MATRIX METHODS FOR ECONOMETRIC RESEARCH

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

PREFACE

There is a strikingly bad connection between matrix algebra as it is being presented in most textbooks and the vast complex of multidimensional problems, which characterises so much of modern econometrics.

To undergraduate students of econometrics, matrix algebra, therefore, seems to be all but an operational device.

An excellent book like Hadley's Linear Algebra, which is widely used by students of econometrics, suffers severely from a lack of applicability to econometric problems. It is primarily the absence of differential calculus, which causes the lack of applicability. Differential calculus can be fruitfully applied to (differentiable) scalar functions of matrices. It can also be applied to matrix functions of matrices, in problems involving Jacobians of transformations, marginal impacts of changes in instruments, asymptotic variances of reduced-form coefficient estimates, etc.

The absence or inadequate treatment of the Kronecker matrix product is also a case in point. Matrix sequences and series, convergence of series, don't receive the attention they deserve.

Some modern econometric textbooks like Goldberger's Econometric Theory or Johnston's Econometric Methods try to build a bridge between matrix theory and econometric problems by providing some elements of operational algebra. In a book like Anderson's An Introduction to Multivariate Statistical Analysis even more can be found.

The fascinating compactness of matrix notation is not made use of thoroughly. Far too often one starts with a matrix expression, breaks it down into elements, transforms these and assembles the results finally again into a matrix expression!

Bodewig in his Matrix Calculus has made some very laudable contributions to making matrix algebra operational. He applies differential calculus and treats matrices as matrices. Unfortunately, he felt compelled to revolutionize matrix notation far too much, so that many of his results are hidden. The present author owes a lot to Bodewig!

It is the purpose of this little booklet to collect some useful results in the fields mentioned above, to supplement them and to present the whole in a coherent way. It should be read in conjunction with a book like Hadley's or the matrix algebra part of Allen's Mathematical Economics.

As has been implicitly stated already, this little book is meant for undergraduate students of econometrics. It is not "mathematical", but written in a pedestrian way.

The author wants to thank professor Alan Walters and Madhu Kanbur of Birmingham University for reading the text.

H.N.

SYNOPSIS

This thesis consists of two parts. The first part is a self-contained survey of matrix methods relevant for econometric research. It consists of four chapters, the most important of which is the fourth. The other chapters are introductory and to a large extent standard. (There are perhaps two non-standard elements: the Kronecker matrix product which plays such a vital role in modern econometrics, and the matrix series). The fourth chapter deals with the interplay of matrices and differential calculus. Its aim is to suppress as much as is possible of the matrix elements when applying differential calculus to matrices. A series of basic theorems is being established.

A varied collection of examples illustrates the methods. They are taken from current econometric literature. A comparison of the procedures used in this thesis, and the traditional procedures is not carried out, in order to safe clarity.

The second part of the thesis consists of eight reprints of articles, (co-)written by the author . The majority of them are connected with matrix methods, some more than the others. Apart from matrix methods, they cover fields of research like consumption theory, business tests, input-output analysis, spatial equilibrium analysis.

C O N T E N T S

Preface	i
Introduction	iv
Chapter 1,	1
1.1 Matrices and vectors; basic definitions and notations	1
1.2 Operations	2
1.3 Partitioning	7
1.4 Special matrices	8
1.5 Determinant	9
1.6 Trace	12
1.7 The adjoint matrix and the inverse matrix. Cramer's rule	13
1.8 Elementary row and column operations	15
1.9 Examples and exercises	19
Chapter 2,	22
2.1 The characteristic equation, roots and vectors	22
2.2 Diagonalisation of matrices, powers of matrices, the Cayley-Hamilton theorem	25
2.3 Quadratic and bilinear forms, Cauchy's inequality, definiteness, rank	28
2.4 Examples and exercises	34
Chapter 3	40
3.1 Matrix sequences and series	40
3.2 Matrix norms	42
3.3 Matrix norms and convergence	43
3.4 Stochastic convergence	46
3.5 Examples and exercises	48
Chapter 4	51
4.1 Introduction	51
4.2 Definitions and assumptions, some basic formulae, scalar functions	51
4.3 Matrix functions	54
4.4 The scalar functions: trace, characteristic root, determinant, quadratic form, bilinear form	56
4.5 Some applications	58
4.6. Jacobians	62
4.7 Differences and approximation, Taylor's theorem	63
4.8 Optimization of scalar functions, Lagrange multipliers	64
4.9 Some applications	65
4.10 A sequel on the generalized inverse	75
4.11 Exercises	77
Literature	80
Index	82

I N T R O D U C T I O N

Although this booklet is meant to be read in conjunction with a book like Hadley's, it turned out to be indispensable to produce the basic definitions, notations, operations and results. Referencing e.g. in proofs would have become hopeless, otherwise.

In the first chapter the reader will find very basic things. Given the operational character of this booklet, vectors will not receive different treatment from matrices. One not too conventional thing is the Kronecker matrix product and another is the 'vec' notation.

A collection of examples and exercises round off this chapter.

In the second chapter the reader will find a summary exposition on characteristic equations, vectors and roots; diagonalisation of matrices; quadratic forms and rank. Also here a collection of examples and exercises will close the chapter.

The third chapter contains definitions of matrix sequences and series. There is also some information on upperbounds. A very short section on stochastic convergence concludes the main part of the chapter.

Purposely, the amount of proofs has been minimized. Most proofs can be found in the books mentioned in references. A combination of examples and exercises is added.

In the fourth chapter the reader will find an exposition on the application of differential calculus to matrices. Firstly, basic definition and notation will be given. The author is aware of the fact that some notations are open to criticism, like $\frac{\partial^2 f}{\partial G^2}$ for the matrix of second order partial derivatives of a scalar function f with respect to the elements g_{ij} of the matrix G. It might be better to use $\nabla^2_G f$. $\nabla_G f$ might be used for $\frac{\partial f}{\partial G}$, then.

Then follow some expositions on the derivatives matrix of matrix functions. The notation $\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$ will be introduced there.

The theorems developed there and before will be applied to some scalar functions; Jacobians will also be examined.

Optimization and approximation and a fair selection of problems from current econometric literature conclude the expository part of the fourth chapter, after a sequel on the generalized inverse.

A long list of examples and exercises, which might seem boring, closes chapter and book.

1.1. Matrices and vectors; basic definitions and notations

Given the purely operational approach of this book, matrices will be defined in the following way:

A matrix is an array of scalars, a_{ij} say, arranged as follows:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

(a_{ij} can be any ordinary number, but also a variable x, and even a function of it, say log x.)

The matrix written here has m×n elements, arranged in m rows and n columns.

If $m = 1$, the matrix reduces to $(a_{11} \cdots a_{1n})$,

If $n = 1$, the matrix reduces to $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}$

If both $m = 1$ and $n = 1$, the matrix reduces to (a_{11}) or, equivalently, to a_{11}.

An array like $(a_{11} \cdots a_{1n})$ is called a row vector, an array like $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}$ is called a column vector.

It is seen that a scalar is a special case of a matrix.

If the matrix has as many rows as columns ($m = n$), it is called square. If $m \neq n$, it is called rectangular

The convenience of matrix notation is that one dispenses with the constituent elements in many cases. Differential and difference calculus can be applied to them.

Matrices will be denoted by capital letters. It is a useful convention to use corresponding capital and lower case letters for the matrix and its elements.

So $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ will be denoted by A.

Vectors will be denoted by lower case letters. A correspondence between elements and letter used to denote the vector is conventional here, too.

A row vector like $(b_1 \cdots b_n)$ will be denoted by b'. A column

vector like $\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ will be denoted by c.

Useful, but not necessarily from a mathematical viewpoint, is the use of the prime ' in denoting a row vector.

Clearly, A can be visualized as an array of m row vectors, or as an array of n column vectors. The following notation will be frequently used in this book:

$\mathbf{A}_{i.}$, the i^{th} row of **A**,

$A_{.j}$, the j^{th} column of **A**.

In the example case there are m constituent row vectors: $\mathbf{A}_{1.}, \ldots\ldots, \mathbf{A}_{m.}$ and n constituent column vectors: $A_{.1}, \ldots\ldots, A_{.n}$

Obviously A can be expressed in its constituent (row or column) vectors.

Therefore $$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1.} \\ \vdots \\ \mathbf{A}_{m.} \end{bmatrix} = \begin{bmatrix} A_{.1} \ldots\ldots A_{.n} \end{bmatrix}$$

Instead of in elements, A is expressed in vectors. A further step is to express A in constituent matrices. (This will be done in the section on partitioning 1.3.)

In certain cases it will be impossible to retain the labelling convention connecting matrix and elements. E.g. in the case of a matrix product AB - to be defined in section 1.2 -. There the notation $(AB)_{ij}$ will be used to denote the ij^{th} element of **AB**.

In some cases, instead of a_{ij} the notation $(A)_{ij}$ will be used.

1.2. Operations

Given the definitions it is desirable to have rules for operation with matrices. Certain rules have been defined in the past, others have been added recently. It should be borne in mind that all rules for operations are determined by the nature of the problem to be attacked.

It is necessary firstly to define identity.

Identity. Two matrices, A and B, are identical if all their corresponding elements a_{ij} and b_{ij} are identical.

Addition. Two matrices, A and B, both having m rows and n columns, (= being of order (m,n)) are added by adding corresponding elements. One cannot add matrices of different orders. They have to be conformable as to addition.

Addition is commutative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, (1)

because $a_{ij} + b_{ij} = b_{ij} + a_{ij}$.

Addition is associative: $(A + B) + C = A + (B + C)$ (2)

because $(a_{ij} + b_{ij}) + c_{ij} = a_{ij} + (b_{ij} + c_{ij})$

Note: There is no need to define subtraction separately. Subtraction of B from A amounts to finding C such that $C + B = A$.

Multiplication

There are two multiplication procedures: ordinary multiplication (in brief: multiplication) and Kronecker multiplication

(Ordinary) multiplication. Two matrices A and B, A having order (m,n), B having order (n,p), if multiplied in this order, give a matrix AB, such that $(AB)_{ij} = \sum^{k} a_{ik}b_{kj}$.

A is called the premultiplier. B is the postmultiplier. From this it appears that multiplication is possible if and only if the premultiplier has as many columns as the postmultiplier has rows.

In the pattern of multiplication AB, A and B are conformable as to multiplication.

Note: In this case BA does not exist, because B has p columns, whereas A has m rows (and we assume $p \neq m$).

Multiplication is in general noncommutative '): $AB \neq BA$ (3)

because $\sum^{k} a_{ik}b_{kj} \neq \sum^{k} b_{ik}a_{kj}$.

If $AB = BA$, A and B are called permutable.

Multiplication is associative: $(AB)C = A(BC)$ (4)

Proof: Let A, B and C have orders (d,e), (e,f) and (f,g) respectively.

The mo^{th} element of AB: $(AB)_{mo} = \sum^{i} a_{mi}b_{io}$

The mp^{th} element of (AB)C is then:

$$\left\{(AB)C\right\}_{mp} = \sum^{o}\left\{\sum^{i} a_{mi}b_{io}\right\}c_{op} = \sum^{oi} a_{mi}b_{io}c_{op}$$

The np^{th} element of BC is: $(BC)_{np} = \sum^{s} b_{ns}c_{sp}$

So the mp^{th} element of A(BC) is:

$$\left\{A(BC)\right\}_{mp} = \sum^{n} a_{mn}\left\{\sum^{s} b_{ns}c_{sp}\right\} = \sum^{ns} a_{mn}b_{ns}c_{sp} = \sum^{oi} a_{mi}b_{io}c_{op}$$

The summation index n has been replaced by i, the summation index s has been replaced by o. This was legitimate, because both n and i assume values 1, ..., e and both s and o assume values 1, ..., f.

Multiplication is left distributive: $A(B+C) = AB + AC$ (5)

and right distributive: $(D+E)F = DF + EF$ (6)

because $\left\{A(B+C)\right\}_{mo} = \sum^{i} a_{mi}(b_{io} + c_{io}) =$

') Here as in future cases we assume that the matrices concerned are conformable as to multiplication.

$$= \sum_i a_{mi} b_{io} + \sum_i a_{mi} c_{io} = (AB)_{mo} + (AC)_{mo}, \text{ and}$$

$$\{(D + E)F\}_{mo} = \sum_i (d_{mi} + e_{mi}) f_{io} = \sum_i d_{mi} f_{io} + \sum_i e_{mi} f_{io} =$$

$$= (DF)_{mo} + (EF)_{mo}$$

Kronecker multiplication. Two matrices, A and B, A having order (m,n), B having order (p,q), if multiplied give a matrix

$$A \boxtimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}, \text{ of order } (m \times p,\ n \times q)$$

It is clear that Kronecker multiplication is always possible.

Kronecker multiplication is in general noncommutative:

$$A \boxtimes B \neq B \boxtimes A \qquad (7)$$

because e.g. $(A \boxtimes B)_{12} = a_{11}b_{12} \neq b_{11}a_{12} = (B \boxtimes A)_{12}$

Kronecker multiplication is associative

$$(A \boxtimes B) \boxtimes C = A \boxtimes (B \boxtimes C) \qquad (8)$$

Proof:

$$(A \boxtimes B) \boxtimes C = \begin{pmatrix} a_{11}b_{11}C & \cdots & a_{11}b_{1n}C & \cdots & a_{1s}b_{11}C & \cdots & a_{1s}b_{1n}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{11}b_{m1}C & \cdots & a_{11}b_{mn}C & \cdots & a_{1s}b_{m1}C & \cdots & a_{1s}b_{mn}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{t1}b_{11}C & \cdots & a_{t1}b_{1n}C & \cdots & a_{ts}b_{11}C & \cdots & a_{ts}b_{1n}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{t1}b_{m1}C & \cdots & a_{t1}b_{mn}C & \cdots & a_{ts}b_{m1}C & \cdots & a_{ts}b_{mn}C \end{pmatrix}$$

and $A \boxtimes (B \boxtimes C) = A \boxtimes \begin{pmatrix} b_{11}C & \cdots & b_{1n}C \\ \vdots & & \vdots \\ b_{m1}C & \cdots & b_{mn}C \end{pmatrix} =$

$$= \begin{pmatrix} a_{11}b_{11}C & \cdots & a_{11}b_{1n}C & \cdots & a_{1s}b_{11}C & \cdots & a_{1s}b_{1n}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{11}b_{m1}C & \cdots & a_{11}b_{mn}C & \cdots & a_{1s}b_{m1}C & \cdots & a_{1s}b_{mn}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{t1}b_{11}C & \cdots & a_{t1}b_{1n}C & \cdots & a_{ts}b_{11}C & \cdots & a_{ts}b_{1n}C \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{t1}b_{m1}C & \cdots & a_{t1}b_{mn}C & \cdots & a_{ts}b_{m1}C & \cdots & a_{ts}b_{mn}C \end{pmatrix}$$

A and B having orders (t,s) and (m,n) respectively.

Kronecker multiplication is left distributive:

$$A \boxtimes (B + C) = A \boxtimes B + A \boxtimes C \qquad (9)$$

and right distributive:

$$(D + E) \boxtimes F = D \boxtimes F + E \boxtimes F \qquad (9')$$

Proof: (for left distributivity only):

$$A \boxtimes (B + C) = \begin{pmatrix} a_{11}(B + C) & \cdots & a_{1s}(B + C) \\ \vdots & & \vdots \\ a_{t1}(B + C) & \cdots & a_{ts}(B + C) \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}B & \cdots & a_{1s}B \\ \vdots & & \vdots \\ a_{t1}B & \cdots & a_{ts}B \end{pmatrix} + \begin{pmatrix} a_{11}C & \cdots & a_{1s}C \\ \vdots & & \vdots \\ a_{t1}C & \cdots & a_{ts}C \end{pmatrix} = A \boxtimes B + A \boxtimes C$$

The following useful theorem can also be proved:

$$(A_1 \boxtimes B_1)(A_2 \boxtimes B_2) \cdots (A_n \boxtimes B_n) = (A_1 A_2 \cdots A_n) \boxtimes (B_1 B_2 \cdots B_n)$$

if $A_1 A_2 \cdots A_n$ and $B_1 B_2 \cdots B_n$ are defined. (10)

Proof: $(A \boxtimes B)(C \boxtimes D) =$

$$= \begin{pmatrix} a_{11}B & \cdots & a_{1g}B \\ \vdots & & \vdots \\ a_{f1}B & \cdots & a_{fg}B \end{pmatrix} \begin{pmatrix} c_{11}D & \cdots & c_{1k}D \\ \vdots & & \vdots \\ c_{g1}D & \cdots & c_{gk}D \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_j a_{1j}c_{j1}BD & \cdots & \sum_j a_{1j}c_{jk}BD \\ \vdots & & \vdots \\ \sum_j a_{fj}c_{j1}BD & \cdots & \sum_j a_{fj}c_{jk}BD \end{pmatrix} = (AC) \boxtimes (BD),$$

because $AC = \begin{pmatrix} \sum_j a_{1j}c_{j1} & \cdots & \sum_j a_{1j}c_{jk} \\ \vdots & & \vdots \\ \sum_j a_{fj}c_{j1} & \cdots & \sum_j a_{fj}c_{jk} \end{pmatrix}$

This proves the theorem.

Notes:

1. A special case of Kronecker multiplication is:

$$\alpha \boxtimes A = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1s} \\ \vdots & & \vdots \\ \alpha a_{t1} & \cdots & \alpha a_{ts} \end{pmatrix},$$

where α is a scalar.

Instead of $\alpha \boxtimes A$ one usually writes αA

2. Both multiplication procedures can be applied to vectors as well. Clearly $b \otimes c' = bc'$, b and c being column vectors. The following operations are useful:

$$(b'c)^2 = b'.cc'.b$$

$$b.c'Ad = b.d'A'c$$

$$b'Ac = \sum_{ij} a_{ij} b_i c_j$$

$$b'Ab = \sum_{ij} a_{ij} b_i b_j$$

A and b, c, d are a matrix and column vectors of appropriate order.

3. (Ordinary) multiplication of a matrix A - of order (m,n) - and a matrix B - of order (n,p) - can be expressed as multiplication of constituent row vectors of A and column vectors of B:

$$AB = \begin{pmatrix} A_{1.} \\ \vdots \\ A_{m.} \end{pmatrix} (B_{.1} \ldots B_{.p}) = \begin{pmatrix} A_{1.}B_{.1} & \ldots & A_{1.}B_{.p} \\ \vdots & & \vdots \\ A_{m.}B_{.1} & \ldots & A_{m.}B_{.p} \end{pmatrix}$$

4. A useful notation is:

$$\text{vec } A = \begin{pmatrix} A_{.1} \\ \vdots \\ A_{.n} \end{pmatrix}, \text{ if A has n columns.}$$

Transposition

Transposition of matrix A - of order (m,n) - is identical with writing down the matrix B - of order (n,m) - such that $b_{ij} = a_{ji}$. B is denoted by A' (read: A transpose). It is clear now why row vectors are denoted by a prime.
The following properties of transposition are important:

$$1.\ (A + B)' = A' + B' \qquad (11)$$

Proof: The ij^{th} element of A + B is $(A + B)_{ij} = a_{ij} + b_{ij}$. Transposition of A + B means giving $(A + B)_{ij}$ the ji^{th} position in the new array.
The ji^{th} element of A' is a_{ij},
the ji^{th} element of B' is b_{ij}.
So the transposed ij^{th} element of A + B is equal to the sum of the transposed ij^{th} elements of A and B.

$$2.\ (AB)' = B'A' \qquad (12)$$

Proof: The ij^{th} element of $AB = \sum_{k} a_{ik} b_{kj}$ will occupy the ji^{th} position in (AB)'. The ji^{th} element of B'A' is $(B'A')_{ji}$ =

$= \sum_{k}^{k} (B')_{jk}(A')_{ki} = \sum_{k}^{k} b_{kj}a_{ik}$. So the transposed ij^{th} element of AB is equal to the ji^{th} element of B'A'

$$3.\ (\mathbf{A} \otimes B)' = A' \otimes B' \qquad (13)$$

Proof: $(A \otimes B)' = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}' =$

$$= \begin{pmatrix} a_{11}B' & \dots & a_{m1}B' \\ \vdots & & \vdots \\ a_{1n}B' & \dots & a_{mn}B' \end{pmatrix} = A' \otimes B'$$

$$4.\ (A')' = A \qquad (14)$$

1.3. Partitioning

Multiplication of matrices can be sensibly simplified by using partitioning procedures. We can subdivide a matrix A - of order (m,n) - into $s \times t$ submatrices A_{ij} of order (m_i, n_j)

$(i = 1 \dots s;\ j = 1 \dots t)$

So $\mathbf{A} = \begin{pmatrix} A_{11} & \dots\dots & A_{1t} \\ \vdots & & \vdots \\ A_{s1} & \dots\dots & A_{st} \end{pmatrix}$

The submatrices A_{ij} have m_i rows, and n_j columns.

Clearly, $\sum_{i} m_i = m$, $\sum_{j} n_j = n$

Take now a matrix B of order (n,p), so that AB exists.

Partition B as far as its <u>rows</u> are concerned in the same way as A has been partitioned with respect to its <u>columns</u>.

$C = AB$ is then a partitioned matrix, whose submatrices C_{il} are sums of products of submatrices A_{ij} and B_{jl}:

$$C_{il} = \sum_{j}^{j} A_{ij}B_{jl}$$

This result is easily verified; the multiplication process is simply split up into its constituents.

If the partitioning has been done according to the above given rule, the submatrices can be handled as scalar elements in the multiplication process.

In section 1.2 it was shown that A can be written as $(A_{.1} \dots A_{.n})$, and that B can be written as $\begin{pmatrix} B_{1.} \\ \vdots \\ B_{n.} \end{pmatrix}$

This is clearly a form of partitioning.

Applying the rules of partitioned multiplication, one gets:

$$AB = (A_{\cdot 1} \ \ldots \ A_{\cdot n}) \begin{pmatrix} B_{1\cdot} \\ \vdots \\ B_{n\cdot} \end{pmatrix} = \sum_i A_{\cdot i} B_{i\cdot}$$

1.4. Special matrices

The unity matrix

The square matrix $\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & 0 \ddots & \vdots \\ \vdots & 0 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix}$ of order n, to be denoted by I_n has the property that any matrix **A** of order (n,m), if premultiplied by I_n is reproduced: $I_n A = A$.
Any matrix B of order (p,n), if postmultiplied by I_n, is reproduced: $BI_n = B$.
Also $I_n C = CI_n$, if C is square of order n.
It is a convention to delete the subscript n, if the order of I is clear from the context.

The null matrix

The matrix of order (m,n) consisting of zero elements only,

$$\begin{pmatrix} 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

to be denoted by 0, is called null matrix. It has the following properties:

$$A + 0 = 0 + A = A$$
$$A0 = 0A = 0 \text{ (orders of 0 as appropriate)}$$

Diagonal matrices

Any square matrix D of order n having elements $d_{ii} \neq 0$, and all $d_{ij} = 0 \quad (i,j = 1 \ldots n)$ is called diagonal.

Idempotent matrices

A (square) matrix **A**, such that $AA = A$, is called idempotent. (We may write this as $A^2 = A$, where A^2 is shorthand notation for AA. In general A ... A (n times) = A^n.)

Nilpotent matrices

A (square) matrix **A**, such that — for some integer $n \geq 1$ — $A^n = 0$, with $A^{n-1} \neq 0$, is called nilpotent.

Symmetric matrices

A (square) matrix such that $A' = A$ is called symmetric. It is obvious, but useful to know that $A + A'$ is symmetric.

Skew-symmetric matrices

A (square) matrix such that $A' = -A$ is called skew-symmetric. The diagonal elements are zero.

Proof: $(A)_{ii} = (-A')_{ii}$ by definition

$\therefore (A)_{ii} = -(A')_{ii}$

Also $(A)_{ii} = (A')_{ii}$,

so $-(A')_{ii} = (A')_{ii}$, which implies $(A')_{ii} = 0$.

1.5. Determinant

To a square matrix A there corresponds a scalar function $|A|$, called "determinant of A". If A is of order n, $|A|$ is called a determinant of order n.

