices ranged from zero for Other transport and communications to 45.3 for Agricultural machinery and 74.0 for Industrial engines. (In the latter case 41 per cent of the industry's output is non-principal products and 33 per cent of the total commodity output is produced in other industries.) The mean difference between $\mathbf{A}_C$ and $\mathbf{A}_I$ was calculated for each of the commodity columns and this was found to be closely correlated with the indices of subsidiary production. It was noticeable, however, that many of the commodities with high indices of subsidiary production did not have as large a mean difference between $\mathbf{A}_C$ and $\mathbf{A}_I$ as a straight linear relation would have suggested. The explanation of this could well be that the differences between the input structures of the commodities involved are quite small thus offsetting the effect of a high index of subsidiary production. As 10 out of 12 commodities involved are in the engineering and vehicles group, where there is a similarity of input structures between principal and subsidiary products, this would seem to be a reasonable explanation.

It would be interesting to extend this study by comparing the index of non-exclusiveness with the differences between **B** and $\mathbf{A}_I$ and also the index of non-specialisation with the differences between **B** and $\mathbf{A}_C$.

It was noted above that the effect of the choice of technology assumptions is less in smaller tables. This can be explained by two factors. Firstly the amount of subsidiary production is reduced by aggregation because where industries indulge in subsidiary production this tends to be in fields similar to their own and with which they will be combined when tables are aggregated. In the 70 × 70 table subsidiary production accounted for 5 per cent of the total but in the 13 × 13 table for only 2.6 per cent. Secondly in the larger tables there is more scope for the appearance of commodities with unusual input structures whereas in smaller tables one would expect more similarity between input structures.

5.7 CONCLUSIONS

This paper has shown how various commodity × commodity tables can be constructed from the commodity × industry absorption matrix and has examined the extent to which the various commodity × commodity tables differ from one another and from the absorption matrix. The extent of these differences depends partly on the amount of subsidiary production and partly on the differences between input structures in the columns of the absorption matrix. In the U.K. tables examined here the amount of subsidiary production was no more than 5 per cent. Further in many of the cases where subsidiary production was quite high there were relatively similar input structures in both the main and subsidiary producing industries.

The frequency distributions of the size of differences show that in the majority of cells the differences between the various matrices (whether in flow or coefficient form) are insignificant and it would seem that many of the cells of subsidiary production in the make matrix can be treated fairly arbitrarily. This is because they are a small proportion of their industry and commodity output and/or the input structures of their industry and commodity are similar. There are, however, a number of cells where the choice of technology is important and where different assumptions lead to quite large differences between the various matrices. Such cells can be treated with a hybrid technology model which allows some to be treated on a commodity technology assumption and some on an industry technology assumption. Closer examination of such cells suggests that some have an input structure which is like neither their industry nor their commodity. As it is unlikely that census data would be available on the inputs into even major subsidiary products the best that can be done is to make an independent estimate of the input structures of such cells, i.e. to apply neither an industry nor a commodity technology solution.

Finally, it should be emphasised that apart from this relatively small number of cells the differences between the cells in the various matrices are insignificant and that the triangles are small. The amount of subsidiary production is small and given the inevitable errors of measurement in the basic absorption matrix and even more so the margin of error attached to any projections, it seems somewhat doubtful whether more sophisticated models will yield significant advantages.

[1] There are two printing errors in S.N.A. paragraphs 3.90 and 3.91 (U.N. [52]). The second row of S.N.A. 3.22 should read

$$\mathbf{q} = [\mathbf{I} - \mathbf{B}(\mathbf{I} + \mathbf{D}_1\mathbf{C}_2 - \widehat{\mathbf{C}_2'\mathbf{i}})^{-1}\mathbf{D}_1]^{-1}\mathbf{e}$$

and the definition of **S** in the third line of para. 3.91 should be post-multiplied by $\mathbf{D}_1$.