It is convenient to define the determinant of an n^{th} order matrix in a recursive way:

For $n = 1$, $|A| = |a| = a$

For $n = 2$, $|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}$

For $n > 2$, the determinant can be defined in terms of determinants of $(n-1)^{th}$ order submatrices, in the following way:

First define the minor m_{ij} of the element a_{ij} in the matrix A of order n; it is the determinant of the $(n-1)^{th}$ order submatrix of A, obtained by deleting the i^{th} row and j^{th} column.

Then define the cofactor c_{ij} of the same element a_{ij}:

$c_{ij} = (-1)^{i+j} m_{ij}$

Finally $|A| = \sum_k a_{ik} c_{ik}$ for any i ($i = 1, \ldots n$).

It is intuitively clear that also

$|A| = \sum_i a_{ik} c_{ik}$ for any k ($k = 1, \ldots n$).

(This proposition is equivalent with $|A'| = |A|$. It will be proved in section 1.8.)

Determinants have, i.a., the following properties: ')

1. The determinant does not change if a multiple of one row is added to another row of a matrix. (15)

Proof: Using the recursive definition and choosing as row of reference any row other than the two mentioned, one gets eventually for the determinant of the amended matrix a

') Instead of "row" one can read "column" as well.

linear combination of 2^{nd} order minors:

$$\begin{vmatrix} a + \alpha d & b + \alpha e \\ d & e \end{vmatrix}, \quad \begin{vmatrix} a + \alpha d & c + \alpha f \\ d & f \end{vmatrix} \quad \text{and so on.}$$

These are equal to $\begin{vmatrix} a & b \\ d & e \end{vmatrix}, \begin{vmatrix} a & c \\ d & f \end{vmatrix}$ and so on.

An identical linear combination of these minors is the determinant of the original matrix.

2. The sum of the determinants of two matrices that differ only in the m^{th} row, is equal to the determinant of the matrix whose rows are identical with those of the original matrices except for the m^{th} row: this is equal to the sum of the m^{th} rows of the original matrices. (16)

Proof: Using the recursive definition and choosing as row of reference any but the m^{th} row, one can write the determinant of the compound matrix as a linear combination of minors like

$$\begin{vmatrix} a + c & b + d \\ e & f \end{vmatrix} = \begin{vmatrix} a & b \\ e & f \end{vmatrix} + \begin{vmatrix} c & d \\ e & f \end{vmatrix}$$

This proves the theorem.

3. If one row of a matrix A is multiplied by a scalar α, say, the determinant of the amended matrix is equal to $\alpha|A|$. (17)

Proof: Expansion according to the multiplied row will show it.

4. If two rows of a matrix A are interchanged, the determinant of the resulting matrix is equal to $-|A|$. (18)

Proof: Let the rows be $A_{i.}$ and $A_{j.}$.

Then $A = \begin{pmatrix} \vdots \\ A_{i.} \\ \vdots \\ A_{j.} \\ \vdots \end{pmatrix}$

Clearly $\begin{vmatrix} \vdots \\ A_{i.} \\ \vdots \\ -A_{j.} \\ \vdots \end{vmatrix} = -|A| = \begin{vmatrix} \vdots \\ A_{i.} - A_{j.} \\ \vdots \\ -A_{j.} \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ A_{i.} - A_{j.} \\ \vdots \\ -A_{j.} - (A_{i.} - A_{j.}) \\ \vdots \end{vmatrix} =$

$$= \begin{vmatrix} \vdots \\ A_{i.} - A_{j.} \\ \vdots \\ -A_{i.} \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ A_{i.} - A_{j.} - A_{i.} \\ \vdots \\ -A_{i.} \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ -A_{j.} \\ \vdots \\ -A_{i.} \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ A_{j.} \\ \vdots \\ A_{i.} \\ \vdots \end{vmatrix}$$

This proves the theorem.

5. A matrix A having two identical rows has zero determinant (19).

Proof: Interchanging these two rows does not change the determinant. According to (18) the determinant changes sign.

So $|A| = -|A|$, which implies $|A| = 0$.

6. $\sum^{k} a_{ik} c_{jk} = 0$ for $i \neq j$. (20)

In words: the sum of products of elements and cofactors corresponding to elements in the same column, but in a different row, vanishes.

Proof:

Expansion of $\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$ according to the j^{th} row

gives $\sum^{k} a_{jk} c_{jk}$.

If one substitutes the i^{th} row for the j^{th} row, one gets:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = 0$$

Expansion according to the new j^{th} row gives $\sum^{k} a_{ik} c_{jk}$, which should be 0, then.

7. If one row of a matrix A is a linear combination of other rows, the determinant of A is zero. (21)

Proof: Let $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{1\cdot} \\ \vdots \\ \mathbf{A}_{n\cdot} \end{pmatrix}$, such that

$$\mathbf{A}_{n\cdot} = \sum_{i=1}^{n-1} \alpha_i \mathbf{A}_{i\cdot}$$

Then $|A| = \begin{vmatrix} \mathbf{A}_{1\cdot} \\ \vdots \\ \mathbf{A}_{n-1\cdot} \\ \sum_{i=1}^{n-1} \alpha_i \mathbf{A}_{i\cdot} \end{vmatrix} = \begin{vmatrix} \mathbf{A}_{1\cdot} \\ \vdots \\ \mathbf{A}_{n-1\cdot} \\ \alpha_1 \mathbf{A}_{1\cdot} \end{vmatrix} + \ldots + \begin{vmatrix} \mathbf{A}_{1\cdot} \\ \vdots \\ \mathbf{A}_{n-1\cdot} \\ \alpha_{n-1} \mathbf{A}_{n-1\cdot} \end{vmatrix} =$

$$= \alpha_1 \begin{vmatrix} \mathbf{A}_{1.} \\ \vdots \\ \mathbf{A}_{n-1.} \\ \mathbf{A}_{n-1.} \end{vmatrix} + \ldots + \alpha_{n-1} \begin{vmatrix} \mathbf{A}_{1.} \\ \vdots \\ \mathbf{A}_{n-1.} \\ \mathbf{A}_{n-1.} \end{vmatrix} = 0,$$

using (16) and (17).

8. $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$, if **A** and B are square. (22)
The proof will be given in section 1.8.

9. $|\mathbf{A}'| = |\mathbf{A}|$ (23)
This proof will also be given in section 1.8.

10. $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n \cdot |\mathbf{B}|^m$, (24)
where **A** has order (m,m)
and B has order (n,n)
The proof will be given in chapter 2, section 1.

11. $|\alpha\mathbf{B}| = \alpha^n |\mathbf{B}|$ where B has order (n,n) (24')
Proof: Put $\mathbf{A} = \alpha$ in (24).

1.6. Trace

A second scalar function associated with a square matrix **A** is the trace of A, tr **A**, whose formal definition is:

$$\text{tr } \mathbf{A} = \sum_i a_{ii}$$

Some theorems on traces can be discovered easily.

1. tr (**A** + B) = tr **A** + tr B. (25)
The proof is obvious.

2. tr **AB** = tr **BA** (26)
Proof: $\text{tr } \mathbf{AB} = \sum_i (\mathbf{AB})_{ii} = \sum_{ik} a_{ik} b_{ki}$

$$\text{tr } \mathbf{BA} = \sum_j (\mathbf{BA})_{jj} = \sum_{jl} b_{jl} a_{lj} = \sum_{ik} a_{ik} b_{ki} = \text{tr } \mathbf{AB}$$

The summation index j has been replaced by k, the index l has been replaced by i. This is permitted because pairwise they have the same range.

3. tr **A'** = tr **A** (27)
Proof: $\text{tr } \mathbf{A}' = \sum_i (\mathbf{A}')_{ii} = \sum_i a_{ii} = \text{tr } \mathbf{A}$

4. tr (**A** ⊗ B) = (tr **A**) . (tr B), where **A** and B are square.') (28)

') Note that **A** ⊗ B can be square, even if A and B are not.

Proof:

$$\text{Let } \mathbf{A} \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}$$

So $\operatorname{tr}(\mathbf{A} \otimes B) = \operatorname{tr} a_{11}B + \ldots + \operatorname{tr} a_{nn}B =$

$= (a_{11} + \ldots + a_{nn}) \operatorname{tr} B = (\operatorname{tr}\mathbf{A}) \cdot (\operatorname{tr} B)$

5. $\operatorname{tr} \mathbf{AB} = (\operatorname{vec} \mathbf{A}')'(\operatorname{vec} B)$ (29)

where $\operatorname{vec} B = \begin{pmatrix} B_{\cdot 1} \\ \vdots \\ B_{\cdot m} \end{pmatrix}$

Proof: $\operatorname{tr} \mathbf{AB} = \sum_i \mathbf{A}_{i\cdot} B_{\cdot i} =$

$$= (A_{1\cdot} \ldots A_{m\cdot}) \begin{pmatrix} B_{\cdot 1} \\ \vdots \\ B_{\cdot m} \end{pmatrix} = \left[(A'_{\cdot 1})' \ldots (A'_{\cdot m})' \right] \begin{pmatrix} B_{\cdot 1} \\ \vdots \\ B_{\cdot m} \end{pmatrix} =$$

$= (\operatorname{vec} \mathbf{A}')'(\operatorname{vec} B).$

1.7. The adjoint matrix and the inverse matrix Cramer's rule

The adjoint matrix and the inverse matrix

If one substitutes for each element a_{ij} of a (square) matrix $\mathbf{A}$ its cofactor c_{ij}, one constructs a matrix $\tilde{\mathbf{A}}$ whose transpose $\tilde{\mathbf{A}}'$ is called the adjoint of $\mathbf{A}$.

$$\text{So } \tilde{\mathbf{A}}' = \begin{pmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & & \vdots \\ c_{1n} & \cdots & c_{nn} \end{pmatrix}$$

Clearly $\mathbf{A}\tilde{\mathbf{A}}' = \tilde{\mathbf{A}}'\mathbf{A} = |\mathbf{A}| I$, according to the definition of $|\mathbf{A}|$ and using (20).

Define then: $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \tilde{\mathbf{A}}'$

One easily sees then: $\mathbf{A}^{-1}\mathbf{A} = \frac{1}{|\mathbf{A}|} \tilde{\mathbf{A}}'\mathbf{A} = I$

$$\mathbf{A}\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \mathbf{A}\tilde{\mathbf{A}}' = I$$

$\mathbf{A}^{-1}$ is called the inverse of A. If $\mathbf{A}^{-1}$ contains infinite elements, owing to $|\mathbf{A}| = 0$, A will be said to have no inverse, or to be singular.

Properties of $\mathbf{A}^{-1}$

1. Uniqueness. There is no other matrix, B say, such that $\mathbf{AB} = I$. There is no other matrix, C say, such that $\mathbf{CA} = I$. (30)

Proof: $\mathbf{A}^{-1} = \mathbf{A}^{-1}I = \mathbf{A}^{-1}\mathbf{AB} = I B = B$ and

$\mathbf{A}^{-1} = I\mathbf{A}^{-1} = \mathbf{CAA}^{-1} = \mathbf{C}I = \mathbf{C}$;

AB and CA have been substituted for I.

2. $(AB)^{-1} = B^{-1}A^{-1}$ (31)

Proof: $(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = I$,
so $B^{-1}A^{-1} = (AB)^{-1}$

3. $(A^{-1})^{-1} = A$ (32)

Proof: Substitute $B = A^{-1}$ in (31),
so $(AA^{-1})^{-1} = (A^{-1})^{-1}A^{-1}$
or $I = (A^{-1})^{-1}A^{-1}$
and then $A = (A^{-1})^{-1}$

4. $(A')^{-1} = A^{-1\,\prime}$ (33)

Proof: $AA^{-1} = I$
so $A^{-1\,\prime}A' = I$ on transposing
which leads to $A^{-1\,\prime}A'(A')^{-1} = (A')^{-1}$
so $A^{-1\,\prime} = (A')^{-1}$

5. $|A^{-1}| = \frac{1}{|A|}$ (22')

Proof: $|AB| = |A|.|B|$ — see (22) —
implies $1 = |I| = |AA^{-1}| = |A| \cdot |A^{-1}|$
so $|A^{-1}| = \frac{1}{|A|}$

6. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (34)

Proof:

$$A^{-1} \otimes B^{-1} = \frac{1}{|A|} \begin{pmatrix} c_{11} & \cdots & c_{n1} \\ \vdots & & \vdots \\ c_{1n} & \cdots & c_{nn} \end{pmatrix} \otimes B^{-1} =$$

$$= \frac{1}{|A|} \begin{pmatrix} c_{11}B^{-1} & \cdots & c_{n1}B^{-1} \\ \vdots & & \vdots \\ c_{1n}B^{-1} & \cdots & c_{nn}B^{-1} \end{pmatrix}$$

Subsequently, $(A \otimes B)(A^{-1} \otimes B^{-1}) =$

$$= \frac{1}{|A|} \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix} \begin{pmatrix} c_{11}B^{-1} & \cdots & c_{n1}B^{-1} \\ \vdots & & \vdots \\ c_{1n}B^{-1} & \cdots & c_{nn}B^{-1} \end{pmatrix} =$$

$$= \frac{1}{|A|} \begin{pmatrix} |A|\,I & & \\ & \ddots & \\ & & |A|\,I \end{pmatrix} = \begin{pmatrix} I & & \\ & \ddots & \\ & & I \end{pmatrix}.$$

This proves the theorem.
Note: As usual c_{ij} stands for the cofactor of a_{ij}

Cramer's rule

An obvious application of matrix inversion is the solution of a system of linear equations.

One can rewrite the system

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= b_n \end{aligned}$$

as $Ax = b$, where $A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$

Clearly premultiplication of $Ax = b$ by A^{-1} will give $x = A^{-1}b$, so x is expressed in b.
Using the definition $A^{-1} = \frac{1}{|A|}\tilde{A}$, one gets $x = \frac{1}{|A|}\tilde{A}\,b$,

so $x_i = \frac{1}{|A|}(\tilde{A}\,b)_i = \frac{1}{|A|}(\tilde{A})_{i.}\,b =$

$$= \frac{1}{|A|}(c_{1i} \ldots c_{ni}) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \frac{1}{|A|}\sum_j^{} c_{ji}b_j =$$

$$= \frac{1}{|A|} \begin{vmatrix} a_{11} & \ldots & a_{1,i-1} & b_1 & a_{1,i+1} & \ldots & a_{1n} \\ \vdots & & & & & & \vdots \\ a_{n1} & \ldots & a_{n,i-1} & b_n & a_{n,i+1} & \ldots & a_{nn} \end{vmatrix}; \qquad (30')$$

the last step becomes clear by expanding the determinant, shown there, with respect to the n^{th} column.

1.8. Elementary row and column operations

When studying properties of determinants in section 1.5, we noticed one operation which did not change the determinant value:
1. Addition of a multiple of one row to another;
and two operations which changed the determinant in an easy way:
2. Multiplication of a row by a scalar,
3. Interchange of two rows.
These three elementary row operations can be represented by matrices, unity matrices upon which these operations have been performed. Let us use matrices of order $(2,2)$ to show this.

Thus premultiplication by $\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ amounts to adding α times the second row to the first,

Premultiplication by $\begin{pmatrix} 1 & 0 \\ 0 & \beta \end{pmatrix}$ amounts to multiplying the second row by β,

premultiplication by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ amounts to interchanging the first and second rows:

$$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + \alpha a_{21} & a_{12} + \alpha a_{22} \\ a_{21} & a_{22} \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ \beta a_{21} & \beta a_{22} \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{pmatrix}$$

These matrices can be generalized for order n. We define

$$E_{ij} = \begin{pmatrix} & & & \\ & & 1 & \\ & & & \end{pmatrix}$$

the unity occupying the ij^{th} position. Clearly $E_{ij}A = B$, such that $B_{i.} = A_{j.}$ and all other rows of B zero.
Then $\alpha E_{ij} + I$ $(i \neq j)$ represents the first elementary row operation: the addition of α times the j^{th} row to the i^{th} row;
$(\alpha-1)E_{ii}+I$ represents the second elementary row operation: the multiplication of the i^{th} row by α;
$E_{ij} + E_{ji} - E_{ii} - E_{jj} + I$ represents the third elementary row operation: the interchange of i^{th} and j^{th} rows.
The determinants of these matrices, which represent these three elementary row operations, are: 1, β and -1.
Elementary <u>column</u> operations and associated matrices can be developed accordingly. (There is no difference in appearance between "row-transformation" matrices and "column-transformation" matrices.)
Clearly $AE_{ij} = C$, such that $C_{.j} = A_{.i}$ and all other columns of C zero. The above given matrices, if used as postmultipliers, will add α times the i^{th} column to the j^{th} column, multiply the i^{th} column by α, interchange i^{th} and j^{th} columns, respectively. These matrices can also be applied to <u>rectangular</u> matrices of appropriate order. They have the same effect.
It is easy to see that any square matrix A, whose rows are linearly independent (that means: no row is a linear combination of other rows), can be reduced to a diagonal matrix by a finite sequence of elementary row operations of the first kind. In other words, A can be made diagonal by repeated premultiplication by transformation matrices of the first kind, to be denoted by E_i; $E_m \; E_{m-1} \ldots\ldots E_1 A = L$, where L is a diagonal matrix, the subscripts are labelling indexes.
A can be "diagonalised" as well by a finite sequence of elementary column operations, which can be represented by transformation matrices F_j; $AF_1 \ldots\ldots F_{n-1} F_n = K$, K again diagonal.

If the rows of A are not linearly independent, a <u>finite combination</u> of elementary row and column operations will diagonalise A: $E_m \ldots\ldots E_1 A F_1 \ldots\ldots F_n = M$, M being diagonal. M will contain some zero diagonal elements.
It follows easily that any nonsingular matrix (a matrix having non-zero determinant) can be reduced to a unity matrix of the same order, by a finite sequence of elementary row operations of first and second kind. (A nonsingular matrix has linearly independent rows, because a matrix with linearly dependent rows is singular.)
This provides us with a straightforward inversion procedure. If the same sequence of elementary row operations is performed on the unity matrix, this will give the inverse:
If $BA = I$, where B is the condensed transformation matrix incorporating the elementary row operations of the first and second kind, and A is the matrix that has to be inverted, then $BI = A^{-1}$.
The reader is now in a position to prove theorems (22) and (23).

<u>Proof of theorem (22)</u>:

$$|A| = |EAF| = |L| = \prod_i \alpha_i,$$

where A has order p,
where E stands for a product of transformation matrices, each of which represents an elementary <u>row</u> operation of the first kind,
where F stands for a product of transformation matrices, each of which represents an elementary <u>column</u> operation of the first kind
and L is a diagonal matrix with p diagonal elements α_i, having determinant $\prod_i \alpha_i$.

Clearly $AB = AFF^{-1}B$,
so $|AB| = |AFF^{-1}B| = |EAFF^{-1}B| = |LF^{-1}B| = \begin{vmatrix} \alpha_1 (F^{-1}B)_{1.} \\ \vdots \\ \alpha_p (F^{-1}B)_{p.} \end{vmatrix} =$

$$= (\prod_i \alpha_i)|F^{-1}B| = (\prod_i \alpha_i)|B| = |A| \cdot |B|,$$

which proves the theorem.
Use has been made of the fact that elementary row operations (1^{st} kind) don't affect the determinant and that multiplication of a row by a scalar α_i results in multiplying the determinant by α_i. Clearly, F^{-1} can be interpreted as a transformation matrix representing an elementary row operation.

Proof of theorem (23):

Let EAF = L,
Then $|A| = |EAF| = |L| = \prod_i \alpha_i$
Also, F'A'E' = L,
and $|A'| = |F'A'E'| = |L| = \prod_i \alpha_i$
This proves the theorem.

1.9. Examples and exercises

1. Prove that a matrix A postmultiplied by a column vector b results in a column vector. Express the elements $(Ab)_i$ in the elements of A and b.

2. Prove that a matrix A premultiplied by a row vector c' results in a row vector. Express the elements of c'A in the elements of A and c'.

3. Prove that a row vector a' postmultiplied by a column vector b results in a scalar.

4. Prove that a row vector a' premultiplied by a column vector c results in a matrix. (let a' have order $(1,n)$ and c have order $(m,1)$)

5. Prove that (ordinary) multiplication and Kronecker multiplication are identical in case 4 and not in case 3:

$$a' \boxtimes b \neq a'b$$
$$c \boxtimes a' = ca'$$

6. The product CD for

$$C = \begin{pmatrix} 5 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 8 & 3 \\ 0 & 0 & 5 & 2 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 1 & -2 & 0 & 0 \\ -2 & 5 & 0 & 0 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & -5 & 8 \end{pmatrix}$$

can easily be determined by partitioned multiplication.

$$CD = \left(\begin{array}{cc:cc} 5 & 2 & 0 & \\ 2 & 1 & & \\ \hdashline 0 & & 8 & 3 \\ & & 5 & 2 \end{array}\right)\left(\begin{array}{cc:cc} 1 & -2 & & \\ -2 & 5 & 0 & \\ \hdashline & 0 & 2 & -3 \\ & & -5 & 8 \end{array}\right) =$$

$$= \left(\begin{array}{cc:cc} \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix} & & 0 & \\ \hdashline & & & \\ 0 & & \begin{pmatrix} 8 & 3 \\ 5 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ -5 & 8 \end{pmatrix} & \end{array}\right) = \left(\begin{array}{c:c} I & 0 \\ \hdashline 0 & I \end{array}\right)$$

7. Find an expression for A^n, if

$$A = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_m \end{pmatrix}$$

8. Find the equations for

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$ to be idempotent.

9. Find the 'easiest' equations for

$B=\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ to be nilpotent.

10. If $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 3 & 9 \end{pmatrix}$, find tr A.
Also find tr $(A \otimes A)$.

11. Find $|A|$ and $|A \otimes A|$ for the previous A.

12. Prove: $x'y = \text{tr } yx'$,
x and y being column vectors.
Use: $\alpha = \text{tr } \alpha$ in case α is a scalar.

13. Prove: $BAA' = CAA'$ implies $BA = CA$.
B and C are (n,m) matrices, A is of order $(m,1)$ $(m > 1)$
Hint: Construct $(BAA' - CAA')(B-C)'$

14. Let A be a symmetric matrix, and B a skew-symmetric one, both of order m. Prove then that AB = BA implies that AB is skew-symmetric.

15. Prove that skew-symmetric matrices of <u>odd</u> order have zero determinants.

16. Solve the equation system

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 4 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

applying Cramer's rule.

17. Prove: $\text{tr } AE_{ij} = a_{ji}$.

18. Prove: $E_{ij}E_{kl} = 0$, if $j \neq k$. $E_{ij}E_{jl} = E_{il}$.

19. Prove: $|\alpha E_{ij} + I| = 1$, if $i \neq j$.

20. Prove: $|(\alpha-1)E_{ii} + I| = \alpha$.

21. Show that AB = 0 implies
1. A = 0 or B = 0
or 2. $|A| = |B| = 0$

22. Determine the inverse of
$\begin{pmatrix} A & o \\ b' & 1 \end{pmatrix}$ by using partitioned multiplication. Here o is a column vector of zeros.

23. Invert $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 3 & 9 \end{pmatrix}$ by elementary row operations.

24. Prove: tr PQ = o, if Q is symmetric and P is skew-symmetric.

25. Prove that $\begin{vmatrix} A & B \\ O & C \end{vmatrix} = |A| \cdot |C|$,

A, B, C being square of order m.

Hint:

Write $\begin{pmatrix} A & B \\ O & C \end{pmatrix} = E_{11} \otimes A + E_{12} \otimes B + E_{22} \otimes C$

Assume $|A| \neq o$. Is the result also true in case $|A| = o$?

26. $E(x_i - Ex_i)(x_j - Ex_j)$ is the covariance of (scalar) random variables x_i and x_j, $E(x_i - Ex_i)^2$ is the variance of x_i. $V = E\left[(x - Ex)(x - Ex)'\right]$ is the variance matrix of the column vector of random variables

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

(E is the expectation operator.) Express the variance of a'x (a linear combination of x_i) in terms of V.

Answer: The variance of a'x, $E\left[a'x - Ea'x\right]^2 = E\left[a'(x - Ex)\right]^2 =$
$= E\left[a'(x - Ex)(x - Ex)'a\right] = a' E\left[(x - Ex)(x - Ex)'\right] a = a'Va$

Suppose $V = \alpha I$, suppose $Ex = o$

Express the expectation of the 'quadratic form' x'Mx in terms of V.

Answer: $Ex'Mx = E \operatorname{tr} x'Mx = E \operatorname{tr} Mxx' = \operatorname{tr} MExx' = \operatorname{tr} M\alpha I = \alpha \operatorname{tr} M$

2.1 The characteristic equation, roots and vectors.

For a square matrix A one can construct the sets of equations:

$$Au = \pi u \quad \text{and} \quad w'A = \pi w',$$

where u is a column vector,
w' is a row vector,
π is a scalar.
u, w' and π are unknowns.
We want to solve these equations, excluding the trivial solutions where $u = o$, $w' = o$.
Thus the solutions are restricted to the case in which $u \neq o$, $w' \neq o$.
$Au = \pi u$ can be written as $(A-\pi I)u=o$. A solution $u = o$ is possible if $(A-\pi I)^{-1}$ exists, so a necessary condition for a solution $u \neq o$ is that $(A-\pi I)^{-1}$ does not exist. This implies $|A-\pi I| = o$.
Linear independence of columns of $A-\pi I$ implies $|A-\pi I| \neq o$, so a sufficient condition for linear dependence of columns of $A-\pi I$ is $|A-\pi I| = o$.
This amounts to solving the determinant equation

$$|A-\pi I| = o,$$

also called the characteristic equation of A.
($w'A = \pi w'$ gives rise to the same conclusion.)

The equation $|A-\pi I| = o$ is an equation in π of the same order as A. If A has order (n,n), the equation has n solutions for π : $\pi_1, \ldots\ldots, \pi_n$, some of which may be multiple.
To each π_i there corresponds a pair of row and column vectors w_i' and u_i (up to a proportionality coefficient: if $w_i'A = \pi_i w_i'$, then also $(\alpha w_i)'A = \pi_i(\alpha w_i)'$).
π_i is called a characteristic root of A,
w_i' is called a characteristic row vector of A,
u_i is called a characteristic column vector of A.

The determinant equation $|A-\pi I| = o$ can be written as
$(-\pi)^n + c_{n-1}(-\pi)^{n-1} + \ldots + c_1(-\pi) + c_o = o$
or in terms of the n solutions as
$(\pi_1 - \pi)\ (\pi_2 - \pi)\ \ldots\ldots\ (\pi_n - \pi) = o$

From the second expression we deduce:

$$c_{n-1} = \sum_i \pi_i$$

$$c_o = \prod_i \pi_i$$

These results are obvious.
If we put $\pi = o$, we get $|A| = c_o$.

It is easy to see that $c_{n-1}(-\pi)^{n-1}$ is part of $\prod_i(a_{ii}-\pi)$ only, on expanding $|A-\pi I|$.
This part is $\sum_i(a_{ii})(-\pi)^{n-1} = (\text{tr } A)(-\pi)^{n-1}$, so tr $A = c_{n-1}$.
Combining the results, one gets:

1. $\sum_i \pi_i = \text{tr } A$ (35)
2. $\prod_i \pi_i = |A|$ (36)

Some more properties of characteristic roots and vectors are easily established.

3. Characteristic row and column vectors corresponding to different characteristic roots, are "orthogonal": their scalar product is zero. (37)

Proof: $Au_i = \pi_i u_i$ and $w_j'A = \pi_j w_j'$ imply:

$w_j'Au_i = \pi_i w_j'u_i$ and $w_j'Au_i = \pi_j w_j'u_i$

So $(\pi_i-\pi_j)w_j'u_i = o$. Because $\pi_i \neq \pi_j$, $w_j'u_i = o$.

4. In case A is symmetric, characteristic row vectors are transposed characteristic column vectors. (38)

Proof: $Au_i = \pi_i u_i$ implies $u_i'A' = \pi_i u_i'$,

so $u_i'A = \pi_i u_i'$, because $A' = A$.

5. In case A is symmetric, all characteristic roots are real. (39)

Proof: Let $Au_j = \pi_j u_j$, where $u_j = a_j+ib_j$,

($i = \sqrt{-1}$, a_j and b_j being real vectors),

then $(a_j-ib_j)'A(a_j+ib_j) = \pi_j(a_j-ib_j)'(a_j+ib_j)$.

Clearly $(a_j-ib_j)'A(a_j+ib_j) = a_j'Aa_j+b_j'Ab_j$,

which is real.

Further $(a_j-ib_j)'(a_j+ib_j) = a_j'a_j+b_j'b_j$, which is also real, so π_j is real.

6. If A has characteristic column vector a_j+ib_j, it also has characteristic column vector a_j-ib_j. (40)

Proof: Let characteristic root be

$\alpha_j+i\beta_j$, then $A(a_j+ib_j) = (\alpha_j+i\beta_j)(a_j+ib_j)$. ($\alpha_j, \beta_j$ real)

From this follows:

$Aa_j = \alpha_j a_j - \beta_j b_j$

$A(ib_j) = \alpha_j(ib_j) + \beta_j(ia_j)$,

so $A(a_j-ib_j) = (\alpha_j-i\beta_j)(a_j-ib_j)$.

Note: a_j-ib_j is associated with the complex conjugate $\alpha_j-i\beta_j$. If $\beta_j = o$ (the characteristic root is real),

both vectors correspond to the same real root α_j.

7. $A \boxtimes B$ has characteristic roots (41)

$\alpha_1\beta_1, \ldots, \alpha_1\beta_m, \alpha_2\beta_1, \ldots, \alpha_2\beta_m, \ldots, \alpha_l\beta_1, \ldots, \alpha_l\beta_m$,

if A has characteristic roots α_i $(i=1\ldots l)$

and B has characteristic roots β_j $(j=1\ldots m)$

$A \boxtimes B$ has characteristic column vectors $u_i \boxtimes x_j$

and characteristic row vectors $y_i' \boxtimes w_j'$,

if A has characteristic column vectors u_i

and characteristic row vectors y_i'

and B has characteristic column vectors x_j

and characteristic row vectors w_j',

Proof: According to theorem (10)

$(A \boxtimes B)(u_i \boxtimes x_j) = Au_i \boxtimes Bx_j$

Clearly $\alpha_i\beta_j(u_i \boxtimes x_j) = \alpha_i u_i \boxtimes \beta_j x_j = Au_i \boxtimes Bx_j$

Further $(y_i' \boxtimes w_j')(A \boxtimes B) = y_i'A \boxtimes w_j'B =$

$\alpha_i y_i' \boxtimes \beta_j w_j' = \alpha_i\beta_j(y_i' \boxtimes w_j')$

The reader is now in a position to prove theorem (24):

$|A \boxtimes B| = |A|^m |B|^l$

Proof: According to (36) $|A \boxtimes B| = \prod^k \pi_k$,

π_k standing for the k^{th} characteristic root of $A \boxtimes B$

According to (41) $\prod^k \pi_k = \prod^{ij}(\alpha_i)^m(\beta_j)^l = |A|^m |B|^l$

8. An idempotent matrix has characteristic roots o or 1 (42)

Proof: Let A be idempotent

and $Au_i = \alpha_i u_i$

Premultiplication by A leads to

$A^2u_i = \alpha_i Au_i = \alpha_i^2 u_i$

$A^2 = A$, so $\alpha_i u_i = \alpha_i^2 u_i$, which gives $\alpha_i(1-\alpha_i)u_i = o$,

with $u_i \neq o$

From this follows $\alpha_i = o$ or $\alpha_i = 1$

9. A nilpotent matrix has characteristic root o only. (43)

Proof: According to the definition of a nilpotent matrix A,

$A^n = 0$, with $A^{n-1} \neq 0$, for some integer $n \geq 1$

From the characteristic equation $Au_i = \alpha_i u_i$ one gets

$A^{n-1}u_i = \alpha_i^{n-1}u_i$ and $A^n u_i = \alpha_i^n u_i$, which implies

$\alpha_i^n u_i = 0$. Because $u_i \neq o$, $\alpha_i = o$.

10. A^n has characteristic roots α_i^n, if A has characteristic roots α_i, n being integer or fractional. (44)

Proof: For positive integer n one gets the result by repeated premultiplication of both sides of the characteristic equation $Au_i = \alpha_i u_i$ say, by A.

Let n be a negative integer. A^{-n} has then characteristic equation $A^{-n}u_i = \alpha_i{}^{-n}u_i$.

Premultiplication by A^n gives: $A^nA^{-n}u_i = \alpha_i{}^{-n}A^nu_i$ so $\alpha_i{}^nu_i = A^nu_i$

Let n be a fraction, either positive or negative, $n = \frac{p}{r}$, $r > o$. Let A^n have a characteristic root π and a characteristic column vector z. Then $A^nz = \pi z$

Repeated premultiplication by A^n (r-1times) gives:

$A^{nr}z = \pi A^{(r-1)n}z$, using $A^{xn}z = \pi^x z$, x positive integer,

hence $A^pz = \pi^r z$

We know that A^p has characteristic roots $\alpha_i{}^p$, so π^r is one of them, hence π is one of $\alpha_i{}^{p/r} = \alpha_i{}^n$.

2.2 Diagonalisation of matrices

1. Matrices having single characteristic roots can be diagonalised by means of their characteristic row and column vectors. (45)

Proof: Let A have n single characteristic roots α_i, characteristic row vectors w_i' and characteristic column vectors u_i, and let these vectors be standardised: $w_i'u_i = 1$

Then WAU = L,

with $W = \begin{pmatrix} w_1' \\ \vdots \\ w_n' \end{pmatrix}$, $U = (u_1 \dots u_n)$, $L = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix}$,

because $w_i'Au_i = \alpha_i w_i'u_i = \alpha_i$ and $w_j'Au_i = \alpha_i w_j'u_i = o$

2. $A^m = UL^mW = \sum \alpha_i{}^m u_i w_i'$ (46)

m being an arbitrary power, integer or fractional.

Proof: $w_i'u_i = 1$ and $w_j'u_i = o$ imply WU = I.

According to theorem (22) $|WAU| = |W| \cdot |A| \cdot |U|$, and according to theorem (36) $|A| = \prod \alpha_i$, also $|L| = \prod \alpha_i$, so $|W| \cdot |U| = 1$ which implies that both W and U have inverses.

Hence $W = U^{-1}$ so UW = I.

Premultiplying WAU by U and postmultiplying by W leads to $A = ULW$ and $A^m = UL^mW$, for m is positive integer.

Application of theorem (44) leads to the proof for negative integer and fractional n.

UL^mW can be rewritten as

$$(u_1 \ldots u_n)\begin{pmatrix}\alpha_1^m & & \\ & \ddots & \\ & & \alpha_n^m\end{pmatrix}\begin{pmatrix}w_1' \\ \vdots \\ w_n'\end{pmatrix} =$$

$$= (\alpha_1^m u_1 e_1' + \ldots + \alpha_n^m u_n e_n')\begin{pmatrix}w_1' \\ \vdots \\ w_n'\end{pmatrix} =$$

$$= (\alpha_1^m u_1 \ldots \alpha_n^m u_n)\begin{pmatrix}w_1' \\ \vdots \\ w_n'\end{pmatrix} =$$

$$= \alpha_1^m u_1 w_1' + \ldots + \alpha_n^m u_n w_n' = \sum \alpha_i^m u_i w_i'$$

Use has been made of partitioned multiplication.

$e_i' = (o \ldots 1 \ldots o)$, the unity occupying the i^{th} position.

3. Let u_1 and u_2 be characteristic column vectors, associated with multiple characteristic root α, the matrix **A** being symmetric. Let $u_1'u_2 \neq o$

$u_1 \neq cu_2$, c being any scalar.

The vectors u_1 and $v_2 = u_2 - u_1 \cdot \dfrac{u_1'u_2}{u_1'u_1}$ will be <u>orthogonal</u> characteristic column vectors of **A**. (47)

<u>Proof</u>:

$$u_1'v_2 = u_1'u_2 - u_1'u_1 \cdot \frac{u_1'u_2}{u_1'u_1} = o$$

$$Av_2 = Au_2 - Au_1 \cdot \frac{u_1'u_2}{u_1'u_1} = \alpha u_2 - \alpha u_1 \cdot \frac{u_1'u_2}{u_1'u_1} =$$

$$= \alpha\left[u_2 - u_1 \cdot \frac{u_1'u_2}{u_1'u_1}\right] = \alpha v_2$$

Of course $x'v_2 = o$, for any characteristic column vector x associated with a different characteristic root β.

<u>Note 1</u>: If α has multiplicity 3, say, and associated characteristic column vectors u_1, u_2 and u_3, the same orthogonalisation procedure is applied, in stages:

$$v_2 = u_2 - u_1 \cdot \frac{u_1'u_2}{u_1'u_1}$$

$$v_3 = u_3 - u_1 \cdot \frac{u_1'u_3}{u_1'u_1}$$

$$w_3 = v_3 - v_2 \cdot \frac{v_2'v_3}{v_2'v_2}$$

u_1, v_2, and w_3 will be orthogonal now. And so on for higher multiplicity.

Note 2: The new orthogonal characteristic column vectors plus the ones, associated with single characteristic roots, can be used to diagonalise **A**. (Remember that A is symmetric.)

4. Cayley-Hamilton theorem

A useful theorem is:
Each matrix satisfies the matrix counterpart of its characteristic equation. (48)

In section 2.1 the characteristic equation was expressed as:

$$(-\pi)^n + c_{n-1}(-\pi)^{n-1} + \ldots + c_1(-\pi) + c_0 = 0$$

The Cayley-Hamilton theorem then asserts:

$$(-A)^n + c_{n-1}(-A)^{n-1} + \ldots + c_1(-A) + c_0 I = 0$$

Proof: 1. For A having non-repeated characteristic roots, we have: $(-A)^m = \dot{\Sigma}(-\pi_i)^m u_i w_i'$. (See theorem (46))

Thus $(-A)^n + c_{n-1}(-A)^{n-1} + \ldots + c_1(-A) + c_0 I =$

$$= \dot{\Sigma}\left[(-\pi_i)^n + c_{n-1}(-\pi_i)^{n-1} + \ldots + c_1(-\pi_i) + c_0\right]u_i w_i' =$$

$$= 0.\dot{\Sigma}u_i w_i' = 0$$

2. For B having repeated characteristic roots, we use the alternative version of the Cayley-Hamilton theorem:

$(\pi_1 I - A)(\pi_2 I - A) \ldots (\pi_n I - A) = 0$, for A having non-repeated roots.
Taking differences one gets:

$$0 = (\pi_1 I - A) \ldots (\pi_n I - A) + \Delta\left[(\pi_1 I - A) \ldots (\pi_n I - A)\right] =$$

$$= \left[(\pi_1 + \Delta\pi_1)I - (A + \Delta A)\right] \ldots \left[(\pi_n + \Delta\pi_n) - (A + \Delta A)\right] \quad ')$$

If we split B in A and ΔA, A having non-repeated characteristic roots π_i, and B having by definition repeated characteristic roots $\beta_i = \alpha_i + \Delta\alpha_i$, we see that B satisfies its own characteristic equation.

5. Powers of matrices, derived by the Cayley-Hamilton theorem

Sometimes powers of matrices can be computed more easily by means of the Cayley-Hamilton theorem. Computation of the inverse is the easiest application. —

') For the application of the difference operator Δ to matrices see chapter 4.

Premultiplication of

$(-A)^n + c_{n-1}(-A)^{n-1} + \ldots + c_1(-A) + c_0 I = 0$ by $(-A)^{-1}$ gives:

$$A^{-1} = \frac{1}{c_0}(-A)^{n-1} + c_{n-1}(-A)^{n-2} + \ldots + c_1 I$$

Fractional powers can sometimes be derived from other fractional powers by making use of theorem (44): if A satisfies the equation $(\pi_1 I - A) \ldots (\pi_n I - A) = 0$, then $A^{1/m}$ satisfies the equation $(\pi_1^{1/m} I - A^{1/m}) \ldots (\pi_n^{1/m} I - A^{1/m}) = 0$ (m integer).

2.3 Quadratic and bilinear forms.

Cauchy's inequality

Definiteness

Rank

1. Quadratic forms. Definiteness

An expression like $f = x'Ax$ is called a quadratic form. It is clearly equal to $\sum^{ij} a_{ij} x_i x_j$. Of course A is a square matrix. It can be assumed to be symmetric. If A is not symmetric, the quadratic form can be rewritten as follows:

$$x'Ax = \tfrac{1}{2}x'Ax + \tfrac{1}{2}x'Ax = \tfrac{1}{2}x'Ax + \tfrac{1}{2}x'A'x = x'\left[\tfrac{1}{2}(A + A')\right]x.$$

$\frac{1}{2}(A + A')$ is clearly symmetric.

If $f = x'Ax > o$ for any $x \neq o$, f and A are called positive definite.

If $f = x'Ax \geqslant o$ for any $x \neq o$, f and A are called semi-positive definite

If $f = x'Ax < o$ for any $x \neq o$, f and A are called negative definite.

If $f = x'Ax \leqslant o$ for any $x \neq o$, f and A are called semi-negative definite.

As we remember, symmetric matrices can always be diagonalised: $M'AM = L$, M being orthogonal, L being diagonal and having the characteristic roots π_i of A as diagonal elements.

Let us define: $x = My$,

then $f = y'M'AMy = y'Ly = \sum^{i} \pi_i y_i^2$

Clearly A is positive definite if, and only if, all $\pi_i > o$,

A is semi-positive definite if, and only if, all $\pi_i \geqslant o$,

A is negative definite if, and only if, all $\pi_i < o$,

A is semi-negative definite if, and only if, $\pi_i \leqslant o$, for all i.

An alternative definition, which will be given without proof, is: The quadratic form x'Ax is positive definite if,and only if,all successive principal minors

$$a_{11}\ ,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}\ ,\ \ldots\ldots,\ |A|$$

are positive. ')

The quadratic form x'Ax is negative definite if, and only if, all successive principal minors are alternatively negative and positive:

$$a_{11} < 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} > 0,\ \ldots\ldots\ ,\ (-1)^n |A| > 0$$

where n is the order of **A**.

An important property, which will be stated without proof, is the following:

The quadratic form x'Ax is positive definite, subject to Bx = o, with all m successive principal minors of B non-zero, where the order of B is (m,n) (m < n), if and only if,

$$(-1)^m \begin{vmatrix} a_{11} \cdots\cdots\cdots a_{1,m+1} & b_{11} \cdots\cdots b_{m1} \\ \vdots \qquad\qquad \vdots & \vdots \qquad\qquad \vdots \\ a_{1,m+1} \cdots a_{m+1,m+1} & b_{1,m+1} \cdots b_{m,m+1} \\ b_{11} \cdots\cdots\cdots b_{1,m+1} & \\ \vdots \qquad\qquad \vdots & \qquad 0 \\ b_{m1} \cdots\cdots\cdots b_{m,m+1} & \end{vmatrix} > 0,$$

$$(-1)^m \begin{vmatrix} a_{11} \cdots\cdots a_{1,m+2} & b_{11} \cdots\cdots b_{m1} \\ \vdots \qquad\qquad \vdots & \vdots \qquad\qquad \vdots \\ a_{1,m+2} \cdots a_{m+2,m+2} & b_{1,m+2} \cdots b_{m,m+2} \\ b_{11} \cdots\cdots b_{1,m+2} & \\ \vdots \qquad\qquad \vdots & \qquad 0 \\ b_{m1} \cdots\cdots b_{m,m+2} & \end{vmatrix} > 0$$

$$\vdots$$

$$(-1)^m \begin{vmatrix} A & B' \\ B & 0 \end{vmatrix} > 0$$

') See Debreu (8), p. 295 et seq.

The quadratic form x'Ax is negative definite, subject to Bx = o, with all m successive principal minors of B non-zero, if, and only if,

$$(-1)^{m+1}\begin{vmatrix} a_{11} & \cdots & a_{1,m+1} & b_{11} & \cdots & b_{m1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{1,m+1} & \cdots & a_{m+1,m+1} & b_{1,m+1} & \cdots & b_{m,m+1} \\ b_{11} & \cdots & b_{1,m+1} & & & \\ \vdots & & \vdots & & 0 & \\ b_{m1} & \cdots & b_{m,m+1} & & & \end{vmatrix} > o$$

$$(-1)^{m+2}\begin{vmatrix} a_{11} & \cdots & a_{1,m+2} & b_{11} & \cdots & b_{m1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{1,m+2} & \cdots & a_{m+2,m+2} & b_{1,m+2} & \cdots & b_{m,m+2} \\ b_{11} & \cdots & b_{1,m+2} & & & \\ \vdots & & \vdots & & 0 & \\ b_{m1} & \cdots & b_{m,m+2} & & & \end{vmatrix} > o$$

$$\vdots$$

$$(-1)^{n}\begin{vmatrix} A & B' \\ B & 0 \end{vmatrix} > o$$

2. Cauchy's inequality

A very useful inequality is (49)

$$(x'y)^2 \leq x'x.y'y$$

x and y being (real) vectors.

Proof: Define A = (x y).

Clearly $A'A = \begin{pmatrix} x'x & x'y \\ x'y & y'y \end{pmatrix}$ is

semi-positive definite, because $u'A'Au = (Au)'Au \geq o$ for $u \neq o$.
So $|A'A| = x'x.y'y - (x'y)^2 \geq o$.
This proves the theorem.
Note: The matrix version is:

$(\text{tr } P'Q)^2 \leq (\text{tr } P'P)\ (\text{tr } Q'Q)$, which can be deduced by defining: $x = \begin{pmatrix} P_{.1} \\ \vdots \\ P_{.n} \end{pmatrix}$, $y = \begin{pmatrix} Q_{.1} \\ \vdots \\ Q_{.n} \end{pmatrix}$

3. A product of quadratic forms, written as one quadratic form

The product $\dot{\Pi} g_i$ of n quadratic forms $g_i = x_i'A_ix_i$ $(i=1...n)$(50) can be rewritten as

$$\dot{\Pi} g_i = (x_1 \otimes x_2 \otimes \ldots \otimes x_n)'(A_1 \otimes A_2 \otimes \ldots \otimes A_n)(x_1 \otimes x_2 \otimes \ldots \otimes x_n)$$

Proof: According to theorem (10)

$$(x_1 \otimes x_2)'(A_1 \otimes A_2)(x_1 \otimes x_2) = (x_1' \otimes x_2')(A_1 \otimes A_2)(x_1 \otimes x_2) =$$
$$= (x_1'A_1x_1) \otimes (x_2'A_2x_2) = (x_1'A_1x_1).(x_2'A_2x_2)$$

From this follows: $\dot{\Pi} g_i = \dot{\Pi}(x_i'A_ix_i) =$

$$= (x_1 \otimes x_2 \otimes \ldots \otimes x_n)'(A_1 \otimes A_2 \otimes \ldots \otimes A_n)(x_1 \otimes x_2 \otimes \ldots \otimes x_n).$$

(The subscript i is a labelling index.)

4. Rank

The rank of a (rectangular) matrix A, r(A) is equal to the number of rows in the biggest square submatrix of A, which has non-zero determinant.
There are some important theorems on ranks:

1. $r(AB) \leq \min\left[r(A), r(B)\right]$ (51)

Proof: Let A have order (m,n), let B have order (n,p), let $r(A)=s$, $r(B) = t$, $t \leq s$.
If $r(AB) > \min\left[r(A), r(B)\right]$, AB should have more than t "linearly independent" column vectors: there should not exist scalars $\alpha_1 \ldots \alpha_k \neq o$, $(k>t)$, such that $\alpha_1 AB_{.1} + \ldots + \alpha_k AB_{.k} = o$. But these scalars exist, because $\beta_1 B_{.1} + \ldots + \beta_k B_{.k} = o$, given the rank of B.
In case $t > s$, we would have expressed AB as $\begin{pmatrix} A_{1.}B \\ \vdots \\ A_{m.}B \end{pmatrix}$, and we would have proved that there exist non-zero scalars $\pi_1 \ldots \pi_l$ $(l > s)$, such that $\pi_1 A_{1.} + \ldots + \pi_l A_{l.} = o$, so $\pi_1 A_{1.}B + \ldots + \pi_l A_{l.}B = o$

Note: $B_{.1} \ldots B_{.k}$ are not necessarily the first k column vectors of B, $A_{1.} \ldots A_{l.}$ are not necessarily the first

l row vectors of A.