[2] Gigantes defines $\mathbf{C}_1 = \mathbf{V}_1'\hat{\mathbf{g}}^{-1}$ whereas here $\mathbf{C}_1 = \mathbf{V}_1'\hat{\mathbf{g}}_1^{-1}$. He obtains in his (28) $\mathbf{A}_1 = \mathbf{B}[\mathbf{C}_1(\widehat{\mathbf{C}_1'\mathbf{i}})^{-1}]^{-1}$. It can be shown that this is the same as $\mathbf{BC}_1^{-1}$ when $\mathbf{C}_1$ is on my definition.

[3] The author was then a member of the Cambridge Department of Applied Economics and collaborated with D. C. Upton of the C.S.O. in the construction of these tables.

[4] The treatment here is somewhat simplified as it ignores the effect of any other elements of subsidiary production.

[5] The 35 sectors were those industries in the Cambridge Growth Project classification; the 13 sectors correspond to those distinguished in the examples in the S.N.A.

[6] By expressing 0.6 as a percentage of 7.0 we obtain a weighted average of the percentage differences between $\mathbf{W}_I$ and $\mathbf{W}_C$, weighted by the size of the cells.

APPENDIX TO CHAPTER 5

The various methods of estimating commodity × commodity tables described in the text are illustrated here by imaginary three-sector tables with *I* indicating Industries and *C* Commodities. The basic data are:

X: Absorption Matrix

$$\begin{array}{cc} & I \\ C & \begin{bmatrix} 10 & 60 & — \\ 40 & 60 & 20 \\ 20 & 30 & 60 \end{bmatrix} \end{array}$$

M: Make Matrix

$$\begin{array}{cc} & C \\ I & \begin{bmatrix} 90 & 10 & — \\ — & 280 & 20 \\ — & 10 & 190 \end{bmatrix} \end{array}$$

Industry outputs, $\mathbf{g} = \mathbf{Mi}$ = 100, 300, 200
Commodity outputs, $\mathbf{q} = \mathbf{M'i}$ = 90, 300, 210.

From these we derive the following:

$\mathbf{B} = \mathbf{X}\hat{\mathbf{g}}^{-1}$ (× 1,000)

$$\begin{array}{cc} & I \\ C & \begin{bmatrix} 100 & 200 & — \\ 400 & 200 & 100 \\ 200 & 100 & 300 \end{bmatrix} \end{array}$$

D: Market Shares

$$\begin{array}{cc} & C \\ I & \begin{bmatrix} 1.000 & 0.033 & \\ & 0.934 & 0.095 \\ & 0.033 & 0.905 \end{bmatrix} \end{array}$$

C: Product Mix

$$C\left\{\overbrace{\begin{array}{ccc} 0.90 & & \\ 0.10 & 0.94 & 0.05 \\ & 0.06 & 0.95 \end{array}}^{I}\right.$$

$\mathbf{C}^{-1}$

$$I\left\{\overbrace{\begin{array}{ccc} 1.111 & — & — \\ -0.118 & 1.068 & -0.056 \\ 0.007 & -0.068 & 1.056 \end{array}}^{C}\right.$$

A5.1 A MATRIX, COMMODITY TECHNOLOGY

In this case we have $\mathbf{A}_C = \mathbf{BC}^{-1}$ and this gives:

$$\mathbf{A}_C = C\left\{\overbrace{\begin{array}{ccc} 87 & 213 & -11 \\ 422 & 207 & 94 \\ 213 & 86 & 311 \end{array}}^{C}\right.$$

This can better be illustrated by calculating **B** from $\mathbf{A}_C$ and **C**. With a commodity technology assumption the inputs into industry j are the weighted average of the inputs into each of the commodities it produces, the weights being the industry's product mix, i.e. $\mathbf{B} = \mathbf{A}_C\mathbf{C}$. Thus for instance

$$b_{1,2} = (87, 213, -11) \times (0, 0.94, 0.06) = 200.$$

The negative entry in $\mathbf{A}_C$ arises because the commodity technology assumption is applied to the production of commodity 2 by industry 3. There is an input of commodity 1 into industry (and thus, commodity) 2 and the solution deducts from column 3 of **B** those inputs relevant to the production of commodity 2. There are, however, no inputs of commodity 1 into industry 3 in **B** and hence a negative entry appears in $\mathbf{A}_C$.

A5.2 A MATRIX, INDUSTRY TECHNOLOGY

This assumes that the input structure of commodities is determined by the industry in which they are produced and hence $\mathbf{A}_I = \mathbf{BD}$:

$$\mathbf{A}_I = C\left\{\overbrace{\begin{array}{ccc} 100 & 190 & 19 \\ 400 & 203 & 110 \\ 200 & 110 & 281 \end{array}}^{C}\right.$$

Here, the entry $a_{1,2}$ is calculated from $(100, 200, 0) \times (0.033, 0.934, 0.033) = 190$. The first column of $\mathbf{A}_I$ is the same as the first column of **B** because all commodity 1 is produced in industry 1. An entry appears at $a_{1,3}$ although $b_{1,3}$ was zero because some commodity 3 is produced in industry 2 and there is an input of commodity 1 into industry 2 in **B**.

CHAPTER 6

Ex ante as a Supplement or Alternative to RAS in Updating Input–Output Coefficients

W. HALDER FISHER

Battelle Memorial Institute, Columbus, Ohio, U.S.A.

6.1 INTRODUCTION

As all practitioners of input–output analysis are painfully aware, conventionally-derived input–output tables tend to be technologically out-dated by the time they are published. When the purpose for which such tables are used is historically analytical, no harm results; but when used for current decision making or for forecasting, their usefulness is seriously impaired. Because of this, much professional attention continues to be lavished on methods of updating input–output coefficients. Probably the most widely used such method—as a scrutiny of the present collection indicates—is RAS, the adjustment by double-proportionality mathematics of a past-year coefficients matrix to fit more recent intermediate input and output marginal values.

There are two admitted shortcomings to the RAS update: (1) the same orders of adjustments are applied to all cells in a given row or column, regardless of how many have, in fact, changed; and (2) no automatic provision can be made for emerging or disappearing interindustry markets. The first of these two shortcomings probably would be considered less serious than the second; but both introduce tabular distortions that should be reduced, if at all possible.

The *ex ante* approach to the updating or forecasting of input–output coefficients, developed at the Columbus Laboratories of the Battelle Memorial Institute, provides either a substitute for or a supplement to the RAS update, as well as a technically more advanced method of forecasting future coefficients.[1] This is an essentially Bayesian exercise that utilizes expert knowledge and judgment concerning industrial input or capital structures, instead of survey statistics, to generate the **A** matrix of a Leontief input–output table. The reader is referred to previous chapters of this book and other sources for a detailed description of this method, which will be only briefly discussed in this Chapter.[2]

It should be emphasised that the *ex ante* method was developed primarily for the purpose of *forecasting* rather than *updating* technical

coefficients. For this reason the resulting updated coefficients are normative and approximate rather than statistically precise; and they probably will not generate a transactions matrix that is closely conformed to the survey statistics of the target year. Nevertheless, they provide a better matrix for 'RAS-ing' than would be available otherwise. We will examine this aspect at some length.

General experience would seem to indicate that reasonably final survey statistics are available from the government's statistical agency—e.g. the Central Statistical Office or the Bureau of the Census—one or two years after the close of the subject year; but the time-lag for an input–output table is more likely to be around six or seven years. Thus, by the time that an input–output table for the year 1967 becomes available, data for a RAS-type update would be available for 1972 or 1973. It is quite likely, however, that significant technological change would have occurred in the productive methods of several industries during the 1967–72 interval. Such changes might even create new interindustry markets—i.e. change specific cells in the **A** matrix from zero to nonzero values—or eliminate old ones.

It is characteristic of the RAS process that every nonzero cell will be adjusted upward or downward, but no zero cell can be given a positive value, and no nonzero cell will be uniquely reduced to zero.[3] This can only be accomplished through the intervention of human judgment, that is, by an *ex ante* procedure. We propose therefore that, before the 'RAS-ing' takes place, the early-year coefficients be updated by the *ex ante* method; then RAS can be performed to achieve more precise statistical conformity with target year data.