2. $r(A+B) \leq r(A) + r(B)$ (52)

Proof: Assume $r(A+B) > r(A)+r(B)$, and let $r(A) = m$ and $r(B) = n$, select then a submatrix of order $p > m+n$, from A+B. Its determinant can be expressed as a sum of determinants, having i rows from A and p-i rows from B. $(o \leq i \leq p)$
It is clear that all these determinants are zero, so the submatrix of order p from A+B is singular.
This implies that $r(A+B) \leq r(A) + r(B)$.

3. $r(AB) = r(A)$, if B is nonsingular. (53)

Proof: According to theorem (51): $r(AB) \leq r(A)$.
From $A = (AB)B^{-1}$ follows: $r(A) \leq r(AB)$, because B and B^{-1} have the same rank. (Their determinants are non zero.)
Thus $r(AB) = r(A)$, if B is nonsingular.

4. $r(A \boxtimes B) = r(A) \times r(B)$ (54)

Proof: According to the definition

$$A \boxtimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & & \vdots \\ a_{n1}B & \dots & a_{nm}B \end{pmatrix}$$

Let A have rank s, and B have rank t. We can therefore select nonsingular submatrices A_x and B_x, of rank s and t respectively. Clearly $A_x \boxtimes B_x$ is a submatrix of $A \boxtimes B$.

As A_x is nonsingular and B_x is nonsingular, then $A_x \boxtimes B_x$ also is nonsingular, according to theorem (24). Its rank is then equal to its order: $s \times t = r(A) \times r(B)$. Clearly no greater nonsingular submatrix of $A \boxtimes B$ can be selected.

5. If the quadratic form $x'Ax$ — A has order n — is transformable into $\sum_{i=1}^{m} b_i y_i^2$, such that each $b_i \neq o$, by a nonsingular transformation $x = Ty$, $r(A) = m$. (55)

Proof: Clearly T'AT and A have identical ranks. (This follows from theorem (53)).

Define $B = \begin{pmatrix} b_1 & & & & \\ & \ddots & & & \\ & & b_m & & \\ & & & o & \\ & & & & \ddots \\ & & & & & o \end{pmatrix}$, with n-m zero diagonal elements; $\sum_{i=1}^{m} b_i y_i^2 = y'By = x'T^{-1}BT^{-1}x = x'Ax$, so $B = T'AT$.

$r(B)=m$, because $\begin{vmatrix} b_1 & & \\ & \ddots & \\ & & b_m \end{vmatrix} \neq o$. So $r(A) = m$.

6. The rank of a symmetric idempotent matrix A is equal to its trace:

$$r(A) = trA \tag{56}$$

The reason is that A has only zero and unity roots; according to theorem (55) r(A) = m, m being equal to the number of unity roots of A, because a quadratic form $x'Ax = y'M'AMy = y'Ly = \sum_{i=1}^{m} y_i^2$. (M is orthogonal.) This proves the theorem.

7. The rank of a symmetric matrix A is greater than or equal to $\frac{(trA)^2}{trA^2}$: (57)

$$r(A) \geq \frac{(trA)^2}{trA^2}$$

Proof: Let A have order n and rank m. Let its nonzero characteristic roots be $\alpha_1 \ldots \alpha_m$. According to Cauchy's inequality:

$$(\sum_{i=1}^{m} \alpha_i)^2 \leq m . \sum_{i=1}^{m} \alpha_i^2, \text{ or } (trA)^2 \leq m\ trA^2.$$

This proves the theorem. (The symmetry of A had to be stipulated to make sure that the characteristic roots are real, so that Cauchy's inequality can be applied.)

Note: In case A is nonsymmetric, we know that $r(AA') \leq r(A)$ and

$$r(AA') \geq \frac{(trAA')^2}{tr(AA')^2} .$$

Combined these results give:

$$r(A) \geq \frac{(trAA')^2}{tr(AA')^2} \tag{58}$$

5. Bilinear forms

An expression like h = x'Ay is called a bilinear form. It is equal to $\sum_{ij} a_{ij} x_i y_j$.

A can no longer be assumed to be symmetric.

2.4 Examples and Exercises

1. To prove that the matrix $B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$, with characteristic roots $\alpha_1 = \alpha_2 = 3$, $\alpha_3 = 7$, satisfies its characteristic equation (Cayley-Hamilton theorem).

Take $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 2\sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$, with characteristic roots $\beta_1 = 3$, $\beta_2 = 5 + \sqrt{7}$, $\beta_3 = 5 - \sqrt{7}$. A will satisfy its characteristic equation: $(\beta_1 I - A)(\beta_2 I - A)(\beta_3 I - A) = 0$.

This implies: $\left[(\beta_1 + \Delta\beta_1)I - (A + \Delta A)\right]\left[(\beta_2 + \Delta\beta_2)I - (A + \Delta A)\right] \times$
$\times \left[(\beta_3 + \Delta\beta_3)I - (A + \Delta A)\right] = (\alpha_1 I - B)(\alpha_2 I - B)(\alpha_3 I - B) = 0,$

with $\Delta A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{3} \\ 0 & 0 & 0 \end{pmatrix}$

$\Delta\beta_1 = 0$
$\Delta\beta_2 = 2 + \sqrt{7}$
$\Delta\beta_3 = -2 - \sqrt{7}$

2. Find the inverse of $B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$ using the Cayley-Hamilton theorem.

According to the theorem:
$(3I - B)(3I - B)(7I - B) = 0$
or $(-B)^3 + 13(-B)^2 + 51(-B) + 63(-B)^0 = 0$
Premultiplication by $(-B)^{-1}$ gives:
$(-B)^2 + 13(-B) + 51(-B)^0 + 63(-B)^{-1} = 0,$
$\therefore B^{-1} = \frac{1}{63}\left[B^2 - 13B + 51I\right] =$
$= \frac{1}{63}\begin{pmatrix} 21 & 0 & 0 \\ 0 & 18 & -3\sqrt{3} \\ 0 & -3\sqrt{3} & 12 \end{pmatrix}$

3. To find the expectation of the product of quadratic forms in the n stochastic variables u_i (having a normal distribution with zero expectation and variance σ^2) : u'Au . u'Bu ')
Clearly $E(u'Au)(u'Bu) = E(u' \otimes u')(A \otimes B)(u \otimes u) =$
$= \mathrm{tr}(A \otimes B)\, E(u \otimes u)(u' \otimes u') = \mathrm{tr}(A \otimes B)E(uu') \otimes (uu').$
$(uu') \otimes (uu')$ consists of n^2 submatrices $u_i u_j\,(uu')$, where

$$Eu_i u_j(uu') = \begin{cases} \sigma^4(E_{ij} + E_{ji}), & \text{if } i \neq j \\ \sigma^4 I + 2\sigma^4 E_{ii}, & \text{if } i = j, \end{cases}$$

') See Nagar (19), p.

E_{ij} having unity element in the ij^{th} position and zeros everywhere else. $(Eu_i^4 = 3\sigma^4)$

This gives: $E(uu') \otimes (uu') = \sigma^4\left\{ I \otimes I + \sum_{i\neq j}^{ij}\left(E_{ij}\otimes(E_{ij} + E_{ji})\right) + \right.$

$\left. + 2\sum^{i}(E_{ii} \otimes E_{ii})\right\}$. So $E(u'Au)(u'Bu) =$

$= \sigma^4 tr(A \otimes B)\left\{I \otimes I + \sum_{i\neq j}^{ij}\left(E_{ij} \otimes (E_{ij} + E_{ji})\right) + 2\sum^{i}(E_{ii} \otimes E_{ii})\right\} =$

$= \sigma^4 tr(A \otimes B) + \sigma^4 \sum_{i\neq j}^{ij} tr\left\{(AE_{ij}) \otimes (BE_{ij})\right\} + \sigma^4 \sum_{i\neq j}^{ij} tr\left\{(AE_{ij})\otimes(BE_{ji})\right\} +$

$+ 2\sigma^4 \sum^{i} tr(AE_{ii}) \otimes (BE_{ii}) = \sigma^4(tr\ A)(tr\ B) + \sigma^4 \sum_{i\neq j}^{ij} a_{ji}b_{ji} + \sigma^4 \sum_{i\neq j}^{ij} a_{ji}b_{ij} +$

$+ 2\sigma^4 \sum^{i} a_{ii}b_{ii} = \sigma^4(tr\ A)(tr\ B) + \sigma^4\ tr\ AB' + \sigma^4\ tr\ AB.$

Without loss of generality we can assume: $B' = B$, so
$E(u'Au)(u'Bu) = \sigma^4(tr\ A)(tr\ B) + 2\ \sigma^4\ tr\ AB.$

4. If **A** is singular, there exists a scalar $\varepsilon \neq o$, such that $A + \varepsilon I$ is nonsingular.

   Proof: Let again $Au = \beta u$, then $(A + \varepsilon I)u = Au + \varepsilon u = (\beta + \varepsilon)u$
   So if A has characteristic roots β, $A + \varepsilon I$ has characteristic roots $\beta + \varepsilon$. If one chooses $\varepsilon \neq -\beta$, $A + \varepsilon I$ is nonsingular.

5. AB and BA have the same characteristic roots, if at least one of the two matrices is nonsingular, both being square.

   Proof: The characteristic equation of AB is:

   $o = |AB - \beta I| = |A(B - \beta A^{-1})| = |A| \cdot |B - \beta A^{-1}|$, assuming $|A| \neq o$

   The characteristic equation of BA is:

   $o = |BA - kI| = |(B - kA^{-1})\ A| = |B - kA^{-1}| \cdot |A|$

   Clearly $|B - \beta A^{-1}| = o$ and $|B - kA^{-1}| = o$ have the same set of solutions, so $\beta = k$

6. AB and BA have the same characteristic roots, even if both matrices are singular.

   Proof: Choose ε, such that $A + \varepsilon I$ is nonsingular, ε in the neighbourhood of zero.
   $(A + \varepsilon I)B$ and $B(A + \varepsilon I)$ will have the same set of characteristic roots, then, however near to zero ε is.

7. If A is symmetric, there exists a scalar ε, such that $A + \varepsilon I$ is positive definite.

   Proof: Let A have characteristic roots $\beta_1 \ldots\ldots \beta_n$.

   Then $A + \varepsilon I$ will have characteristic roots $\beta_1 + \varepsilon \ldots\ldots$ $\ldots\ldots \beta_n + \varepsilon$. Both A and $A + \varepsilon I$ will have real roots,

because they are symmetric. (See theorem (39))
We have to choose ε in such a way that $\beta_i + \varepsilon > 0$ for all i.
The choice $\varepsilon > -\beta_{min}$ will be the correct choice.

8. Give a lower bound for the rank of $\mathbf{A} \otimes B$, with

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix}$$

Applying theorems (54) and (57), we get:

$$r(A \otimes B) = r(A) \times r(B) \geqslant \frac{(trA)^2}{tr\ A^2} \times \frac{(tr\ B)^2}{tr\ B^2} = \frac{36}{28} \times \frac{36}{52} = 0.89.$$

This gives: $r(A \otimes B) \geqslant 1$.
In fact $r(A \otimes B) = p(A) \times p(B) = 2 \times 3 = 6$.
This result shows that theorem (57) is rather weak. It gains relevance if the diagonal elements (strongly) dominate the off-diagonal elements in absolute value.

9. To prove that for the characteristic roots β_i of $K = (A + B)^{-1}A$, where A is semi-positive definite and B is positive definite: $0 \leq \beta_i < 1$

Proof: $Ku = \beta u$, u being the characteristic column vector of K, corresponding to the characteristic root β.
After substitution, premultiplication by A + B and premultiplication by u', we get:
$u'Au = \beta u'Au + \beta u'Bu$
Obviously β can be zero, because A is semi-positive definite. β cannot be less than zero: $(1 - \beta)u'Au = \beta u'Bu$ would have semi-positive left-hand side and negative right-hand side.
β cannot be equal to one; this would imply $u'Bu = 0$.
β cannot exceed one; this would imply that $(1 - \beta)u'Au = \beta u'Bu$ has semi-negative left-hand side and positive right-hand side.
So $0 \leq \beta < 1$

10. Prove that xy' has m - 1 zero characteristic roots and one characteristic root y'x, x and y being column vectors of order (m,1)

Proof: Obviously xy' has rank 1. It has one non-zero characteristic root tr xy' = y'x

11. Prove that $|I + xy'| = 1 + y'x$, x and y being column vectors of order (m,1)

Proof: I + xy' has m - 1 characteristic roots 1, and one characteristic root 1 + y'x (See ex. 10)
$|I + xy'|$ being equal to the product of characteristic roots, we find $|I + xy'| = 1 + y'x$

12. Approximate $|I + tA|$

Answer: Let A have characteristic roots α_i, $i = 1 \ldots n$, then $I + tA$ has characteristic roots $1 + t\alpha_i$, $i = 1 \ldots n$

$|I + tA| = \prod^{i}(1 + t\alpha_i) = 1 + t\sum^{i}\alpha_i$ + terms in t^2, t^3 a.s.o. An approximation to $|I + tA|$ is then $1 + t \operatorname{tr} A$, which neglects terms involving powers of t.

13. Simplify $|A \otimes I + I \otimes A|$, A being nonsingular of order n. Assume $|A| \neq 0$. In case $|A| = 0$, take $A + \varepsilon I_n$, ε being non-zero and not equal to any $-\alpha_i$. (α_i characteristic roots of A.)

$(A \otimes I)^{-1}(A \otimes I + I \otimes A) = I + (A^{-1} \otimes I)(I \otimes A) =$
$= I + A^{-1} \otimes A$

$\therefore |A \otimes I + I \otimes A| = |A \otimes I| \cdot |I + A^{-1} \otimes A| =$
$|A|^n \cdot |I + A^{-1} \otimes A|$

If A has characteristic roots $\alpha_1 \ldots \alpha_n$, then $I + A^{-1} \otimes A$ has characteristic roots:

$1 + \frac{\alpha_i}{\alpha_j}$ $(i,j = 1 \ldots n)$

So $|I + A^{-1} \otimes A| = 2^n|A| \cdot \prod^{j} \prod_{i \neq j}^{i} (\alpha_i + \alpha_j)$ $i,j = 1 \ldots n$

For $n = 2$, we have $|A \otimes I + I \otimes A| = 4|A|(\operatorname{tr} A)^2$

For $n = 3$, we have $|A \otimes I + I \otimes A| =$
$8/9\,|A|\left\{(\operatorname{tr} A)^3 - \operatorname{tr} A^3\right\}^2$

14. Show that $|A \otimes I + I \otimes A'| = |A \otimes I + I \otimes A|$

15. Diagonalise
$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$$

A has characteristic roots $\beta_1 = \beta_2 = 3$, $\beta_3 = 7$. Characteristic column vectors corresponding to root 3 satisfy the equation

$$(0 \quad 1 \quad \sqrt{3}) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

Non-orthogonal solutions are

$u_1 = \begin{pmatrix} 0 \\ -\frac{1}{2}\sqrt{3} \\ \frac{1}{2} \end{pmatrix}$ and $v_1 = \begin{pmatrix} 1 \\ -\frac{1}{2}\sqrt{3} \\ \frac{1}{2} \end{pmatrix}$; u_1 and $u_2 =$

$= v_1 - u_1 \cdot \frac{u_1'v_1}{u_1'u_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ are orthogonal, of standardised length.

Characteristic column vectors corresponding to root 7, satisfy the equation $\begin{pmatrix} -7 & 0 & 0 \\ 0 & -3 & \sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$

This leads to $u_3 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2}\sqrt{3} \end{pmatrix}$. $P = \begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2}\sqrt{3} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2}\sqrt{3} \end{pmatrix}$ will diagonalise **A**.

$P'AP = L$, where $L = \begin{pmatrix} 3 & & \\ & 3 & \\ & & 7 \end{pmatrix}$

16. Find A^n, for $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & \sqrt{3} \\ 0 & \sqrt{3} & 6 \end{pmatrix}$

From ex. 15 we derive:

$$A = PLP'$$

$$\therefore A^n = PL^nP' =$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ -\frac{1}{2}\sqrt{3} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2}\sqrt{3} \end{pmatrix} \begin{pmatrix} 3^n & & \\ & 3^n & \\ & & 7^n \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{2}\sqrt{3} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{pmatrix} =$$

$$= \begin{pmatrix} 3^n & 0 & 0 \\ 0 & \frac{1}{4}(3^{n+1}+7^n) & \frac{1}{4}\sqrt{3}(7^n-3^n) \\ 0 & \frac{1}{4}\sqrt{3}(7^n-3^n) & \frac{3}{4}(3^{n-1}+7^n) \end{pmatrix}$$

17. $\operatorname{tr}(XX')^2 \leq (\operatorname{tr} XX')^2$.

Proof: Let XX' have characteristic roots $\beta_1 \ldots\ldots \beta_n$, $(XX')^2$ then has characteristic roots $\beta_1^2 \ldots\ldots \beta_n^2$.

Clearly $\operatorname{tr}(XX')^2 = \sum_{i=1}^{n} \beta_i^2 \leq (\sum_{i=1}^{n} \beta_i)^2 = (\operatorname{tr} XX')^2$, because all $\beta_i \geq 0$.

18. Solve the system of linear difference equations

$x_t = Ax_{t-1} + b$, with

$$\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 8 \\ 16 \end{pmatrix}$$

Put $x_t = \sum^{i} \alpha_i \beta_i^t u_i + (I - \mathbf{A})^{-1} b$

β_i being i^{th} characteristic root of **A**,

u_i being i^{th} characteristic column vector of **A**,

α_i being an arbitrary constant.

Clearly $\mathbf{A}x_{t-1} = \mathbf{A}\sum^{i} \alpha_i \beta_i^{t-1} u_i + A(I - A)^{-1} b =$

$= \sum^{i} \alpha_i \beta_i^{t-1} Au_i + A(I - A)^{-1} b = \sum^{i} \alpha_i \beta_i^t u_i + A(I - A)^{-1} b =$

$= x_t - (I - A)^{-1} b + A(I - A)^{-1} b = x_t - (I - A)(I - A)^{-1} b =$

$= x_t - b.$

Solution: $x_t = \alpha_1 5^t \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 (-1)^t \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -5 \\ -1 \end{pmatrix}$

Note: If $|\beta_i| < 1$ for all i, $\lim_{t \to \infty} x_t = (I-A)^{-1} b$

19. Necessary for statistical independence of stochastic variables x_i and x_j is: $Ex_ix_j = (Ex_i).(Ex_j)$
Using the results of ex. 3 prove that necessary for statistical independence of u'Au . u'Bu (the u_i having a normal distribution with zero expectation and variance σ^2) is: tr AB = o

20. Prove that a (real) symmetric matrix of order 2 has no multiple roots.

3.1. Matrix sequences and series

In complete analogy with their scalar counterparts we can define matrix sequences and series. (Of course also vector sequences and series can be defined. We will not mention them separately.) [1]

If to each positive integer n (n = 1,2,3, ...) there corresponds a matrix S_n of order (p,q), then the set $\{S_n\}$ of matrices $S_1, S_2, S_3, \ldots S_n$ is called a matrix sequence.

A matrix sequence $\{S_n\}$ is said to be bounded if real finite constants k_{ij} exist such that $|(S_n)_{ij}| \leq k_{ij}$ for every positive integer n.

If no such constants exist, the sequence is unbounded. A matrix sequence $\{S_n\}$ is said to tend to the limit Σ, if for any pre-assigned set of positive numbers ε_{ij}, there exists a number N - which may depend on the ε_{ij} - such that for every $n > N$, $|(S_n)_{ij} - \sigma_{ij}| < \varepsilon_{ij}$, where

$$\Sigma = \begin{pmatrix} \sigma_{11} \cdots \sigma_{1q} \\ \vdots \quad\quad \vdots \\ \sigma_{p1} \cdots \sigma_{pq} \end{pmatrix}$$

So $\lim_{n\to\infty} S_n = \Sigma$

A matrix series $A_1 + A_2 + A_3 + \ldots$, or $\sum_{k=1}^{\infty} A_k$ can be defined, where

$$\begin{aligned} S_1 &= A_1 \\ S_2 &= A_1 + A_2 \\ &\vdots \\ S_n &= \sum_{k=1}^{n} A_k \end{aligned}$$

A matrix series $\sum_{k=1}^{\infty} A_k$ is convergent if the underlying sequence $\{S_n\}$ is tending to a limit.

If $\{S_n\}$ is tending to the limit Σ, we say that Σ is the sum of the convergent series $\sum_{k=1}^{\infty} A_k$. We shall write $\sum_{k=1}^{\infty} A_k = \Sigma$

Necessary and sufficient for the convergence of a matrix series is that all series of corresponding matrix elements converge. The sums of these series are the elements of the sum of the matrix series. Because of this, all results for scalar series can be translated into results for matrix series.

For example, if $\sum_{k=1}^{\infty} A_k$ and $\sum_{k=1}^{\infty} B_k$ are both convergent, they may be added term by term, i.e.

$$\sum_{k=1}^{\infty} (A_k + B_k) = \sum_{k=1}^{\infty} A_k + \sum_{k=1}^{\infty} B_k$$

If $\sum_{k=1}^{\infty} A_k$ converges and if c is an arbitrary scalar constant, the series $\sum_{k=1}^{\infty} (cA_k)$ converges.

So far there is no need for a special treatment of matrix series. The concept of a matrix power series easily arises. This asks for

[1] For scalar sequences and series see Burington and Torrance (6), pp. 446 et seq.

matrix treatment. The reason is the repeated multiplication of elements.

A matrix power series in (square)A, with real numbers c_n, is written as

$$\sum_{k=0}^{\infty} c_k A^k$$

In fact, the power series is the most interesting case and we shall examine only this type of series.

As a prologue we shall discuss some relevant aspects of scalar power series.

A scalar power series $\sum_{k=0}^{\infty} c_k a^k$ is for all a convergent if

$$\lim_{k\to\infty} |c_k|^{1/k} = 0$$

If $\lim_{k\to\infty} |c_k|^{1/k} = +\infty$, the power series nowhere converges.

If $0 < \lim_{k\to\infty} |c_k|^{1/k} < +\infty$, the power series converges absolutely for every $|a| < \frac{1}{\lim_{k\to\infty} |c_k|^{1/k}}$; it diverges for every

$$|a| > \frac{1}{\lim_{k\to\infty} |c_k|^{1/k}} .$$

We shall introduce now the norm concept. This will permit us to switch from matrix power series to scalar ones. It will give sufficiency conditions for convergence of matrix power series. There will be a minimum of proofs, in order not to blur the results.

3.2. Matrix norms

A norm of a matrix Z is a nonnegative number, to be denoted by $\|Z\|$, which satisfies the following conditions:

1. $\|Z\| > 0$, for $Z \neq 0$
2. $\|Z\| = 0$, for $Z = 0$
3. $\|cZ\| = |c| . \|Z\|$ for any scalar c
4. $\|Z + Y\| \leq \|Z\| + \|Y\|$
5. $\|ZX\| \leq \|Z\| . \|X\|$

(The sums and products are supposed to exist.) ')

The conditions are not sufficient in themselves to determine $\|Z\|$. In fact, quite a few definitions satisfy these conditions. Some frequently used matrix norms are:

1. $\|Z\|_1 = \max_i \sum_j |z_{ij}|$, the row sum norm,
2. $\|Z\|_2 = \max_j \sum_i |z_{ij}|$, the column sum norm,
3. $\|Z\|_3 = (\mathrm{tr}\ Z'Z)^{\frac{1}{2}}$
4. $\|Z\|_4 = (nm)^{\frac{1}{2}} \max_{ij} |z_{ij}|$
5. $\|Z\|_5 = \beta^{\frac{1}{2}} = \max_x \left(\frac{x'Z'Zx}{x'x}\right)^{\frac{1}{2}}$

where Z has order (n,m),
β is the maximum characteristic root of $Z'Z$,
and x is a column vector.

These norms also apply to vectors, because vectors can be interpreted as matrices. The reader can easily find the expressions of the five norms for row and column vectors.

The fifth relationship:

$$\|ZX\| \leq \|Z\| . \|X\|$$

is of special relevance, if applied to a matrix Z and a column vector x. It then reads:

$$\|Zx\|_i \leq \|Z\|_i . \|x\|_i \qquad (i = 1, 2, 3, 4, 5.)$$

If applied to a bilinear form x'Zy, the fifth relationship gives:

$$|x'Zy| = \|x'Zy\|_i \leq \|x'\|_i . \|Z\|_i . \|y\|_i \quad (i=1,2,3,4,5)$$

Still more interesting is the relationship

$$\|Zx\|_i \leq \|Z\|_j . \|x\|_i \qquad (i \text{ and } j \text{ different}),$$

which can be obtained for certain combinations of matrix and vector norms. If it obtains, the matrix and vector norms are called compatible.