6.2 PREPARATION FOR *EX ANTE*

The *ex ante* update is applied to the direct technical coefficients, not to the transaction values; therefore the first step must consist in constructing a complete column vector of coefficients for each sector. This is of course accomplished by dividing total input into every entry in the column of sector input transactions, including *value added*. In the *ex ante* exercise, the work is performed by the column, not by the row, and each sector's column of direct coefficients (the a_{ij}'s) should be entered into a separate columnar worksheet containing several blank columns to the right.

The researcher should be prepared to explain all the sector definitions and the input–output conventions to the expert (see below) during the interview, because the expert is much more likely to be an engineer or production-process oriented technical economist than he is a macroeconomist. If the table being updated is not a pure-technology, product-to-product table, the researcher should be prepared to explain to the expert all the entries which result from the transfer of secondary outputs. Otherwise, these entries, since they do not correspond to the true process inputs with which the expert will be familiar, are likely to cause much confusion during the interview.

6.3 SELECTING THE EXPERT

The expert is the key to a successful *ex ante* operation and must be selected with great care. He should combine several kinds of knowledge about the industry or group of industries under study. Our experience has been that when the national economy has been divided into 50 or more sectors, following the usual sectoral concepts, it will not be too difficult to find single experts who can deal with an entire sector. The greater the number of sectors, the easier it is to locate a single expert for each. On the other hand, when the economy is divided into about 20 sectors, it is extremely difficult to find single experts for each. Subsector experts will have to be used and the entire exercise becomes much more complicated.

The expert should be familiar with the sector from the input (or production) rather than from the output (or marketing) point of view. He should know what technological innovations are being studied in the laboratories and in the pilot plants, and he should know the rates at which they are emerging and diffusing into industry practice. If the expert is being used for updating—in contrast to forecasting—his knowledge need not go beyond actual sector practices; but for forecasting he must be able to anticipate the practices of the future.

In highly industrialized economies there are many places where experts can be found. Among the best (but by no means all) places to look are:

1. Broad-spectrum research institutions (such as Battelle)
2. The editors of trade periodicals
3. The faculties of technical and engineering schools
4. Large accounting or consulting firms that serve specific industries
5. Large investment houses (in the research departments)
6. Trade associations
7. Government agencies, especially in statistical, research, or planning areas
8. Industry, itself.

A word of warning is in order about selecting experts from the industry under study. All too often, industry personnel think too strictly in 'own-company' terms. This may adversely affect their contribution in one of two ways: Either they will (a) withhold information for fear of revealing trade secrets, or will (b) think and talk strictly in terms of what their company does, rather than what the industry or sector does.

6.4 THE INTERVIEW

The *ex ante* interview takes place in two phases: First, the researcher–interviewer briefs the expert on the problem under study, the input–output conventions and assumptions, and the sector definitions; then the expert is engaged in a dialogue-in-depth designed to provide replicable information concerning the technological changes that have taken place between the year originally described and the target year of the updating. (Of course, in a forecasting exercise the target year is in the future and some degree of replicability may be lost because of the necessity for judgment as well as knowledge.)

It may be advisable (if time and money budgets allow) for the two phases of the interview to be separated by an interval. This will allow the expert to refresh his knowledge and to bring together, often for transmission to the researcher, technical or statistical information concerning the changes that have occurred. This interval should be one of days, rather than weeks, however, because the expert may forget important elements of his briefing and revert to a more customary and less input–output-oriented way of thinking.

During the second part of the interview, the researcher should specifically focus the expert's attention on *every* cell in the column of coefficients. Nonzero entries should be examined to determine if they have changed during the update interval; and zero cells should be examined to determine if new interindustry markets have emerged. It is particularly important to identify emerging or disappearing markets. Next in importance is the estimation of changes in the ratio of *value added* to *total output*. And, third in importance is the estimation of major changes in the sizes of particular coefficients. The best possible approximations should be entered into the worksheet; but there is no need to balance the entire column. It is only necessary to get approximate coefficient values that bear the proper relationships to each other. The entire column can then be normalized (to sum to unity) by a simple program.

For sake of completeness and defensibility of information, it is best to obtain and record the reason for every significant change. In addition to notes taken by the researcher–interviewer, it is often advisable to tape-record the entire dialogue.