Equally indexed matrix and vector norms are by definition compatible. Further $\|Z\|_4 = (nm)^{\frac{1}{2}} \max_{ij} |z_{ij}|$ is compatible with $\|x\|_1 = \max_k |x_k|$, $\|x\|_2 = \sum_k |x_k|$ (if Z has not more rows than columns, thus $n \leq m$) and $\|x\|_3 = (x'x)^{\frac{1}{2}} . \|Z\|_5 = \beta^{\frac{1}{2}}$ is compatible with $\|x\|_3$.

') See Faddeev and Faddeeva (12), pp. 103 et seq.

(These compatibility relationships can be extended to bilinear forms, of course.)

It is desirable to find the matrix norm which is less than or equal to any compatible matrix norm, and of course being compatible itself. Such a matrix norm is called subordinate. The formal definition of a subordinate matrix norm $\|Z\|_{sub}$ is:

$$\|Z\|_{sub} = \max_x \frac{\|Zx\|}{\|x\|}$$

It follows that if y is a vector such that $\frac{\|Zy\|}{\|y\|} = \|Z\|_{sub}$, and if

$$\|Zy\| \leq \alpha \|y\|,$$

then $\|Z\|_{sub} \leq \alpha$,

α being any compatible matrix norm.

Clearly a subordinate matrix norm is compatible:

$$\frac{\|Zx\|}{\|x\|} \leq \max_x \frac{\|Zx\|}{\|x\|}, \text{ so } \|Zx\| \leq \|Z\|_{sub} \cdot \|x\|$$

Without proofs we state:

$\|Z\|_1$, $\|Z\|_2$ and $\|Z\|_5$ are subordinate matrix norms associated with $\|x\|_1$, $\|x\|_2$ and $\|x\|_5$ respectively.

A useful theorem is: (59)

The modulus of every characteristic root does not exceed any matrix norm.

Proof: Let $\|A\| = \alpha$. Take $B = \frac{1}{\alpha+\varepsilon} A \qquad \varepsilon > 0$.

Then $\|B\| = \frac{1}{\alpha+\varepsilon} \|A\| = \frac{\alpha}{\alpha+\varepsilon}$, so $\lim_{n\to\infty} B^n = 0$, because $\|B^n\| \leq \|B\|^n$

Examine $Bu = \beta u$, then $B^n u = \beta^n u$

Clearly $\lim_{n\to\infty} \beta^n u = \lim_{n\to\infty} B^n u = o$, for $u \neq o$, so $|\beta| < 1$

But $\beta = \frac{1}{\alpha+\varepsilon} k$, k being characteristic root of A.

Then $\frac{1}{\alpha+\varepsilon} |k| = |\beta| < 1$

So $|k| < \alpha+\varepsilon$

This leads to $|k| \leq \alpha = \|A\|$, because we can take ε as small as we wish.

3.3. Matrix norms and convergence

The following theorem logically follows from the definitions for norms, limits and convergence:

If $\|S_n - \Sigma\|$ tends to the limit o, the sequence $\{S_n\}$ tends to the limit Σ, and so the series $\sum_{k=1}^{\infty} A_k$ converges to Σ. (60)

As explained already, matrix power series will be examined thoroughly. Let us inspect the series $I+B+B^2+\ldots+B^n+\ldots$

We might guess that its sum is $(I-B)^{-1}$ under certain conditions. Under which? Examine $\|I + B + B^2 + \ldots + B^n - (I - B)^{-1}\| = \|(I - B)^{-1}(I - B^{n+1}) - (I - B)^{-1}\| = \|(I - B)^{-1}B^{n+1}\| \leq \|(I - B)^{-1}\| . \|B^{n+1}\| \leq \|(I - B)^{-1}\| . \|B\|^{n+1}$,

because $(I - B)(I + B + B^2 + \ldots + B^n) = I - B^{n+1}$, and if $(I - B)^{-1}$ exists.

If $\|B\| < 1$, then $\lim_{n\to\infty} \|B\|^{n+1} = 0$ and $(I - B)^{-1}$ exists. (All characteristic roots of B are less than 1.)

So $\|I + B + B^2 + \ldots + B^n - (I - B)^{-1}\|$ tends to the limit 0 and

$I + B + B^2 + \ldots\ldots = (I - B)^{-1}$.

This way of reasoning can be followed only exceptionally, if $\sum_{k=1}^{\infty} A_k$ and $\sum$ are "easy" expressions.

Applying conditions 3, 4 and 5 of matrix norms, we can express $\|\sum_{k=0}^{\infty} c_k A^k\|$ in terms of $\|A\|$:

$$\left\| \sum_{k=0}^{\infty} c_k A^k \right\| \leq \sum_{k=0}^{\infty} |c_k| . \|A\|^k .$$

The right hand side expression is a scalar power series in $\|A\|$. Applying the rules for scalar convergence, we can state that

$\sum_{k=0}^{\infty} |c_k| . \|A\|^k$ is for all $\|A\|$ convergent, if

$$\lim_{k\to\infty} |c_k|^{1/k} = 0$$

If $0 < \lim_{k\to\infty} |c_k|^{1/k} < +\infty$, the power series $\sum_{k=0}^{\infty} |c_k| . \|A\|^k$ converges for every $\|A\| < \dfrac{1}{\lim_{k\to\infty} |c_k|^{1/k}}$

In case $\sum_{k=0}^{\infty} |c_k| . \|A\|^k$ converges, $\sum_{k=0}^{\infty} c_k A^k$ has a limit, given the nature of the matrix norm. In this way we can find <u>sufficiency</u> conditions for convergence of matrix power series. If the scalar power series does not converge, we cannot infer that the matrix power series does not converge, however.

Familiar matrix power series are:

1. $I + rA + \frac{r(r-1)}{2!} A^2 + \frac{r(r-1)(r-2)}{3!} A^3 + \ldots\ldots = (I + A)^r$ for
   - any $\|A\|$ and positive integer r
   - $\|A\| < 1$ and any r
2. $I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \ldots = e^A$ for any $\|A\|$
3. $A - \frac{1}{2}A^2 + \frac{1}{3}A^3 - \ldots = \log(I + A)$ for $\|A\| < 1$

It is clear that $\frac{1}{m}\left\|\sum_{k=0}^{\infty} c_k \mathbf{A}^k\right\|_4$ can lead to a <u>convenient</u> upperbound for the elements of the matrix power series $\sum_{k=0}^{\infty} c_k \mathbf{A}^k$, m being the order of **A**. Whether it is <u>efficient</u> depends on $\|\mathbf{A}\|_4$ and on A itself. Sometimes better upperbounds can be determined, making use of the special character of A (e.g. $\mathbf{A} \geqslant 0$).
Applying the formula for the upperbound to the three matrix power series that have been given above, we get, assuming $\|\mathbf{A}\|_4 < 1$:

1a. $\frac{1}{m}\|(I+A)^r\|_4 \leqslant \frac{1}{m}\|I+A\|_4{}^r \leqslant \frac{1}{m}\left\{\|I\|_4 + \|A\|_4\right\}^r = \frac{1}{m}\left\{m + \|A\|_4\right\}^r$
if r is positive integer.

1b. $\frac{1}{m}\|(I+A)^{-r}\|_4 = \frac{1}{m}\left\|\left\{(I+A)^r\right\}^{-1}\right\|_4 \leqslant \frac{1}{m}\,\frac{1}{1-\|I-(I+A)^r\|_4} + 1 - \frac{1}{m} \leqslant$

$\leqslant \frac{1}{m}\cdot\frac{1}{\left\{1-\|A\|_4\right\}^r} + 1 - \frac{1}{m}$, if r is negative integer or negative fractional.

Use has been made of:

$$\|I-(I+A)^r\|_4 = \left\|rA + \frac{r(r-1)}{2!}A^2 + \frac{r(r-1)(r-2)}{3!}A^3 + \ldots\right\|_4 \leqslant$$

$$\leqslant r\|A\|_4 - \frac{r(r-1)}{2!}\|A\|_4^2 + \frac{r(r-1)(r-2)}{3!}\|A\|_4^3 - \ldots =$$

$$= -\left\{1-\|A\|_4\right\}^r + 1 < 1 \text{ , if } \|A\|_4 < 1$$

1c. $\frac{1}{m}\|(I+A)^r\|_4 = \frac{1}{m}\left\|I+rA + \frac{r(r-1)}{2!}A^2 + \frac{r(r-1)(r-2)}{3!}A^3 + \ldots\right\|_4 \leqslant$

$$\leqslant \frac{1}{m}\left\{\|I\|_4 + r\|A\|_4 - \frac{r(r-1)}{2!}\|A\|_4^2 + \frac{r(r-1)(r-2)}{3!}\|A\|_4^3 - \ldots\right\} =$$

$$= \frac{1}{m}\left\{-1 + r\|A\|_4 - \frac{r(r-1)}{2!}\|A\|_4^2 + \frac{r(r-1)(r-2)}{3!}\|A\|_4^3 - \ldots + m + 1\right\} =$$

$$= -\frac{1}{m}\left\{1 - \|A\|_4\right\}^r + 1 + \frac{1}{m} \text{ , if r is positive fractional.}$$

2. $\frac{1}{m}\|e^{\mathbf{A}}\|_4 = \frac{1}{m}\left\|I+A+\frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \ldots\right\|_4 \leqslant$

$$\leqslant \frac{1}{m}\left\{\|I\|_4 + \|A\|_4 + \frac{1}{2!}\|A\|_4^2 + \frac{1}{3!}\|A\|_4^3 + \ldots\right\} =$$

$$\frac{1}{m}\left\{e^{\|A\|_4} + m - 1\right\} = \frac{1}{m}e^{\|A\|_4} + 1 - \frac{1}{m}$$

3. $\frac{1}{m}\|\log(I+A)\|_4 = \frac{1}{m}\left\|A - \frac{1}{2}A^2 + \frac{1}{3}A^3 - \ldots\right\|_4 \leqslant$

$$\leqslant \frac{1}{m}\left\{\|A\|_4 + \frac{1}{2}\|A\|_4^2 + \frac{1}{3}\|A\|_4^3 + \ldots\right\} = -\frac{1}{m}\log\left\{1-\|A\|_4\right\}$$

3.4. Stochastic convergence

A sequence of stochastic matrices $\{S_n\}$ tends stochastically to the limit Σ, if for any pre-assigned set of positive numbers ε_{ij} and δ_{ij} there exists a number N - which may be dependent on the ε_{ij} - such that for every $n > N$:

$$\Pr\left\{\left|(S_n)_{ij} - \sigma_{ij}\right| < \varepsilon_{ij}\right\} > 1 - \delta_{ij} \quad ')$$

In other words: The probability (Pr) that $(S_n)_{ij}$ is arbitrarily close to σ_{ij}, can be made arbitrarily close to 1.

The matrix series $\sum_{k=1}^{\infty} A_k$, where $S_n = \sum_{k=1}^{n} A_k$ converges stochastically to Σ, if $\{S_n\}$ is stochastically tending to the limit Σ.

What has been said about (numerical) convergence in sections 3.1 and 3.3, can be rephrased here in probability terms for stochastic convergence.

It is clear that, if $\|S_n - \Sigma\|$ tends stochastically to the limit o, the sequence $\{S_n\}$ tends stochastically to the limit Σ, and the series $\sum_{k=1}^{\infty} A_k$ converges stochastically to Σ.

Let us examine stochastic matrix power series. As we saw earlier:

$$\left\| \sum_{k=0}^{\infty} c_k A^k \right\| \leq \sum_{k=0}^{\infty} |c_k| \cdot \|A\|^k$$

It can be stated then:

If $\sum_{k=0}^{\infty} |c_k| \cdot \|A\|^k$ converges stochastically to a finite scalar, $\sum_{k=0}^{\infty} c_k A^k$ converges stochastically to a finite matrix.

$\sum_{k=0}^{\infty} |c_k| \cdot \|A\|^k$ will converge stochastically to a finite scalar for all $\|A\|$, if $\lim_{k\to\infty} |c_k|^{1/k} = o$. So

$I + rA + \frac{r(r-1)}{2!} A^2 + \frac{r(r-1)(r-2)}{3!} A^3 + \ldots$ converges stochastically to $(I+A)^r$ for any $\|A\|$ and positive integer r.

$I + A + \frac{1}{2!} A^2 + \frac{1}{3!} A^3 + \ldots$ converges stochastically to e^A for any $\|A\|$.

If we set: $\Pr\{\|A\| < 1\} = 1$, the power series $\sum_{k=0}^{\infty} |c_k| \cdot \|A\|^k$ will converge stochastically for $o < \lim_{k\to\infty} |c_k|^{1/k} < +\infty$. This is an almost trivial condition!

A relaxed approach is: Let the elements a_{ij} of A follow a (normal) distribution with zero mean and variance of order m^h. (Take m as the sample size.) This means that the a_{ij} lie with probability

') See Wilks (28), p. or Anderson (3), p. .

arbitrarily close to 1 in an interval with non-stochastic limits of which the largest in absolute value is of the order $m^{\frac{1}{2}h}$. [2]) The a_{ij} are said to be of the order $m^{\frac{1}{2}h}$ in probability (i.p.). It is known that if a is of order m^h i.p. and if b is of order m^k i.p., ab is of order m^g i.p., with $g \leq h+k$, a+b is of order m^l i.p., with $l \leq \max(h,k)$

Clearly $\|A\|_1$ and $\|A\|_2$ are of order m^g i.p., with $g \leq \frac{1}{2}h$, $\|A\|_3$ is of order $m^{\frac{1}{2}h}$ i.p., and so on.

Given a favorable value of h (= 1, say), powers of matrix norms will be of continuously decreasing orders in probability.

[2]) See Theil (29), p. 4 et seq.

3.5

Examples and exercises

1. $I + \frac{1}{2}A - \frac{1}{8}A^2 + \frac{1}{16}A^3 - \frac{5}{128}A^4 + \ldots$ can be used to approximate $(I+A)^{\frac{1}{2}}$ if $\|A\|<1$.

For $A = \begin{pmatrix} 0.4 & 0.2 \\ 0 & 0.1 \end{pmatrix}$, we get $(I+A)^{\frac{1}{2}} \sim \begin{pmatrix} 1.184 & 0.090 \\ 0 & 1.049 \end{pmatrix}$ using the first four terms of the series.

If we square this matrix, we get $\begin{pmatrix} 1.402 & 0.201 \\ 0 & 1.100 \end{pmatrix}$, whereas

$I+A = \begin{pmatrix} 1.4 & 0.2 \\ 0 & 1.1 \end{pmatrix}$

For the maximum element of $(I+A)^{\frac{1}{2}}$ we find the upperbound 1.275.

2. $I - \frac{1}{3}A - \frac{1}{9}A^2 - \frac{5}{81}A^3 - \frac{10}{243}A^4 - \ldots$ can be used to approximate $(I-A)^{1/3}$ if $\|A\|<1$.

For $A = \begin{pmatrix} 0.4 & 0.2 \\ 0 & 0.1 \end{pmatrix}$ we get $(I-A)^{1/3} \sim \begin{pmatrix} 0.843 & -0.810 \\ 0 & 0.965 \end{pmatrix}$ using the first five terms.

If we cube it, we get $\begin{pmatrix} 0.607 & -0.200 \\ 0 & 0.898 \end{pmatrix}$ whereas $I-A = \begin{pmatrix} 0.6 & -0.2 \\ 0 & 0.9 \end{pmatrix}$

For the maximum element of $(I-A)^{1/3}$ we find the upperbound 1.0358. Realizing that $A \geq 0$, we find a better upperbound: 1

3. e^A can be approximated by $I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \ldots$ for any $\|A\|$.

For $A = \begin{pmatrix} 0.4 & 0.2 \\ 0 & 0.1 \end{pmatrix}$ we find, summing the first four terms:

$e^A \sim \begin{pmatrix} 1.491 & 0.257 \\ 0 & 1.105 \end{pmatrix}$, whereas the maximum element of e^A has as upperbound 1.613

4. Prove that e^A has characteristic roots e^{β_i} if A has characteristic roots β_i.

5. Prove $|e^A| = e^{\mathrm{tr}\, A}$.

6. Prove that e^A is always nonsingular.

7. Prove $\|Z+Y\|_i \leq \|Z\|_i + \|Y\|_i \qquad (i = 1, 2, 3, 4)$ ')

8. Prove $\|ZX\|_i \leq \|Z\|_i \cdot \|X\|_i \qquad (i = 1, 2, 3, 4)$ ')

9. Show that $\|x\|_3 = \|x\|_5$; $\|x\|_1$ and $\|x\|_4$ are proportional.

') The proofs for the fifth matrix norm require differential calculus. This will be introduced in chapter 4.

10. Prove that $\|Z\|_4$ is compatible with $\|x\|_2$ (if Z has not more rows than columns) and with $\|x\|_3$.

11. Prove that $\|Z\|_1$, $\|Z\|_2$ and $\|Z\|_5$ are subordinate matrix norms associated with $\|x\|_1$, $\|x\|_2$ and $\|x\|_3$ respectively.

12. Give an upperbound for the sum of the elements of column vector Ax, with $A = \begin{pmatrix} 1 & 3 \\ 5 & 8 \end{pmatrix}$, $x \geqslant o$, $\sum_i x_i = 10$.

13. Give an upperbound for the greatest characteristic root of $A = \begin{pmatrix} 1 & 6 \\ 2 & 20 \end{pmatrix}$

14. Give a lowerbound for the smallest characteristic root of $A = \begin{pmatrix} 1 & 6 \\ 2 & 20 \end{pmatrix}$

15. Give an upperbound for x'Ay, given $\sum_i x_i = 10$, $\max_i y_i = 2$, $x, y \geqslant o$.

16. Let $c = (I-A)p$ be the n-sector input-output relationship:
with p the production vector
c the final demand vector
A the input coefficients matrix.
Let the system be aggregated into an m-sector system:
p becomes Kp
c becomes Kc,

$$\text{with } K = \begin{pmatrix} e'_{r_1} & & \\ & \ddots & \\ & & e'_{r_m} \end{pmatrix} \quad \text{and}$$

$$e'_{r_i} = (\underbrace{1 \ldots 1}_{r_i \text{ elements}})$$

$$\sum_i r_i = n$$

If we adopt as aggregated input coefficients matrix B, we can define the prediction error

$$K(I-A)p - (I-B)Kp = (BK-KA)p,$$

the difference between <u>predicted</u> sustainable final demand (I-B)Kp and actually sustainable final demand K(I-A)p in the aggregate system.
Give an upperbound for the prediction error.

17. Prove that no squared real characteristic roots of A exceed the maximum characteristic root of A'A.

18. Prove that all characteristic roots α of matrix **A** are located in the intervals

$$|\alpha - a_{ii}| \leq \sum^{j\neq i} |a_{ij}| \qquad (i = 1 \ldots n)$$

Proof: From the equation for the characteristic column vector u

$$\mathbf{A}u = \alpha u$$

we derive

$$|\alpha - a_{ii}| \leq \sum^{j\neq i} |a_{ij}| \cdot \frac{|u_j|}{|u_i|} \quad (i = 1 \ldots n)$$

Take $|u_i| = \|u\|_1$,

then $|\alpha - a_{ii}| \leq \sum^{j\neq i} |a_{ij}| \cdot \frac{|u_j|}{\|u\|_1} \leq \sum^{j\neq i} |a_{ij}|$, which gives the desired result.

19. Prove that $|A| \neq o$, if $|a_{ii}| > \sum^{j\neq i} |a_{ij}| \qquad (i = 1 \ldots n)$

Proof: Using the result of ex. 18 we get

$$|\alpha - a_{ii}| \leq \sum^{j\neq i} |a_{ij}| < |a_{ii}| \qquad (i = 1 \ldots n)$$

or $|\alpha - a_{ii}| < |a_{ii}|$. Clearly $\alpha \neq o$

20. Show that $\mathbf{A} = \begin{pmatrix} -2 & o & 1 \\ 1 & 3 & 1 \\ 2 & 1 & -4 \end{pmatrix}$ is nonsingular.

4.1. Introduction

In econometrics one often meets with problems involving matrices (vectors) and differential calculus. One class of problems derives from partial differentiation of (differentiable) scalar functions of matrices with respect to the matrix elements. Examples are: characteristic root, determinant, trace, quadratic form, bilinear form.
Another class derives from partial differentiation of elements of a matrix function with respect to elements of the argument matrix. An example is $\frac{\partial (A^2)_{ij}}{\partial (A)_{kl}}$

The matrix elements can be dependent on one or more parameters, which leads to more complicated problems.
In this chapter it will be shown how the concepts of partial derivatives, differentials (and differences) can be combined with matrices.
First some definitions will be given.

4.2. Definitions and assumptions. Some basic formulae. Scalar functions

Let $G = \begin{pmatrix} g_{11} & \cdots & g_{1m} \\ \vdots & & \vdots \\ g_{n1} & \cdots & g_{nm} \end{pmatrix}$ be a matrix with <u>independent</u> elements g_{ij}.

(This implies non-repeatedness of the elements.) ')

Let then $dG = \begin{pmatrix} dg_{11} & \cdots & dg_{1m} \\ \vdots & & \vdots \\ dg_{n1} & \cdots & dg_{nm} \end{pmatrix}$

From this definition we derive:

1. $d(GH) = (dG)H + G(dH)$ (61)

<u>Proof</u>: $d(GH)_{ij} = d(\sum^{\kappa} g_{ik}h_{kj}) = \sum^{\kappa} d(g_{ik}h_{kj}) = \sum^{\kappa} (dg_{ik})h_{kj} + \sum^{\kappa} g_{ik}(dh_{kj}) = (dG)_{i.}H_{.j} + G_{i.}(dH)_{.j}$

This proves the theorem.

2. $d(G^{-1}) = -G^{-1}(dG)G^{-1}$ (62)

<u>Proof</u>: This follows from (61) by substituting $H = G^{-1}$

3. $d(G+H) = dG + dH$ (63)

The proof is obvious.

4. $d\ trG = tr\ dG$ (64)

<u>Proof</u>: $d\ trG = d\sum_{i} g_{ii} = \sum_{i} dg_{ii} = tr\ dG$

') If this assumption is not fulfilled, the formulae for the derivatives matrices have to be modified. This is not difficult. It simply amounts to rearranging elements.

Let f be a (differentiable) scalar function of the elements of G, let $\frac{\partial f}{\partial g_{ij}}$ and $\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}$ exist.

Let then $\frac{\partial f}{\partial G} = \begin{pmatrix} \frac{\partial f}{\partial g_{11}} & \cdots & \frac{\partial f}{\partial g_{1m}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial g_{n1}} & \cdots & \frac{\partial f}{\partial g_{nm}} \end{pmatrix}$

and

$$\frac{\partial^2 f}{\partial G^2} = \begin{pmatrix} \frac{\partial^2 f}{\partial g_{11}^2} & \cdots & \frac{\partial^2 f}{\partial g_{n1}\partial g_{11}} & \cdots & \frac{\partial^2 f}{\partial g_{1m}\partial g_{11}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}\partial g_{11}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial g_{11}\partial g_{n1}} & \cdots & \frac{\partial^2 f}{\partial g_{n1}^2} & \cdots & \frac{\partial^2 f}{\partial g_{1m}\partial g_{n1}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}\partial g_{n1}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial g_{11}\partial g_{1m}} & \cdots & \frac{\partial^2 f}{\partial g_{n1}\partial g_{1m}} & \cdots & \frac{\partial^2 f}{\partial g_{1m}^2} & \cdots & \frac{\partial^2 f}{\partial g_{nm}\partial g_{1m}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial g_{11}\partial g_{nm}} & \cdots & \frac{\partial^2 f}{\partial g_{n1}\partial g_{nm}} & \cdots & \frac{\partial^2 f}{\partial g_{1m}\partial g_{nm}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}^2} \end{pmatrix}$$

($\frac{\partial^2 f}{\partial G^2}$ is frequently called the Hessian matrix of f.)

The following theorems can be proved:

5. $df = \mathrm{tr}H(dG)'$ implies $H = \frac{\partial f}{\partial G}$ (65)

and vice versa.