6.5 PREPARATION FOR 'RAS'

After the interview, the researcher normalizes the new column of coefficients so that they all, purchased inputs and value added, add to precisely unity. The updated coefficients can be reviewed with one or more other experts, if desired, and 'fine-tuned'; however, if the table is to be further adjusted by RAS, this is probably a redundant precaution.

In order to prepare them for the RAS computations, however, the column of coefficients must be converted back to money-transactions form and reimbedded into a complete intermediate transactions matrix. This is done *on a first-approximation basis* by multiplying the entire column of coefficients by the *original* value of the sector's total output. In other words, a column of values is obtained that shows what the original year's inputs would have been if the total output had been produced with the updated technology.

Obviously a better update will be obtained if *every* column is subjected to the above-described *ex ante* revision. However, if resources do not permit a full revision, it will still be beneficial to revise as many sectors in this manner as those resources can support.

After the conversion back to first approximation transactions values the intermediate matrix is subjected to a standard RAS updating. The coefficients implicit in the resulting transactions table will represent the best possible marriage of technological knowledge *and* statistical data.

[1] See W. Halder Fisher and Cecil H. Chilton, "Developing *ex ante* input-output flow and capital coefficients" in *Input-Output Techniques* (A. Bródy and A. P. Carter, Eds., North Holland, 1972) pp. 393-405.

[2] In addition to the above-cited paper, see the two following Battelle publications: *A Businessman's Introduction to Input-Output* (especially pp. 12-15) and *The Development Planner's Introduction to Input-Output* (especially pp. 25-29).

[3] Obviously, if the marginal sum of a row or column vector is changed to zero, *every* cell in that row or column will so change during RAS. But this is not the same thing as changing selected cells to zero. And this procedure is not symmetrical: changing a marginal value from zero to positive *would not* correspondingly alter the vector.

REFERENCES

[1] Allen, R. I. G., "Some experiments with the RAS method of updating input-output coefficients," *Oxford Bulletin of Economics and Statistics* Vol. 36, No. 3, 1974.

[2] Almon, C., "Investment in input-output models and the treatment of secondary products", in A. P. Carter and A. Brody (eds.), *Applications of Input-Output Analysis*, North-Holland, Amsterdam, 1970.

[3] Almon, C., "Recent methodological advances in input-output in the United States and Canada", paper presented to the Fourth International Conference on Input-Output Techniques, Geneva, January 1968.

[4] Arrow, K. J. and M. Hoffenberg, *A Time Series Analysis of Inter-Industry Demands*, North-Holland, Amsterdam, 1959.

[5] Bacharach, M., *Biproportional Matrices and Input-Output Change*, Cambridge University Press, Cambridge, 1970.

[6] Barker, T. S., "Social Accounting Matrices 1954-1968: the commodity accounts", Cambridge, Department of Applied Economics, Growth Project Paper 358, 1971.

[7] Beckerman, W., "The pattern of growth of output, employment and productivity, 1960 to 1975", in W. Beckerman and others, *The British Economy in 1975*, Cambridge University Press, Cambridge, 1965.

[8] Bénard, J., "Reseau des échanges internationaux et plannification ouverte", *Economie Appliquée*, Vol. 16, 1963.

[9] Bluet, J. C. and others, "Application d'un modèle gravitationnel aux flux d'échanges internationaux de véhicules automobiles", *Bulletin de CEPREL*, No. 7, 1966.

[10] Brauers, W. K., "The PES method and the variation of the current input coefficients", paper presented to the Econometric Society, Amsterdam, September 1968.

[11] Cambridge, Department of Applied Economics, *Input-Output Relationships 1954-1966*, Volume 3 of *A Programme for Growth*, Chapman and Hall, London, 1963.

[12] Cambridge, Department of Applied Economics, *The Demand for Fuel 1948-1975*, Volume 8 of *A Programme for Growth*, Chapman and Hall, London, 1968.

[13] Cambridge, Department of Applied Economics, *Exploring 1972*

with Special Reference to the Balance of Payments, Volume 9 of *A Programme for Growth*, Chapman and Hall, London, 1970.

[14] Deming, W. E. and F. F. Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *Annals of Mathematical Statistics*, Vol. 11, 1940.

[15] Evans, G. J. and R. M. Lindley, "The use of RAS and related models in manpower forecasting", *Economics of Planning*, forthcoming.

[16] Fontela, E. and others, "Forecasting technical coefficients and changes in relative prices", in A. P. Carter and A. Brody (eds), *Applications of Input-Output Analysis*, North-Holland, Amsterdam, 1970.

[17] Friedlander, D., "A technique for estimating a contingency table, given the marginal totals and some supplementary data", *Journal of the Royal Statistical Society*, Series A, Vol. 124, 1961.

[18] Geary, R. C., "A method of estimating the elements of an inter-industry matrix, knowing the row and column totals", *Economic and Social Review*, Vol. 4, No. 4, 1973.

[19] Ghosh, A., *Experiments with Input–Output Models*, Cambridge University Press, Cambridge, 1964.

[20] Gigantes, T., "The representation of technology in input–output models" in A. P. Carter and A. Brody (eds.), *Contributions to Input–Output Analysis*, North-Holland, Amsterdam, 1970.

[21] Gigantes, T. and T. Matuszewski, "Technology in input–output models", paper presented to the Fourth International Conference on Input–Output Techniques, Geneva, 1968.

[22] Grandville, O. de la, E. Fontela and A. Gavus, "A note on European projections of foreign trade", *Economia Internazionale*, Vol. 21, No. 4, 1968.

[23] Green, M. J., "Investment matrices for plant and machinery: 1963 and 1968", *Economic Trends*, No. 214, 1971.

[24] Henry, E. W., "A note on the logarithmic transformation of the RAS", *Economic and Social Review*, Vol. 4, No. 4, 1973.

[25] Henry, E. W., "Relative Efficiency of RAS versus least squares methods of updating input–output structures, as adjudged by application to Irish data", *Economic and Social Review*, Vol. 5, No. 1, 1973.

[26] Johansen, L., "Explorations in long-term projections for the Norwegian economy", *Economics of Planning*, Vol. 8, Nos. 1–2, 1968.

[27] Kouevi, A. F., "Essai d'application prospective de la méthode RAS au commerce international", *Bulletin du CEPREL*, No. 5, 1965.

[28] Lecomber, J. R. C., "A generalisation of RAS", Cambridge, Department of Applied Economics, Growth Project Paper 196, 1964.

[29] Lecomber, J. R. C., "RAS projections when two or more matrices are known", *Economics of Planning*, Vol. 9, No. 3, 1969

[30] Leontief, W. W., *The Structure of the American Economy, 1919–1929*, Harvard University Press, Cambridge, Massachusetts, 1941.

[31] Linnemann, H., *An Econometric Study of International Trade Flows*, North-Holland, Amsterdam, 1966.

[32] Matuszewski, T., P. R. Pitts and J. A. Sawyer, "Linear programming estimates of changes in input–output coefficients", *Canadian Journal of Economics and Political Science*, Vol. 30, No. 2, 1964.

[33] Omar, F. H., *The Projection of Input–Output Coefficients with Application to the United Kingdom*, University of Nottingham, Ph.D. thesis, 1967.

[34] Paelinck, J. and J. Waelbroeck, "Etude empirique sur l'évolution de coefficients 'input–output' ", *Economie Appliquée*, Vol. 16, No. 1, 1963.

[35] Schneider, H. M., *An Evaluation of Two Alternative Methods for Updating Input–Output Tables*, Harvard College, B.A. dissertation, 1965.

[36] Stephan, F. F., "An iterative method of adjusting sample frequency tables when expected marginal totals are known", *Annals of Mathematical Statistics*, Vol. 13, 1942.

[37] Stone, R., *Input–Output and National Accounts*, Organisation for Economic Cooperation and Development, Paris, 1961.

[38] Stone, R. and J. A. C Brown, "A long-term growth model for the British economy", in R. C. Geary (ed.), *Europe's Future in Figures*, North-Holland, Amsterdam, 1962.

[39] Stone, R. and C. S. Leicester, "An exercise in projecting industrial needs for labour", Cambridge, Department of Applied Economics, 1966.