Proof: $df = \sum\sum \frac{\partial f}{\partial g_{ij}} dg_{ij} = \mathrm{tr}\ \frac{\partial f}{\partial G}(dG)'$

Further $df = \mathrm{tr}H(dG)'$.

Identity for any dG proves the theorem.

6. $Z'\frac{\partial \beta}{\partial Zb} = \frac{\partial \beta}{\partial b}$, (66)

β being a scalar function of column vector $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$

Z being a matrix of constants $Z = \begin{pmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & & \vdots \\ z_{m1} & \cdots & z_{mn} \end{pmatrix}$

Proof:

$$\frac{\partial \beta}{\partial Zb} = \begin{pmatrix} \frac{\partial \beta}{\partial \sum z_{1i} b_i} \\ \vdots \\ \frac{\partial \beta}{\partial \sum z_{mi} b_i} \end{pmatrix}$$

$$Z' \frac{\partial \beta}{\partial Zb} = \begin{pmatrix} \sum^{j} \frac{\partial \beta}{\partial \sum^{i} z_{ji} b_i} z_{j1} \\ \vdots \\ \sum^{j} \frac{\partial \beta}{\partial \sum^{i} z_{ji} b_i} z_{jn} \end{pmatrix} = \begin{pmatrix} \frac{\partial \beta}{\partial b_1} \\ \vdots \\ \frac{\partial \beta}{\partial b_n} \end{pmatrix} = \frac{\partial \beta}{\partial b} ,$$

because $\frac{\partial \sum^{i} z_{ji} b_i}{\partial b_k} = z_{jk}$, $\frac{\partial \beta}{\partial b_k} = \sum^{j} \frac{\partial \beta}{\partial \sum^{i} z_{ji} b_i} \cdot \frac{\partial \sum^{i} z_{ji} b_i}{\partial b_k}$

7. $d^2 f = \mathrm{tr}(dG)' A (dG) B$ implies (67)

$(B \otimes A') + (B' \otimes A) - \mathrm{diag}\,(B' \otimes A) = \frac{\partial^2 f}{\partial G^2}$

and vice versa.

(diag $(B' \otimes A)$ is the diagonal part matrix of $B' \otimes A$)

Proof: Obviously A and B are (n,n) and (m,m) matrices, respectively.

$$d^2 f = \mathrm{tr}(dG)' A (dG) B = \mathrm{tr} \begin{bmatrix} (dG)'_{1.} \\ \vdots \\ (dG)'_{m.} \end{bmatrix} A \left[(dG)_{.1} \ldots (dG)_{.m} \right] B =$$

$$= \sum^{ij} (dG)'_{j.} A (dG)_{.i} b_{ij} = (\mathrm{vec}\, dG)'(B' \otimes A)(\mathrm{vec}\, dG) =$$

$$= (\mathrm{vec}\, dG)'(B \otimes A') + (B' \otimes A) - \mathrm{diag}(B' \otimes A)\,(\mathrm{vec}\, dG)$$

On the other hand

$$d^2 f = (\mathrm{vec}\, dG)' \frac{\partial^2 f}{\partial G^2} (\mathrm{vec}\, dG)$$

Identity for any dG proves the theorem. We had to rewrite the quadratic form of the one but last line of the first part of the proof, because $B' \otimes A$ is not necessarily symmetric, whereas $\frac{\partial^2 f}{\partial G^2}$ is. (Young's theorem)

Let the elements g_{ij} depend onthe parameter k. We can define then

$$\frac{\partial G}{\partial k} = \begin{pmatrix} \frac{\partial g_{11}}{\partial k} & \cdots & \frac{\partial g_{1m}}{\partial k} \\ \vdots & & \vdots \\ \frac{\partial g_{n1}}{\partial k} & \cdots & \frac{\partial g_{nm}}{\partial k} \end{pmatrix} \quad \text{and} \quad \frac{\partial^2 G}{\partial k^2} = \begin{pmatrix} \frac{\partial^2 g_{11}}{\partial k^2} & \cdots & \frac{\partial^2 g_{1m}}{\partial k^2} \\ \vdots & & \vdots \\ \frac{\partial^2 g_{n1}}{\partial k^2} & \cdots & \frac{\partial^2 g_{nm}}{\partial k^2} \end{pmatrix} \quad \text{and so on.}$$

One can easily prove the following theorems:

8. $dG = (dk) \frac{\partial G}{\partial k}$ (68)

There is no need for a formal proof.

9. $\frac{\partial GH}{\partial k} = \frac{\partial G}{\partial k} H + G \frac{\partial H}{\partial k}$ (69)

Proof: $\frac{\partial (GH)_{ij}}{\partial k} = \frac{\partial \sum^{\rho} g_{ip} h_{pj}}{\partial k} = \sum^{\rho} \frac{\partial g_{ip}}{\partial k} h_{pj} + \sum^{\rho} g_{ip} \frac{\partial h_{pj}}{\partial k} =$

$= \left(\frac{\partial G}{\partial k}\right)_{i.} H_{.j} + G_{i.} \left(\frac{\partial H}{\partial k}\right)_{.j}$

10. $\frac{\partial f}{\partial k} = \operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial G}{\partial k}$ (70)

Proof: $\frac{\partial f}{\partial k} = \sum^{ij} \frac{\partial f}{\partial g_{ij}} \frac{\partial g_{ij}}{\partial k} = \operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial G}{\partial k}$

11. $\frac{\partial^2 f}{\partial k^2} = (\operatorname{vec} \frac{\partial G}{\partial k})' \frac{\partial^2 f}{\partial G^2} (\operatorname{vec} \frac{\partial G}{\partial k}) + \operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial^2 G}{\partial k^2}$ (71)

Proof: $\frac{\partial^2 f}{\partial k^2} = \frac{\partial}{\partial k} \left[\operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial G}{\partial k} \right] = \frac{\partial}{\partial k} \left(\operatorname{vec} \frac{\partial f}{\partial G}\right)' . \operatorname{vec} \frac{\partial G}{\partial k} +$

$+ \operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial^2 G}{\partial k^2} = \left(\operatorname{vec} \frac{\partial G}{\partial k}\right)' \frac{\partial^2 f}{\partial G^2} \operatorname{vec} \frac{\partial G}{\partial k} +$

$+ \operatorname{tr} \left(\frac{\partial f}{\partial G}\right)' \frac{\partial^2 G}{\partial k^2}$

4.3 Matrix functions

The second kind of function of a matrix is a matrix function, say F(G).

Let F(G) be of order (s,t) , let G be of order (n,m)

An elementary result concerns dF(G). Using theorem (61) dF(G) can be expressed in dG.

E.g. d(KGG'L) = K(dG)G'L + KG(dG)'L, K and L being matrix constants.

Higher-order differentials can be derived in the same way.

E.g. $d^2(KGG'L) = K(dG)(dG)'L + K(dG)(dG)'L = 2K(dG)(dG)'L$

This is very straightforward.

A more complicated problem is finding the matrix of partial derivatives:

$$\begin{pmatrix} \frac{\partial F(G)_{11}}{\partial g_{11}} & \cdots & \frac{\partial F(G)_{s1}}{\partial g_{11}} & \cdots & \frac{\partial F(G)_{1t}}{\partial g_{11}} & \cdots & \frac{\partial F(G)_{st}}{\partial g_{11}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial F(G)_{11}}{\partial g_{n1}} & \cdots & \frac{\partial F(G)_{s1}}{\partial g_{n1}} & \cdots & \frac{\partial F(G)_{1t}}{\partial g_{n1}} & \cdots & \frac{\partial F(G)_{st}}{\partial g_{n1}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial F(G)_{11}}{\partial g_{1m}} & \cdots & \frac{\partial F(G)_{s1}}{\partial g_{1m}} & \cdots & \frac{\partial F(G)_{1t}}{\partial g_{1m}} & \cdots & \frac{\partial F(G)_{st}}{\partial g_{1m}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial F(G)_{11}}{\partial g_{nm}} & \cdots & \frac{\partial F(G)_{s1}}{\partial g_{nm}} & \cdots & \frac{\partial F(G)_{1t}}{\partial g_{nm}} & \cdots & \frac{\partial F(G)_{st}}{\partial g_{nm}} \end{pmatrix}$$

For this derivatives matrix the notation $\frac{\partial \operatorname{vec} F(G)}{\partial \operatorname{vec} G}$ will be used.

It follows then that:

$$\frac{\partial y}{\partial x} = \frac{\partial \text{vec } y}{\partial \text{vec } x} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_m} & \cdots & \frac{\partial y_n}{\partial x_m} \end{pmatrix}, \text{ if } \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Theorem (65) can succesfully be applied to establish $\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$

First we determine $dF(G)_{ij}$. This will involve expressions of dG or (dG)' Without loss of generality we can write for each expression involving dG:

$\{M(dG)N\}_{ij} = M_{i.}(dG)N_{.j} = \text{tr}(M_{i.})'(N_{.j})'(dG)'$

For each expression involving (dG)' we find:

$\{R(dG)'Q\}_{ij} = R_{i.}(dG)'Q_{.j} = \text{tr}(Q_{.j}R_{i.}(dG)'$

$\frac{\partial F(G)_{ij}}{\partial G}$ can therefore be written as a sum of expressions like $(M_{i.})'(N_{.j})'$ and $Q_{.j}R_{i.}$, because of theorem (65).

We shall arrange these terms in the manner of $\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$:

$$\left[\text{vec } \frac{\partial F(G)_{11}}{\partial G} \,..\, \text{vec } \frac{\partial F(G)_{s1}}{\partial G} \,..\, \text{vec } \frac{\partial F(G)_{1t}}{\partial G} \,..\, \text{vec } \frac{\partial F(G)_{st}}{\partial G}\right]$$

For $(M_{i.})'(N_{.j})'$ we then get:

$$\left[\text{vec}(M_{1.})'(N_{.1})' .. \text{vec}(M_{s.})'(N_{.1})' .. \text{vec}(M_{1.})'(N_{.t})' .. \text{vec}(M_{s.})'(N_{.t})'\right] =$$

$= N \otimes M'$

For $Q_{.j}R_{i.}$ we get:

$$\left[\text{vec } Q_{.1}R_{1.} .. \text{vec } Q_{.1}R_{s.} .. \text{vec } Q_{.t}R_{1.} .. \text{vec } Q_{.t}R_{s.}\right] =$$

$$= \begin{bmatrix} Q \otimes (R')_{1.} \\ \vdots \\ Q \otimes (R')_{n.} \end{bmatrix}$$

From these two subresults we derive: (72)

$$\frac{\partial \text{vec } F(G)}{\partial \text{vec } G} = \sum^{k} N_k \otimes M'_k + \sum^{l} \begin{bmatrix} Q_1 \otimes (R'_1)_{1.} \\ \vdots \\ Q_1 \otimes (R'_1)_{n.} \end{bmatrix} \quad (k = 1 .. K;\ l = 1 .. L),$$

K being the number of expressions $M(dG)N_{ij}$ that contribute to $dF(G)_{ij}$, L being the number of expressions $R(dG)'Q_{ij}$ that contribute to $dF(G)_{ij}$. k and l are simply labelling indices. They have no further meaning.

It follows that $\frac{\partial^2 f}{\partial G^2}$ can be found from $\frac{\partial f}{\partial G}$ directly, applying theorem (72).

$\frac{\partial^2 f}{\partial G^2} = \frac{\partial \text{vec } \frac{f}{G}}{\partial \text{vec } G}$ will be of order (nm, nm).

4.4. The scalar functions: trace, characteristic root, determinant, quadratic form, bilinear form.

The formulae given so far will help us in determining derivatives and differentials for the scalar functions mentioned above. The trace case has been dealt with already in theorem (64) and (65). We shall deal with the others now.

1. Characteristic root

Let the equations for characteristic row and column vectors and characteristic root of A be:

$Au_i = \alpha_i u_i \qquad w_i'A = \alpha_i w_i' \qquad \text{with } w_i'u_i = 1$

Applying theorem (61), we get:

$(dA)u_i + A(du_i) = (d\alpha_i)u_i + \alpha_i(du_i)$

Premultiplication by w_i' gives:

$w_i'(dA)u_i + w_i'A(du_i) = (d\alpha_i)w_i'u_i + \alpha_i w_i'(du_i),$ ( )

which reduces to:

$$d\alpha_i = w_i'(dA)u_i = \text{tr } w_i u_i'(dA)' \tag{73}$$

According to theorem (65):

$$\frac{\partial \alpha_i}{\partial A} = w_i u_i' \tag{74}$$

If A is symmetric:

$$\frac{\partial \alpha_i}{\partial A} = 2u_i u_i' - \text{diag}(u_i u_i') \tag{75}$$

In case A is symmetric, we can find an expression for $d^2\alpha_i$

Differentiating $d\alpha_i = u_i'(dA)u_i$ again, we find:

$d^2\alpha_i = (du_i)'(dA)u_i + u_i'(dA)du_i = 2(du_i)'(dA)u_i$

We need an expression for $(du_i)'$. This can be found as follows: differentiating $U'A = LU'$, and postmultiplying by U, and using $AU = UL$, we get:

$U'(dA)U - dL = L(dU)'U - (dU)'UL.$

Multiplying ij^{th} elements by $\frac{1}{\alpha_i - \alpha_j}$ $(i \neq j)$, multiplying ii^{th} elements by o, we get:

$$\begin{pmatrix} 0 & \frac{u_1'(dA)u_2}{\alpha_1-\alpha_2} & \cdots\cdots & \frac{u_1'(dA)u_m}{\alpha_1-\alpha_m} \\ \frac{-u_2'(dA)u_1}{\alpha_1-\alpha_2} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{u_{m-1}'(dA)u_m}{\alpha_{m-1}-\alpha_m} \\ \frac{-u_m'(dA)u_1}{\alpha_1-\alpha_m} & \cdots & \frac{-u_m'(dA)u_{m-1}}{\alpha_{m-1}-\alpha_m} & 0 \end{pmatrix} = (dU)'U$$

We shall denote the left-hand side matrix by S. Clearly S is skew-symmetric. Use has been made of $(dU)'U + U'(dU) = 0$. As usual L stands for the diagonal matrix of characteristic roots α_i. $U = (u_1 \ \ldots \ u_m)$

We conclude: $(dU)' = SU'$ and $(du_i)' = S_i.U'$.

So $d^2\alpha_i = 2S_i.U'(dA)u_i$ (76)

Further $\sum_i d^2\alpha_i = 2 \operatorname{tr} SU'(dA)U = 0$, because $S' = -S$ and $(U'(dA)U)' = U'(dA)U$

2. Determinant

Given the fact that $|A| = \prod_i \alpha_i$, we determine:

$$d|A| = d(\prod_i \alpha_i) = \sum_j d\alpha_j \prod_{i\neq j} \alpha_i = \sum_j \operatorname{tr} w_j u_j'(dA)' \prod_{i\neq j} \alpha_i =$$

$$= \operatorname{tr} \sum_j \frac{1}{\alpha_j} w_j u_j'(dA)' \prod_i \alpha_i = |A| \ \operatorname{tr}(A')^{-1}(dA)' \quad (77)$$

Use has been made of: $\sum_j \frac{1}{\alpha_j} w_j u_j' = (A')^{-1}$

Application of (65) gives:

$$\frac{\partial |A|}{\partial A} = |A| \ (A')^{-1} \quad (78)$$

If A is symmetric

$$\frac{\partial |A|}{\partial A} = |A| \left(2(A)^{-1} - \operatorname{diag}(A)\right) \quad (79)$$

3. Quadratic and bilinear forms

Let us examine $f = x'Ay$ and $g = x'Ax$

Application of theorem (61) gives:

$df = (dx)'Ay + x'(dA)y + x'A(dy)$

$dg = (dx)'Ax + x'(dA)x + x'A(dx)$,

from which follows, according to theorem (65):

$$\frac{\partial f}{\partial x} = Ay, \frac{\partial f}{\partial y} = A'x, \frac{\partial f}{\partial A} = xy', \frac{\partial g}{\partial x} = (A+A')x, \frac{\partial g}{\partial A} = xx' \qquad (80)$$

If A is symmetric, some of these results have to be modified slightly: $\frac{\partial f}{\partial A} = xy' + yx' - \text{diag}(xy')$; $\frac{\partial g}{\partial A} = 2xx' - \text{diag}(xx')$.

4.5. Some applications

1. Find $\frac{\partial \text{tr } A'KA}{\partial A}$, if K' = K.

Clearly d tr A'KA = 2 tr KA(dA)', so according to (65)

$\frac{\partial \text{tr } A'KA}{\partial A} = 2KA$

2. Determine then $\frac{\partial^2 \text{tr } A'KA}{\partial A^2}$

We first establish

$$d\left[\frac{\partial \text{tr } A'KA}{\partial A}\right]_{ij} = d(2KA)_{ij} = 2K_{i.}(dA)I_{.j}$$

So according to (72) $\frac{\partial^2 \text{tr } A'KA}{\partial A^2} = 2I \otimes K$

3. Let $A = \begin{pmatrix} C \\ B \end{pmatrix}$ and $P = -BC^{-1}$

Find $\frac{\partial \text{vec } P}{\partial \text{vec } A}$ [1])

We know: $dP_{ij} = d(-BC^{-1})_{ij} = -\left[(dB)C^{-1}\right]_{ij} + \left[BC^{-1}(dC)C^{-1}\right]_{ij} =$

$= -I_{i.}(dB)C^{-1}_{.j} + (BC^{-1})_{i.}(dC)C^{-1}_{.j}$

According to (72):

$$\frac{\partial \text{vec } P}{\partial \text{vec } B} = -C^{-1} \otimes I$$

$$\frac{\partial \text{vec } P}{\partial \text{vec } C} = C^{-1} \otimes (BC^{-1})' = -C^{-1} \otimes P'$$

Thanks to the structure of A, we can assemble the two sub-results into:

$$\frac{\partial \text{vec } P}{\partial \text{vec } A} = -C^{-1} \otimes \begin{pmatrix} P' \\ I \end{pmatrix}$$

(I is a unity matrix having as many rows and columns as B has rows.)

4. Asymptotic standard errors of characteristic roots of an econometric equation system

A well-known theorem relates the asymptotic variance of a scalar function a of a vector π and the asymptotic variance matrix of the vector, Y, in the following way:

$$\text{var } a = \left(\frac{\partial a}{\partial \pi}\right)' Y \frac{\partial a}{\partial \pi} ,$$

$\frac{\partial a}{\partial \pi}$ being evaluated at the asymptotic expectation of a. [2]) [3])

[1]) For the formulation of the problem see Goldberger (13), p.369
[2]) See e.g. Goldberger (13), p. 124
[3]) See Neudecker and van de Panne (20) for the result of this subsection.

Let us apply this theorem to the characteristic roots α_i of the system $y_t = Ay_{t-1} + c_t$ [1])

We then get:

$$\text{var } \alpha_i = (\text{vec } w_i u_i')' Y (\text{vec } w_i u_i'),$$

Y being the asymptotic variance matrix of vec A,

$\frac{\partial \alpha_i}{\partial A} = w_i u_i'$ evaluated at the asymptotic expectation of α_i.

5. To find $\frac{\partial \log |W|}{\partial B}$ [2])

for $W = E'E$

$E = Y - XB$

W positive definite.

Applying theorems (62) and (77), we get:

$d \log |W| = \text{tr } W^{-1} dW = 2 \text{ tr}(X'XB - X'Y)W^{-1} (dB)'$,

because $dW = (dE)'E + E'dE = -(XdB)'(Y-XB) - (Y-XB)'(XdB)$

Theorem (65) then gives:

$$\frac{\partial \log |W|}{\partial B} = 2(X'XB - X'Y)W^{-1}$$

6. Let $S = \frac{1}{T} \begin{bmatrix} (y_1 - Z_1\beta_1)' \\ \vdots \\ (y_M - Z_M\beta_M)' \end{bmatrix} \left[(y_1 - Z_1\beta_1) \ldots (y_M - Z_M\beta_M)\right]$

find then $q_i = -\frac{1}{2} \frac{\partial \log |S|}{\partial \beta_i}$ $\quad (i = 1 \ldots M)$

and $\frac{\partial q_i}{\partial \beta_k}$ $\quad (k = 1 \ldots M)$ [3])

Substituting $u_i = y_i - Z_i\beta_i$, we get

$d \log |S| = \text{tr } S^{-1} dS = -\frac{2}{T} \text{tr}(u_1 \ldots u_M)S^{-1} d(Z_1\beta_1 \ldots Z_M\beta_M)'$,

because $dS = -\frac{1}{T} d(Z_1\beta_1 \ldots Z_M\beta_M)'(u_1 \ldots u_M) -$

$-\frac{1}{T} (u_1 \ldots u_M)' d(Z_1\beta_1 \ldots Z_M\beta_M)$,

$(S^{-1})' = S^{-1}$

and $d \log |S| = \text{tr } S^{-1} dS$

Applying theorem (65), we get:

$$\frac{\partial \log |S|}{\partial (Z_1\beta_1 \ldots Z_M\beta_M)} = -\frac{2}{T} (u_1 \ldots u_M)S^{-1}$$

1) Originally it was: $y_t^{\times} = A_1 y_{t-1}^{\times} + \ldots + A_r y_{t-r}^{\times} + B_0 c_t^{\times} + \ldots + B_q c_{t-q}^{\times}$
The variables have been time-transformed.

2) For the formulation of the problem see Goldberger (13), p. 205

3) For the problem see Rothenberg and Leenders (24), p.64 and 65.

so $\frac{\partial \log |S|}{\partial \alpha_i \beta_i} = \frac{-2}{T} (u_1 \dots u_M)(S^{-1})_{.i}$

Applying theorem (66), we get:

$$\frac{\partial \log |S|}{\partial \beta_i} = - \frac{2}{T} Z_i' (u_1 \dots u_M)(S^{-1})_{.i}$$

Obviously: $q_i = \frac{1}{T} Z_i' (u_1 \dots u_M)(S^{-1})_{.i}$

Then, $dq_i = - \frac{1}{T} Z_i' d(\alpha_1 \beta_1 \dots Z_M \beta_M)(S^{-1})_{.i} -$

$- \frac{1}{T} Z_i' (u_1 \dots u_M) S^{-1} (dS)(S^{-1})_{.i} =$

$= - \frac{1}{T} Z_i' d(Z_1 \beta_1 \dots Z_M \beta_M)(S^{-1})_{.i} +$

$+ \frac{1}{T^2} Z_i' (u_1 \dots u_M) S^{-1} d(Z_1 \beta_1 \dots Z_M \beta_M)' (u_1 \dots u_M)(S^{-1})_{.i} +$

$+ \frac{1}{T^2} Z_i' (u_1 \dots u_M) S^{-1} (u_1 \dots u_M)' d(Z_1 \beta_1 \dots Z_M \beta_M)(S^{-1})_{.i} =$

$= - \frac{1}{T} Z_i' (\sum^j s^{ji} Z_j d\beta_j) +$

$+ \frac{1}{T^2} Z_i' (u_1 \dots u_M) \left\{ \sum^j (S^{-1})_{.j} (\alpha_j d\beta_j)' \right\} (u_1 \dots u_M)(S^{-1})_{.i} +$

$+ \frac{1}{T^2} Z_i' (u_1 \dots u_M) S^{-1} (u_1 \dots u_M)' (\sum^j s^{ji} Z_j d\beta_j) = - \frac{1}{T} \sum^j s^{ji} Z_i' Z_j d\beta_j +$

$+ \frac{1}{T^2} Z_i' (u_1 \dots u_M) \sum^j (S^{-1})_{.j} (S^{-1})_{i.} (u_1 \dots u_M)' Z_j d\beta_j +$

$+ \frac{1}{T^2} \sum^j s^{ji} Z_i' (u_1 \dots u_M) S^{-1} (u_1 \dots u_M)' Z_j d\beta_j \qquad (j = 1 \dots M)$

From this we conclude, applying theorem (72):

$$\frac{\partial q_i}{\partial \beta_k}' = - \frac{1}{T} s^{ki} Z_i' Z_k + \frac{1}{T^2} Z_i' (u_1 \dots u_M)(S^{-1})_{.k} (S^{-1})_{i.} (u_1 \dots u_M)' Z_k +$$

$$+ \frac{1}{T^2} s^{ki} Z_i' (u_1 \dots u_M) S^{-1} (u_1 \dots u_M)' Z_k$$

As usual s^{ij} is the ij^{th} element of S^{-1}.