[40] Theil, H., *Economics and Information Theory*, North-Holland, Amsterdam, 1967.

[41] Theil, H., and C. B. Tilanus, "The demand for production factors and the price sensitivity of input–output predictions", *International Economic Review*, Vol. 5, No. 3, 1964.

[42] Tilanus, C. B., *Input–Output Experiments, The Netherlands 1948–1961*, Rotterdam University Press, Rotterdam, 1966.

[43] United Kingdom, Central Statistical Office, *Annual Abstract of Statistics*, HMSO, London, published annually.

[44] United Kingdom, Central Statistical Office, "Commodity analysis of central government current expenditure on goods and services", *Economic Trends*, No. 214, 1971.

[45] United Kingdom, Central Statistical Office, *Input–Output Tables for the United Kingdom 1954*, Studies in Official Statistics No. 8, HMSO, London, 1961.

[46] United Kingdom, Central Statistical Office, *Input–Output Tables for the United Kingdom 1963*, Studies in Official Statistics No. 16, HMSO, London, 1970.
[47] United Kingdom, Central Statistical Office, *National Income and Expenditure*, HMSO, London, published annually.
[48] United Kingdom, Central Statistical Office, "Provisional input–output tables for 1968", *Economic Trends*, No. 207, 1971.
[49] United Kingdom, Central Statistical Office, "Provisional summary input–output tables for 1963", *Economic Trends*, No. 154, 1966.
[50] United Kingdom, Central Statistical Office, "Summary input–output tables for 1970", *Economic Trends*, No. 247, 1974.
[51] United Kingdom, Department of Economic Affairs, "Use of an input–output model in the preparation of the British National Plan", in United Nations, Economic Commission for Europe, *Macro-Economic Models for Planning and Policy Making*, United Nations, New York, 1967.
[52] United Nations, *A System of National Accounts*, Studies in Methods, Series F, No. 2, New York, 1968.
[53] Upton, D. C., "Expanded provisional input–output tables for 1963", *Economic Trends*, No. 178, 1968.
[54] Waelbroeck, J., "Une nouvelle méthode d'analyse des matrices d'échanges internationaux", *Cahiers Economique de Bruxelles*, No. 21, 1e trimestre, 1964.
[55] Waelbroeck, J., "Les échanges des pays du Marché Commun avec les pays du monde extérieur", in Collège de l'Europe, *Intégration Europeène et Réalité Economique*, Bruges, 1964.
[56] Watanabe, T. and S. Shishido, "Planning applications of the Leontief model in Japan", in A. P. Carter and A. Brody (eds.), *Applications of Input–Output Analysis*, North-Holland, Amsterdam, 1970.
[57] Wigley, K. H., "Production methods and time trends in input–output coefficients" in W. F. Gossling (ed.), *Input–Output in the United Kingdom*, Frank Cass, London, 1970.
[58] Woodward, V. H., "A 70 sector input–output table for 1954", Cambridge, Department of Applied Economics, Growth Project Paper 342, 1971.

INDEX

NOTES

Books of Similar Interest

CAPITAL COEFFICIENTS AND DYNAMIC INPUT-OUTPUT MODELS

Edited by Dr. W.F.Gossling

INPUT-OUTPUT AND THROUGHPUT: PROCEEDINGS OF THE 1971 NORWICH CONFERENCE

Edited by Dr. W.F.Gossling

PRODUCTIVITY TRENDS IN A SECTORAL MACRO-ECONOMIC MODEL

by Dr. W.F.Gossling

Distributed by Frank Cass & Co Ltd., 67 Great Russell Street, London, W.C.1., and International Scholarly Book Services Inc., P.O. Box 4347, Portland, Oregon, 97208, U.S.A.

★ ★ ★

Occasional Papers of the Input-Output Research Association

No. 1: COMBINATORIAL THEORY OF DEMAND
by S.N.Afriat

No. 2: SOME PRODUCTIVE CONSEQUENCES OF ENGELS' LAW
by W.F.Gossling

INPUT-OUTPUT PUBLISHING CO., 3 WYNDHAM PLACE, LONDON, W1H 1AP