7. Capital growth matrix ')

Let

$$A = \begin{pmatrix} p_1 & \dots & p_{n-1} & p_n \\ d_1 & \bullet & \circ & \circ \\ \circ & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \circ & \circ & d_{n-1} & \circ \end{pmatrix}$$

Let A have elements

p_i: the productivity rate of a capital stock of age i $(i = 1 \dots n)$

d_j: "survival" rate of a capital of age j $(j = 1 \dots n-1)$

$1 - d_j$: the depreciation rate of a capital of age j

') See Enzer (11), p.186, 191 et seq.

Under certain circumstances A has a non-repeated positive characteristic root r, greater in absolute value than any other root. r is the growth rate of the system. If $r>1$, it expands.

We are interested in the interrelationships between the $\frac{\partial r}{\partial p_i}$ and $\frac{\partial r}{\partial d_j}$, and their dependence on the characteristic row vector u' and column vector x, corresponding to r.

It is known that $\frac{\partial r}{\partial A} = ux'$

If we focus attention on the non zero elements of A only, we can write

$$\begin{pmatrix} \frac{\partial r}{\partial p_1} & \cdots & \frac{\partial r}{\partial p_{n-1}} & \frac{\partial r}{\partial p_n} \\ \frac{\partial r}{\partial d_1} & o & o & o \\ o & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ o & o & \frac{\partial r}{\partial d_{n-1}} & o \end{pmatrix} = \begin{pmatrix} u_1x_1 & \cdots & u_1x_n \\ \vdots & & \vdots \\ \vdots & & \vdots \\ u_nx_1 & \cdots & u_nx_n \end{pmatrix}$$

We can derive now:

i. the growth rate increases due to increases in productivity rates p_i depend on x_i and u_i only; the ordering $x_1 > x_2 > \cdots > x_n$ (for $r>1$) implies the ordering $\frac{\partial r}{\partial p_1} > \frac{\partial r}{\partial p_2} > \cdots > \frac{\partial r}{\partial p_n}$ (if $u_1 > o$)

ii. For a given age category of capital stock, j say, the growth rate increase due to an increase in productivity p_j is greater than the growth rate increase due to an equivalent increase in the survival rate d_j, if $u_1 > u_2 > \cdots > u_n$, and if $x_1 > x_2 > \cdots > x_n$: $\frac{\partial r}{\partial p_j} > \frac{\partial r}{\partial d_j}$

iii. Between any pair of survival rates, the increment in the growth rate is greater for increases in the survival rate of younger capital than for increases in the survival rate of older capital: i.e.

$$\frac{\partial r}{\partial d_j} > \frac{\partial r}{\partial d_k} \qquad (j < k)$$

if $u_1 > u_2 > \cdots > u_n$ and $x_1 > x_2 > \cdots > x_n$

iv. The growth rate r is a linear homogeneous function of the p_i and d_j because

$$r = u'Ax = \sum_{i=1}^{n} u_i p_i x_i + \sum_{i=1}^{n-1} u_{i+1} d_i x_i$$

So a proportional change in all productivity and survival rates produces an equal proportional change in the growth rate r.

4.6. Jacobians

With the help of theorem (72), the Jacobians of (linear) transformations can be determined. For the transformation from the matrix G to the matrix F(G), both having order (n,m) , the Jacobian J is:

$$J = \left| \frac{\partial \text{vec } F(G)}{\partial \text{vec } G} \right|^{-1}$$

$\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$ is square, of course, in these cases, so its order is $(n.m \;,\; n.m)$. In addition to non-repeatedness of the elements of G, we have to stipulate non-repeatedness of the elements of F(G), in order to avoid singularity of $\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$.

It is a standard result that

$$\left| \frac{\partial \text{vec } G}{\partial \text{vec } F(G)} \right| = \left| \frac{\partial \text{vec } F(G)}{\partial \text{vec } G} \right|^{-1} \qquad (81)$$

Proof: $\left[d \text{ vec } F(G) \right]' = (d \text{ vec } G)' \frac{\partial \text{vec } F(G)}{\partial \text{vec } G}$

$$(d \text{ vec } G)' = \left[d \text{ vec } F(G) \right]' \frac{\partial \text{vec } G}{\partial \text{vec } F(G)}$$

so: $(d \text{ vec } G)' = (d \text{ vec } G)' \frac{\partial \text{vec } F(G)}{\partial \text{vec } G} \frac{\partial \text{vec } G}{\partial \text{vec } F(G)}$,

for any dG. This implies:

$$\frac{\partial \text{vec } F(G)}{\partial \text{vec } G} \frac{\partial \text{vec } G}{\partial \text{vec } F(G)} = I \text{ and} \left| \frac{\partial \text{vec } G}{\partial \text{vec } F(G)} \right| = \left| \frac{\partial \text{vec } F(G)}{\partial \text{vec } G} \right|^{-1}$$

It can easily be seen that theorem (81) is a special case of the following theorem, which will be given without proof:

If Y is a function of U, U is a function of V, V is a function of X,

$$\left| \frac{\partial \text{vec } Y}{\partial \text{vec } X} \right| = \left| \frac{\partial \text{vec } Y}{\partial \text{vec } U} \right| \cdot \left| \frac{\partial \text{vec } U}{\partial \text{vec } V} \right| \cdot \left| \frac{\partial \text{vec } V}{\partial \text{vec } X} \right| \qquad (82)$$

The requirements concerning non-repeatedness of elements will not always be met.

Examples are:

1. $F(G) = GG'$, with $G' \neq G$. If the order of G is (p,p) F(G) will have the same order. F(G) is symmetric, it has only $\frac{p(p+1)}{2}$ non-repeated elements. G can also have only $\frac{p(p+1)}{2}$ non-repeated elements. This leads to the requirement that G be triangular, with $p + (p-1) + (p-2) + \ldots + 1 = \frac{p(p+1)}{2}$ non-repeated elements:

$$\begin{pmatrix} g_{11} & o & \cdots & o \\ g_{21} & g_{22} & o & \cdots & o \\ g_{31} & g_{32} & g_{33} & o & \cdots & o \\ \vdots & \vdots & \vdots & & & \vdots \\ g_{p-1,1} & g_{p-1,2} & g_{p-1,3} & \cdots & g_{p-1,p-1} & o \\ g_{p1} & g_{p2} & g_{p3} & \cdots & & g_{pp} \end{pmatrix}$$

$\frac{\partial \text{vec } GG'}{\partial \text{vec } G}$ computed according to (72) will contain repeated columns (corresponding with repeated elements of GG') and meaningless rows (corresponding with "non-elements" of G). $\frac{\partial \text{vec } GG'}{\partial \text{vec } G}$ will have to be reduced to get rid of these rows and columns, in order to avoid singularity.

2. $F(G) = KGK'$, with $G' = G$. Let the order of G be (p,p), F(G) is symmetric. G has only $\frac{p(p+1)}{2}$ non-repeated elements and F(G) also has only $\frac{p(p+1)}{2}$ non-repeated elements.

$\frac{\partial \text{vec } KGK'}{\partial \text{vec } G}$ computed according to (72) will contain repeated rows and columns, corresponding with repeated elements of G and KGK' respectively.

$\frac{\partial \text{vec } KGK'}{\partial \text{vec } G}$ will have to be reduced.

Deemer and Olkin (9), following a different road, have given procedures to find $\left|\frac{\partial \text{vec } F(G)}{\partial \text{vec } G}\right|$ in case of repeated elements. The reader is advised to read their articles.

4.7. Differences and approximation; Taylor's theorem

Obviously the difference operator Δ can be applied to matrices as well.

$$\Delta G = \begin{pmatrix} \Delta g_{11} & \cdots & \Delta g_{1m} \\ \vdots & & \vdots \\ \Delta g_{n1} & \cdots & \Delta g_{nm} \end{pmatrix}$$ and so on up to $\Delta^p G$

The following theorems can be proved:

1. $\Delta(G+H) = \Delta G + \Delta H$ (83)

2. $\Delta(GH) = (\Delta G)H + G(\Delta H) + (\Delta G)(\Delta H)$ (84)

Proof: $\Delta(GH)_{ij} = \Delta(\sum^{k} g_{ik}h_{kj}) = \sum^{k} \Delta(g_{ik}h_{kj}) = \sum^{k} (\Delta g_{ik})h_{kj} +$

$+ \sum^{k} (g_{ik})\Delta h_{kj} + \sum^{k} (\Delta g_{ik})(\Delta h_{kj}) = (\Delta G)_{i.} H_{.j} + G_{i.} (\Delta H)_{.j} +$

$+ (\Delta G)_{i.} (\Delta H)_{.j}$

This proves the theorem.

Taylor's theorem can usefully be applied to scalar functions f of matrices G.
The easiest representation is in terms of differentials:

$$\Delta f = df + \frac{1}{2!} d^2 f + \frac{1}{3!} d^3 f + \ldots\ldots, \qquad (85)$$

where df, $d^2 f$, $d^3 f$, etc. can be expressed in $dG = \Delta G$.
If third and higher order terms are not used, Δf can be expressed in ΔG as follows:

$$\Delta f = \mathrm{tr} \frac{\partial f}{\partial G}' (\Delta G) + \frac{1}{2!} (\mathrm{vec}\, \Delta G)' \frac{\partial^2 f}{\partial G^2} (\mathrm{vec}\, \Delta G) + \ldots. \qquad (86)$$

If the elements of ΔG are small, Taylor's theorem can be used for approximation purposes, involving only few terms.
Taylor's theorem can also be applied to matrix functions $F(G)$ of matrices G:

$$\Delta F(G) = dF(G) + \frac{1}{2!} d^2 F(G) + \frac{1}{3!} d^3 F(G) + \ldots\ldots, \qquad (87)$$

if $dF(G)$ can be defined. Again $dG = \Delta G$.
Taylor's theorem cannot be applied to e.g. $G^{\frac{1}{2}}$, because $dG^{\frac{1}{2}}$ cannot be determined in general:
$dG = (dG^{\frac{1}{2}})G^{\frac{1}{2}} + G^{\frac{1}{2}}(dG^{\frac{1}{2}})$ cannot be solved in general.
It can be applied to e.g. G^{-1}:

$$\Delta G^{-1} = \sum_{i=1}^{\infty} (-1)^i (G^{-1}\Delta G)^i G^{-1}$$

The matrix series will be convergent if $\| G^{-1}\Delta G \| < 1$.
Sufficient is also: $\| G^{-1} \| \, . \, \| \Delta G \| < 1$.

Theorems (86) and (87) can be combined to give an occasionally better approximation of Δf in terms of first-order partial derivatives and first differences:

$$\Delta f = \mathrm{tr} \frac{\partial f}{\partial G}' \Delta G + \frac{1}{2!} (\mathrm{vec}\, \Delta G)' \Delta \frac{\partial f}{\partial \mathrm{vec}\, G} + \ldots., \text{ because} \qquad (88)$$

$$\Delta \frac{\partial f}{\partial \mathrm{vec}\, G} = d \frac{\partial f}{\partial \mathrm{vec}\, G} + \ldots. = \frac{\partial^2 f}{\partial G^2} \mathrm{vec}\, dG + \ldots. = \frac{\partial^2 f}{\partial G^2} \mathrm{vec}\, \Delta G + \ldots$$

There is no need to determine the Hessian, therefore.
In case f is a scalar function of a column vector g, the expression reduces to:

$$\Delta f = \frac{\partial f}{\partial g}' \Delta g + \frac{1}{2!} (\Delta g)' \Delta \frac{\partial f}{\partial g} + \ldots. = (\Delta g)' \left\{ \frac{\partial f}{\partial g} + \frac{1}{2!} \Delta \frac{\partial f}{\partial g} \right\} + \ldots.') \qquad (89)$$

4.8. Optimization of scalar functions; Lagrange multipliers

It is well-known that the first-order condition for an optimum of a scalar function f is:

$$df = o$$

The second-order condition for a $\begin{matrix}\text{maximum} \\ \text{minimum}\end{matrix}$ is $d^2 f \begin{matrix}< o \\ > o\end{matrix}$

') For this theorem see Kloek (16), p. 120 and 121.

Optimization of scalar functions of matrices (vectors) is a straightforward procedure with the help of the theorems of sections 2 and 4 of this chapter. The only thing we have to cope with is equality-constrained optimization. Here Lagrange multipliers come in.
In order to deal with multiple side relations Lagrange matrix multipliers are used.
Suppose the side relation is $AX = B$, A and B being constants. B has order (n,m). The matrix expression to be joined to the scalar function which is to be optimized, is then: tr $M(AX-B)$, M being a Lagrange matrix multiplier. The amended scalar function can be optimized unconditionally now.
It is easy to see that tr $M(AX-B)$ stands for the $n \times m$ side relations:

$$(AX)_{ij} = (B)_{ij}$$

$$\text{tr } M(AX-B) = \sum_{j}\Big[M(AX-B)\Big]_{jj} = \sum_{ij} m_{ji}(AX-B)_{ij} = \sum_{ij} m_{ji}\Big[(AX)_{ij} - (B)_{ij}\Big]$$

The optimum conditions in case of constraints are:

$df = o$, and $d^2f \begin{matrix} < o \\ > o \end{matrix}$ subject to $dF(G) = 0$, for a $\begin{matrix} \text{maximum} \\ \text{minimum} \end{matrix}$, if the constraint is $F(G) = B$, B being a constant.
From (86) we see that d^2f is a quadratic form in vec ΔG. The results on quadratic forms of section 2.3 can be applied. In case $\frac{\partial^2 f}{\partial G^2}$ is $\begin{matrix} \text{positive} \\ \text{negative} \end{matrix}$ definite in a constrained $\begin{matrix} \text{minimization} \\ \text{maximization} \end{matrix}$ problem, there is no need to consider the constraint $dF(G) = 0$. [1])

4.9. Some applications

1. To find the minimizing B for $\log|W|$ of application 5 in section 4.5. Clearly $d \log|W| = o$, so $(X'XB - X'Y)W^{-1} = 0$, which leads to $B = (X'X)^{-1}X'Y$.
Further: $d^2 \log|W| = 2 \text{ tr } X'X(dB)W^{-1}(dB)' = 2 \text{ tr}(XdB)'W^{-1}(XdB)$ in the extremum, which is a positive definite quadratic form, if X has order (m,n) and the rank of X is n.

2. Minimize tr $A'AV$ with respect to A, under the condition: $AX=I$. A is of order (m,n), V is positive definite and symmetric, X has rank m. [2])
The formulation of the problem is:
Minimize $f = \text{tr } A'AV - \text{tr } M(AX-I)$, M being a Lagrange matrix multiplier.
The first-order condition is:
$df = d \text{ tr } A'AV - d \text{ tr } M(AX-I) = \text{tr}(2AV - M'X')(dA)' = o$,

[1]) See Allen (1), p. 495 et seq.
[2]) For the formulation of the problem see Johnston (15), p. 180

so $2AV - M'X' = 0$

Postmultiplication by $V^{-1}X$ leads to: $2I = M'X'V^{-1}X$, so $M' = 2(X'V^{-1}X)^{-1}$.

Substituting this into $2AV - M'X' = 0$, leads to:

$$2AV = 2(X'V^{-1}X)^{-1}X'$$

$$\text{so: } A = (X'V^{-1}X)^{-1}X'V^{-1}$$

The second-order condition is:

$$d^2f = d\ \mathrm{tr}(AV + AV' - M'X')(dA)' = 2\ \mathrm{tr}(dA)V(dA)' > 0.$$

It is obvious that $\mathrm{tr}(dA)V(dA)'$ is a sum of positive definite quadratic forms. This establishes the minimum.

3. To minimize $2\ \mathrm{tr}\ Q^2 + (1 - \mathrm{tr}\ Q)^2$ with respect to Q under the condition: $QX = 0$.

X is a matrix of order (m,n), with rank n. Q is symmetric. [1])

The formulation of the problem is:

Minimize $f = 2\ \mathrm{tr}\ Q^2 + (1 - \mathrm{tr}\ Q)^2 - \mathrm{tr}\ MQX$

The first-order condition is:

$df = 2\ \mathrm{tr}(dQ^2) - 2(1 - \mathrm{tr}\ Q)\ \mathrm{tr}(dQ) - \mathrm{tr}\ M(dQ)X =$

$= 4\ \mathrm{tr}\ QdQ - 2(1 - \mathrm{tr}\ Q)\mathrm{tr}\ dQ - \mathrm{tr}\ XMdQ =$

$= \mathrm{tr}\ 4Q - 2(1 - \mathrm{tr}\ Q) - XM\ dQ = 0$, so: $4Q - 2(1 - \mathrm{tr}\ Q)I = XM$

Premultiplication by X' leads to: $X'XM = -2(1 - \mathrm{tr}\ Q)X'$, so

$M = -2(1 - \mathrm{tr}\ Q)(X'X)^{-1}X'$.

We substitute this result into:

$$4Q - 2(1 - \mathrm{tr}\ Q)I = XM, \text{ and get}$$

$$4Q - 2(1 - \mathrm{tr}\ Q)I = -2(1 - \mathrm{tr}\ Q)X(X'X)^{-1}X'$$

This gives: $4\ \mathrm{tr}\ Q - 2(1 - \mathrm{tr}\ Q)m + 2(1 - \mathrm{tr}\ Q)n = 0$, so

$$\mathrm{tr}\ Q = \frac{m-n}{2+m-n}$$

Finally we find: $Q = \frac{1}{2+m-n}\left[I - X(X'X)^{-1}X'\right]$

The second-order condition is:

$d^2f = 4\ \mathrm{tr}(dQ)^2 + 2(\mathrm{tr}\ dQ)^2 > 0$

This condition is fulfilled because dQ is symmetric.

4. Write the (n,m) matrix Z as

$$Z = tc' + W,$$

and determine t and c such that $\mathrm{tr}\ WW'$ is minimum. [2])

Clearly: $d\ \mathrm{tr}\ WW' = d(\mathrm{tr}\ ZZ' - 2t'Zc + t't.c'c) =$

$= -2(dt)'Zc - 2t'Z(dc) + 2(dt)'t.c'c + 2t't(dc)'c =$

$= 2(dt)'(-Zc + c'c.t) + 2(dc)'(-Z't + t't.c) = 0$

So: $Zc = c'c.t$

$Z't = t't.c$ are the first-order minimum conditions. They

[1]) See Theil and Schweitzer (24), p. 19

[2]) This problem arises in Principal components analysis. See Kloek (16), pp. 52, 53 and 54 for a lengthy proof in two stages.

lead to:

$$Z\left\{\frac{1}{t't}Z't\right\} = c'c.t \text{ or } ZZ't = c'c.t't.t$$

Minimizing t is the characteristic column vector, associated with the largest characteristic root c'c ($t't = 1$).

That we need the largest characteristic root follows easily from:

$\mathrm{tr}\ WW' = \mathrm{tr}\ ZZ' - c'c$, using $Zc = c'c.t$ and $t't = 1$.

5. To minimize [1]

$$|S|$$

with respect to all α_i and β_i, $(i = 1 \ldots G)$ and to determine the minimizing α_i and β_i values, where $S = (u_1 \ldots u_G)'(u_1 \ldots u_G)$,

$$u_i = y_i - (Y_i\alpha_i + Z_i\beta_i),$$

$$U = (u_1 \ldots u_G).$$

The minimization goes as follows:

$$d|S| = |S|\ \mathrm{tr}(S)^{-1}dS =$$

$$= |S|\ \mathrm{tr}(S)^{-1}\left\{U'd(u_1 \ldots u_G) + d(u_1 \ldots u_G)'U\right\} =$$

$$= -2|S|\ \mathrm{tr}\ U(S)^{-1}\begin{bmatrix} d\alpha_1'Y_1' + d\beta_1'Z_1' \\ \vdots \\ d\alpha_G'Y_G' + d\beta_G'Z_G' \end{bmatrix} =$$

$$= -2|S|\ \mathrm{tr}\ U(S)^{-1}_{\cdot 1}\left\{d\alpha_1'Y_1' + d\beta_1'Z_1'\right\} - \ldots -$$

$$-2|S|\ \mathrm{tr}\ U(S)^{-1}_{\cdot G}\left\{d\alpha_G'Y_G' + d\beta_G'Z_G'\right\} =$$

$$= -2|S|\ \mathrm{tr}\ Y_1'U(S)^{-1}_{\cdot 1}d\alpha_1' - 2|S|\ \mathrm{tr}\ Z_1'U(S)^{-1}_{\cdot 1}d\beta_1' - \ldots -$$

$$-2|S|\ \mathrm{tr}\ Y_G'U(S)^{-1}_{\cdot G}d\alpha_G' - 2|S|\ \mathrm{tr}\ Z_G'U(S)^{-1}_{\cdot G}d\beta_G' = 0$$

So $Y_i'U(S)^{-1}_{\cdot i} = 0$

$$Z_i'U(S)^{-1}_{\cdot i} = 0 \text{ for all } i.$$

A little rearranging will give then:

$$\begin{bmatrix} Y_i' \\ Z_i' \end{bmatrix}(y_1 \ldots y_G)\ (S)^{-1}_{\cdot i} =$$

$$\begin{bmatrix} Y_i' \\ Z_i' \end{bmatrix}\left((Y_1\ Z_1)\begin{bmatrix}\alpha_1 \\ \beta_1\end{bmatrix} \ldots (Y_G\ Z_G)\begin{bmatrix}\alpha_G \\ \beta_G\end{bmatrix}\right)(S)^{-1}_{\cdot i}$$

Collecting all expressions, we get

[1]) For the formulation of this problem (the derivation of three-stage least squares), see Chow (7), p. 548).

$$\begin{pmatrix} \begin{bmatrix} Y_1' \\ Z_1' \end{bmatrix} (y_1 \ \dots \ y_G)(S)^{-1}_{\cdot 1} \\ \vdots \\ \begin{bmatrix} Y_G' \\ Z_G' \end{bmatrix} (y_1 \ \dots \ y_G)(S)^{-1}_{\cdot G} \end{pmatrix} =$$

$$= \begin{pmatrix} s^{11} \begin{bmatrix} Y_1' \\ Z_1' \end{bmatrix} (Y_1 \quad Z_1) \dots\dots s^{G1} \begin{bmatrix} Y_1' \\ Z_1' \end{bmatrix} (Y_G \quad Z_G) \\ \vdots \qquad\qquad\qquad \vdots \\ s^{1G} \begin{bmatrix} Y_G' \\ Z_G' \end{bmatrix} (Y_1 \quad Z_1) \dots\dots s^{GG} \begin{bmatrix} Y_G' \\ Z_G' \end{bmatrix} (Y_G \quad Z_G) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_G \\ \beta_G \end{pmatrix}$$

s^{kl} stands of course for the kl^{th} element of S^{-1}.
The right-hand side column vector can now be determined. It is the vector of three stage least squares coefficients.

6. Linear programming.
Let the problem be:

$$\text{maximize } b'x$$
$$\text{subject to } Ax \leq c$$
$$x \geq o$$

How can we allow for variations in c, b or A? If we assume that the basic solution variables of the original solution will still be basic in the perturbated problem, we can derive an answer to the problem in the following way:
Let the initial simplex tableau be:

$$\begin{pmatrix} P & Q & o & c \\ -b_1' & -b_2' & 1 & \alpha \end{pmatrix}$$ and the final one:

$$\begin{pmatrix} I & P^{-1}Q & o & P^{-1}c \\ o & b_1'P^{-1}Q - b_2' & 1 & b_1'P^{-1}c + \alpha \end{pmatrix}$$

Clearly P and $-b_1'$ correspond to final basic variables, Q and $-b_2'$ correspond to final non-basic variables. This was the original problem.
In case of the perturbated problem:

$$\text{maximize } (b+\Delta b)'x$$
$$\text{subject to } Ax \leq c+\Delta c$$
$$x \geq o$$

the final simplex tableau will be:

$$\begin{pmatrix} I & (P+\Delta P)^{-1}(Q+\Delta Q) & o & (P+\Delta P)^{-1}(c+\Delta c) \\ o & (b_1+\Delta b_1)'(P+\Delta P)^{-1}(Q+\Delta Q) - (b_2+\Delta b_2)' & 1 & (b_1+\Delta b_1)'(P+\Delta P)^{-1}(c+\Delta c) + \alpha \end{pmatrix}$$

Approximate values of the basic solution variables will be:

$\left[I - P^{-1}\Delta P + (P^{-1}\Delta P)^2 - \ldots\right] P^{-1}(c + \Delta c)$, if

$\|P^{-1}\Delta P\| < 1$ or $\|P^{-1}\| . \|\Delta P\| < 1$

For a feasible solution $(P+\Delta P)^{-1}(c+\Delta c)$ should be positive.
Approximate value of the function will be:

$(b_1 + \Delta b_1)' \left[I - P^{-1}\Delta P + (P^{-1}\Delta P)^2 - \ldots\right] P^{-1}(c + \Delta c) + \alpha$, according to the same conditions.

Necessary and sufficient for the optimum is:

$(b_1 + \Delta b_1)'(P + \Delta P)^{-1}(Q + \Delta Q) - (b_2 + \Delta b_2)' \geqslant 0$

Approximated to first-order changes:

$b_1' \left[Q + \Delta Q - P^{-1}(\Delta P)Q\right] + (\Delta b_1)'Q \geqslant b_2' + (\Delta b_2)'$

is necessary and sufficient for the optimum.

The approximated conditions should be handled with care. It is possible that approximated conditions hold, whereas the complete conditions don't. The reader can easily judge this by comparing the values of the original problem and the approximated values of the perturbated problem, and seeing what the orders of the changes are.

7. Least squares

The least-squares estimate for B in

$$Y = XB \text{ is } B = (X'X)^{-1}X'Y$$

How will B be affected by moderate changes in Y and X?

Approximating ΔB up to first-order changes, we find:

$$\Delta B = dB = d\left[(X'X)^{-1}X'Y\right] =$$

$$= -(X'X)^{-1}\left[(dX)'X + X'dX\right](X'X)^{-1}X'Y +$$

$$+ (X'X)^{-1}(dX)'Y + (X'X)^{-1}X'dY =$$

$$= -(X'X)^{-1}(\Delta X)'XB - (X'X)^{-1}X'(\Delta X)B + (X'X)^{-1}(\Delta X)'Y +$$

$$+ (X'X)^{-1}X'\Delta Y$$

From this follows: $B + \Delta B =$

$$= (X'X)^{-1}\left[X'X - (\Delta X)'X - X'(\Delta X)\right]B + (X'X)^{-1}\left[(\Delta X)'Y + X'\Delta Y\right] + \ldots$$

8. Find the Jacobian of the linear transformation

$$F(G) = KGL,$$

K being square of order m, L being square of order n, G having order (m,n).

First of all, $d(KGL)_{ij} = \left\{K(dG)L\right\}_{ij}$,

so $\frac{\partial \text{vec } KGL}{\partial \text{vec } G} = L \otimes K'$,

and $J = \left|L \otimes K'\right|^{-1} = |L|^{-m} \times |K|^{-n}$

9. Minimize

$$\text{tr } AA' - 2 \text{ tr } A$$

with respect to A, subject to $A'X = 0$ ')

X has order (n,m), and rank m.

We define

$$f = \text{tr } AA' - 2 \text{ tr } A - \text{tr } GA'X,$$

G being a Lagrangean matrix multiplier.

Then

$df = \text{tr}(dA)A' + \text{tr } A(dA)' - 2 \text{ tr } dA - \text{tr } G(dA)'X =$

$= \text{tr}(2A - 2I - XG)(dA)' = o,$

so: $2A - 2I - XG = 0.$

Premultiplying both sides by X' leads to

$$-2X' - X'XG = 0, \text{ and hence}$$

$$G = -2(X'X)^{-1}X'$$

Substituting this G value into $2A - 2I = XG = 0$:

$$A = I - X(X'X)^{-1}X'$$

The second-order differential is:

$$d^2f = 2 \text{ tr}(dA)(dA)',$$

which is positive for any $dA \neq 0$.

This establishes the minimum.

10. Maximize

$$\text{tr } J'A, \text{ with respect to A}$$

and subject to $A'X = 0$ and $A'A = I$. ')

X again has order (n,m) and rank m. $J = \begin{pmatrix} 0 \\ I \end{pmatrix}$, 0 having order $(m,n-m)$, I having order $(n-m,n-m)$. We define:

$$f = \text{tr } J'A - \text{tr } G_1A'X - \text{tr } G_2(A'A - I),$$

G_1 and G_2 being Lagrangean matrix multipliers.

Then:

$df = \text{tr } J'(dA) - \text{tr } G_1(dA)'X - \text{tr } G_2(dA)'A - \text{tr } G_2A'dA =$

$= \text{tr}(J - XG_1 - AG_2 - AG_2')(dA)' = o,$ which leads to

$J - XG_1 - AG_2 - AG_2' = 0$

Multiplication by X' gives:

$$G_1 = (X'X)^{-1}X'J$$

Hence $J - X(X'X)^{-1}X'J - A(G_2 + G_2') = 0$

Premultiplication by A' leads to:

$$A'J = G_2 + G_2'$$

Substituting this expression for $G_2 + G_2'$ into

$J - X(X'X)^{-1}X'J - A(G_2 + G_2') = 0$:

$\left[I - X(X'X)^{-1}X' - AA'\right]J = 0$

Moreover $A'J = J'A$, because $A'J = G_2 + G_2'$ is symmetric.

The reader is referred to Theil's paper for the determination of A. Suffice it here to state that $A = \begin{pmatrix} A \\ A_1 \end{pmatrix}$, with A_1 having

') See Theil (26), p. 1067 et seq.

positive characteristic roots. So $A'J = \begin{pmatrix} A_0' & A_1' \end{pmatrix}\begin{pmatrix} 0 \\ I \end{pmatrix} = A_1'$

We shall focus attention on the second-order differential, d^2f, now:

$d^2f = -\mathrm{tr}(dA)(G_2 + G_2')(dA)' = -\mathrm{tr}(dA)A'J(dA)' = -\mathrm{tr}(dA)A_1'(dA)'$, which is negative.

This establishes the maximum.

11. Minimize

$$f = \mathrm{tr}(C\underline{O}C' - 2C\underline{O}_* + \underline{O}_{**}) + 2\,\mathrm{tr}\,(CF - F_*)N$$

with respect to C. $\underline{O}$ is symmetric, and positive definite. [1])

$df = d\,\mathrm{tr}(C\underline{O}C' - 2C\underline{O}_* + \underline{O}_{**}) + 2d\,\mathrm{tr}\,(CF - F_*)N =$

$= \mathrm{tr}\,d(C\underline{O}C' - 2C\underline{O}_* + \underline{O}_{**}) + 2\,\mathrm{tr}\,d\,(CF - F_*)N =$

$= \mathrm{tr}\left[(dC)\underline{O}C' + C\underline{O}(dC)' - 2(dC)\underline{O}_*\right] + 2\,\mathrm{tr}\,(dC)FN =$

$= 2\,\mathrm{tr}(\underline{O}C' - \underline{O}_* + FN)dC = 0.$

So $\underline{O}C' - \underline{O}_* + FN = 0$

Further $CF - F_* = 0$ (this was in fact a side relation)

$d^2f = 2\,d\,\mathrm{tr}(\underline{O}C' - \underline{O}_* + FN)dC =$

$= 2\,\mathrm{tr}\,\underline{O}(dC)'dC =$

$= 2\,\mathrm{tr}(dC)'\underline{O}dC > 0.$

12. A^{-1} can be defined by the relation

$$x'A^{-1}x = \max_y (2x'y - y'Ay) \qquad [2])$$

if A is positive definite.

Proof: Define $f = 2x'y - y'Ay$

Maximizing f amounts to setting

$df = 2x'dy - 2y'Ady = 0$

$\therefore\ x' = y'A$ and

$y = (A')^{-1}x$

Furthermore $d^2f = -2(dy)'Ady < 0$, because A is positive definite. We can conclude that

$f_{max} = 2x'(A')^{-1}x - x'A^{-1}A(A')^{-1}x = x'A^{-1}x$

13. Let $f = \log(a'g)$, a being a column vector constant. Approximate Δf.

$df = \frac{1}{a'g}\,a'dg$, so $\frac{\partial f}{\partial g} = \frac{1}{a'g}\,a$

$d\,\frac{\partial f}{\partial g} = -\frac{1}{(a'g)^2}(a'dg)a = -\frac{1}{(a'g)^2}\,aa'dg$

so: $\frac{\partial^2 f}{\partial g^2} = -\frac{1}{(a'g)^2}aa'$

Approximation according to theorem (86):

$$\Delta f = \frac{1}{a'g}\,a'\Delta g - \frac{1}{2!}\frac{1}{(a'g)^2}\,(\Delta g)'aa'(\Delta g) + \ldots.$$

[1]) See Nagar and Kakwani (18), p.

[2]) See Bellman (4), p. 92.

Sufficient for convergence is: $\left|\frac{a'\Delta g}{a'g}\right| < 1$

Approximation according to theorem (89):

$$\Delta f = \frac{1}{a'g}\, a'\Delta g + \frac{1}{2!}\,\frac{1}{a'(g+\Delta g)}\, a'\Delta g - \frac{1}{2!}\,\frac{1}{a'g}\, a'\Delta g + \ldots. =$$

$$= \tfrac{1}{2}\left\{\frac{1}{a'(g+\Delta g)} + \frac{1}{a'g}\right\} a'\Delta g + \ldots.$$

14. Maximize:

$$w = a'x + b'y + \tfrac{1}{2}(x'Ax + y'By + x'Cy + y'C'x)$$

with respect to x and y, subject to $y - Rx = s$. ')

A and B are symmetric.

We have to maximize unconditionally:

$\overline{w} = a'x + b'y + \frac{1}{2}(x'Ax + y'By + x'Cy + y'C'x) - \alpha'(y - Rx - s)$

Thus $d\overline{w} = a'dx + b'dy + x'Adx + y'Bdy + x'Cdy + y'C'dx - \alpha'dy + \alpha'Rdx = o$,

which leads to $a' + x'A + y'C' + \alpha'R = o$

$$b' + y'B + x'C - \alpha' = o$$

Together with the constraint we have:

$$(x' \quad y' \quad \alpha')\begin{pmatrix} A & C & R' \\ C' & B & -I \\ R & -I & 0 \end{pmatrix} = -(a' \quad b' \quad s')$$

The second order maximum condition is:

$d^2\overline{w} = (dx)'Adx + (dy)'Bdy + (dx)'Cdy + (dy)'C'dx < o$,

subject to $dy - Rdx = o$.

There is no need to apply the principal minor conditions of section 2.3, as is being done by Theil, because we can substitute $dy = Rdx$ into $d^2\overline{w}$:

$d^2\overline{w} = (dx)'(A + R'BR + CR + R'C')dx < o$.

The second-order requirement is therefore:

A + R'BR + CR + R'C' be negative definite.

15. In example 14 we derived the solution for (x' y' α'). From the equation the "optimal reaction functions"

$$x^o = x(R,s;a,b,A,B,C)$$

$$y^o = y(R,s;a,b,A,B,C) \text{ can be derived.}$$

Theil then investigates the problem what the effects are of "infinitesimal changes in the arguments of the optimal reaction functions". The question is answered by differentiating the above-given equation. (We transpose it.)

$$d\begin{pmatrix} A & C & R' \\ C' & B & -I \\ R & -I & 0 \end{pmatrix}\cdot\begin{pmatrix} x^o \\ y^o \\ \alpha^o \end{pmatrix} + \begin{pmatrix} A & C & R' \\ C' & B & -I \\ R & -I & 0 \end{pmatrix} d\begin{pmatrix} x \\ y \\ \alpha \end{pmatrix} = -d\begin{pmatrix} a \\ b \\ s \end{pmatrix},$$

') See for this problem Theil (25), p. 37 et seq.

which gives:

$$\begin{pmatrix} dA & dC & (dR)' \\ (dC)' & dB & 0 \\ dR & 0 & 0 \end{pmatrix} \begin{pmatrix} x^o \\ y^o \\ \alpha^o \end{pmatrix} + \begin{pmatrix} A & C & R' \\ C' & B & -I \\ R & -I & 0 \end{pmatrix} \begin{pmatrix} dx \\ dy \\ d\alpha \end{pmatrix} = - \begin{pmatrix} da \\ db \\ ds \end{pmatrix},$$

from which follows:

$$\begin{pmatrix} dx \\ dy \\ d\alpha \end{pmatrix} = - \begin{pmatrix} (xx) & (xy) & (x.) \\ (yx) & (yy) & (y.) \\ (.x) & (.y) & (..) \end{pmatrix} \begin{pmatrix} da + (dA)x^o + (dC)y^o + (dR)'\alpha^o \\ db + (dC)'x^o + (dB)y^o \\ ds + (dR)x^o \end{pmatrix}$$

The peculiar (.,x,y) matrix is the inverse of $\begin{pmatrix} A & C & R' \\ C' & B & -I \\ R & -I & 0 \end{pmatrix}$;

the submatrices of the inverse have the same orders as the corresponding submatrices of the original matrix.

The author then gives the partial derivatives of the optimal reaction functions

$\frac{\partial x_h}{\partial s_j}$, $\frac{\partial y_i}{\partial s_j}$, $\frac{\partial x_h}{\partial r_{jk}}$; $\frac{\partial y_i}{\partial r_{jk}}$, $\frac{\partial x_h}{\partial a_k}$, $\frac{\partial x_h}{\partial b_j}$, which result is quite obscure.

Applying theorem (72) we get the compact expressions:

$$\frac{\partial x}{\partial \text{vec } R} = - x^o \otimes (x.)' - \begin{pmatrix} \alpha^o \otimes (xx)'_{1.} \\ \vdots \\ \alpha^o \otimes (xx)'_{m.} \end{pmatrix}$$

$$\frac{\partial x}{\partial s} = -I \otimes (x.)' = - (x.)'$$

$$\frac{\partial x}{\partial a} = -I \otimes (xx)' = - (xx)'$$

$$\frac{\partial x}{\partial b} = -I \otimes (xy)' = - (xy)'$$

$$\frac{\partial x}{\partial \text{vec } C} = - y^o \otimes (xx)' - \begin{pmatrix} x^o \otimes (xy)'_{1.} \\ \vdots \\ x^o \otimes (xy)'_{n.} \end{pmatrix}$$

$$\frac{\partial y}{\partial \text{vec } R} = - x^o \otimes (y.)' - \begin{pmatrix} \alpha^o \otimes (yx)'_{1.} \\ \vdots \\ \alpha^o \otimes (yx)'_{m.} \end{pmatrix}$$

$$\frac{\partial y}{\partial s} = - I \otimes (y.)' = - (y.)'$$

$$\frac{\partial y}{\partial a} = - I \otimes (yx)' = - (yx)'$$

$$\frac{\partial y}{\partial b} = - I \otimes (yx)' = - (yx)'$$

$$\frac{\partial y}{\partial \text{vec } C} = - y^{\circ} \otimes (yx)' - \begin{pmatrix} x^{\circ} \otimes (yy)'_{1\cdot} \\ \vdots \\ x^{\circ} \otimes (yy)'_{n\cdot} \end{pmatrix}$$

The derivation of $\frac{\partial x}{\partial \text{vec } A}$, $\frac{\partial x}{\partial \text{vec } B}$, $\frac{\partial y}{\partial \text{vec } A}$ and $\frac{\partial y}{\partial \text{vec } B}$ is more complicated because A and B are <u>symmetric</u>.

If A and B were <u>nonsymmetric</u>, we would have:

$$\frac{\partial x}{\partial \text{vec } A} = - x^{\circ} \otimes (xx)'$$

$$\frac{\partial x}{\partial \text{vec } B} = - y^{\circ} \otimes (xy)'$$

$$\frac{\partial y}{\partial \text{vec } A} = - x^{\circ} \otimes (yx)'$$

$$\frac{\partial y}{\partial \text{vec } B} = - y^{\circ} \otimes (yy)'$$

The derivatives matrices for <u>symmetric</u> A and B can be found by combining rows, which correspond to now <u>identical</u> elements of A and B.

An example can make this clear:

Let $\frac{\partial u}{\partial \text{vec } W} = \begin{pmatrix} \frac{\partial u_1}{\partial w_{11}} & \frac{\partial u_2}{\partial w_{11}} \\ \frac{\partial u_1}{\partial w_{21}} & \frac{\partial u_2}{\partial w_{21}} \\ \frac{\partial u_1}{\partial w_{12}} & \frac{\partial u_2}{\partial w_{12}} \\ \frac{\partial u_1}{\partial w_{22}} & \frac{\partial u_2}{\partial w_{22}} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \\ s_{31} & s_{32} \\ s_{41} & s_{42} \end{pmatrix}$

If $W' = W$, $\frac{\partial u}{\partial \text{vec } W} = \begin{pmatrix} \frac{\partial u_1}{\partial w_{11}} & \frac{\partial u_2}{\partial w_{11}} \\ \frac{\partial u_1}{\partial w_{12}} & \frac{\partial u_2}{\partial w_{12}} \\ \frac{\partial u_1}{\partial w_{12}} & \frac{\partial u_2}{\partial w_{12}} \\ \frac{\partial u_1}{\partial w_{22}} & \frac{\partial u_2}{\partial w_{22}} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} + s_{31} & s_{22} + s_{32} \\ s_{21} + s_{31} & s_{22} + s_{32} \\ s_{41} & s_{42} \end{pmatrix}$

Formal expressions for $\frac{\partial x}{\partial \text{vec } A}$, $\frac{\partial x}{\partial \text{vec } B}$, $\frac{\partial y}{\partial \text{vec } A}$, $\frac{\partial y}{\partial \text{vec } B}$, which incorporate the above-mentioned transformations, can easily be derived. This will not be done here.

4.10 A sequel on the generalized inverse

In most optimization problems which have been worked out in section 4.9, adequate rank assumptions had to be made in order to be able to reach a solution.
In application 1 we assumed that $(X'X)^{-1}$ existed, in application 2 we assumed that $(X'V^{-1}X)^{-1}$ existed, and so on.
What can be done in cases where these assumptions are not warranted?
The " generalized inverse" as developed by, among others, C.R. Rao can bring relief here. ') Some of Rao's results will be reported here, and applied.
The formal definition of a generalized inverse is:

> Consider an (m,n) matrix of any rank. A generalized inverse of A is an (n,m) matrix A^- such that for any vector y for which $Ax = y$ is a consistent equation, $x = A^-y$ is a solution.

It can be proved that the existence of A^- is warranted if $AA^-A = A$ and vice versa:

> Suppose A^- exists. Choose $y = A_{.i}$. The equation $Ax = A_{.i}$ is clearly consistent, and hence $x = A^-A_{.i}$ is a solution, that is $AA^-A_{.i} = A_{.i}$ for all i, so $AA^-A = A$.
>
> Conversely if A^- is such that $AA^-A = A$ and $Ax = y$ is a consistent equation, then $AA^-Ax = Ax$ or $A(A^-y) = y$. Hence $x = A^-y$ is a solution.

Clearly necessary and sufficient that $Ax = y$ is consistent is that $AA^-y = y$.
$AA^-y = y \rightarrow A^-y$ is a solution of $Ax = y$
$Ax = y$ is consistent $\rightarrow AA^-Ax = AA^-y$ or $Ax = AA^-y$ of $y = AA^-y$.
The generalized inverse A^- is not necessarily unique. One way of constructing it is the following:

> Given A, there exist non-singular matrices B and C, such that $BAC = D$, or $A = B^{-1}DC^{-1}$, where D is a matrix with elements d_{ij}, such that $d_{ij} = o$ for $i \neq j$. Define $(D^-)'$ as the matrix obtained by replacing the non-zero elements of D by their reciprocals.
> Then $DD^-D = D$
> And $A^- = CD^-B$, because $AA^-A =$
> $= B^{-1}DC^{-1}CD^-BB^{-1}DC^{-1} = B^{-1}DD^-DC^{-1} = B^{-1}DC^{-1} = A$
> Moreover for this A^- : $A^-AA^- = A^-$, because
> $CD^-BB^{-1}DC^{-1}CD^-B = CD^-DD^-B = CD^-B = A^-$

') See Rao (23), p. 263 et seq.

Some more properties follow easily:

1. One choice of $(A')^- = (A^-)'$
2. $A(A'A)^-A'A = A$
3. $A'A(A'A)^-A' = A'$
4. $A(A'A)^-A'$ is symmetric, idempotent and of rank equal to that of A.

This generalized inverse A^- can be applied in all cases where we cannot make appropriate rank assumptions.

Some cases will be dealt with now. (They have also been examined by Rao, but without differential calculus.)

1. Let L be a symmetric positive definite matrix of order (m,m), B be of order (m,k) and u be a column vector of order m

Then $B'x=u$

$\min \quad x'Lx = u'(B'L^{-1}B)^-u$

Proof: Let $f = x'Lx - \alpha'(B'x - u)$,

α' being a Lagrange row vector multiplier.

Then $df = 2x'Ldx - \alpha'B'dx = o$,

so $Lx = B\alpha$

Premultiplication by $B'L^{-1}$ leads to:

$u = B'x = B'L^{-1}B\alpha$, and subsequently to

$\alpha = (B'L^{-1}B)^-u$

Substituting back we get:

$Lx = B(B'L^{-1}B)^-u$, so

$x = L^- B(B'L^{-1}B)^-u$ and $x'Lx = u'(B'L^{-1}B)^- u$

2. Minimize $x_1'x_1$ subject to $B'x = u$, where $x' = (x_1' \; x_2')$

Procedure: $f = x_1'x_1 - \alpha'(B_1' \; B_2')\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where $B' = (B_1' \; B_2')$

Then $df = 2x_1'dx_1 - \alpha'B_1'dx_1 - \alpha'B_2'dx_2 = o$

which leads to $\quad 2x_1 = B_1\alpha$

$\quad o = B_2\alpha$

Premultiplying the first equation by B' gives:

$$\begin{pmatrix} B_1'B_1 & B_2' \\ B_2 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{2}\alpha \\ x_2 \end{pmatrix} = \begin{pmatrix} u \\ o \end{pmatrix}, \text{ using } B_1'x_1 + B_2'x_2 = u,$$

and the solution $x_1 = B_1C_1u$

$x_2 = C_2'u$, if

we define $\begin{pmatrix} B_1'B_1 & B_2' \\ B_2 & 0 \end{pmatrix}^- = \begin{pmatrix} C_1 & C_2 \\ C_2' & C_3 \end{pmatrix}$

This gives the constrained minimum value of

$x_1'x_1 = u'C_1B_1'B_1C_1u = u'C_1(u - B_2'C_2'u) = u'C_1u$, because

$$\begin{pmatrix} B_1'B_1 & B_2' \\ B_2 & 0 \end{pmatrix}\begin{pmatrix} C_1 & C_2 \\ C_2' & C_3 \end{pmatrix}\begin{pmatrix} u \\ o \end{pmatrix} = \begin{pmatrix} u \\ o \end{pmatrix} \text{ (consistency requirement)}$$

4.11

Exercises

1. The linear programming problem:

Maximize $f = 3x + 2y$

subject to $x + 2y \leq 5$

$x + 3y \leq 6$

$x, y \geq 0$

has solution: $x = 5 \quad y = 0 \quad f = 15$

Approximate then the solution for:

Maximize $3{,}1x + 2.2y$

subject to $1.2x + 3y \leq 5.3$

$1.3x + 4y \leq 6.1$

$x, y \geq 0$

2. The solution of

$$\mathbf{A}x = b$$

with $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix}$

$b = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

is $x = \begin{pmatrix} 3 \\ -0.5 \end{pmatrix}$

Approximate the solution of:

$$\overline{\mathbf{A}}x = \overline{b}$$

with $\overline{\mathbf{A}} = \begin{pmatrix} 1.1 & 2.1 \\ 2.1 & 6.2 \end{pmatrix}$

$\overline{b} = \begin{pmatrix} 2.1 \\ 3.1 \end{pmatrix}$

3. Find the marginal impact of changes in input coefficients and final demand upon production:

$\frac{\partial p}{\partial \text{vec } \mathbf{A}}$ and $\frac{\partial p}{\partial c}$, if

$p = (I-A)^{-1}c$ is the relationship between

production p,

final demand c,

input coefficients matrix $\mathbf{A}$.

4. Determine $\frac{\partial \alpha_i}{\partial G}$ in the 'point' G_o,

$$G_o = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

5. Approximate the characteristic roots of

$$G = \begin{pmatrix} 2.2 & \sqrt{2} \\ \sqrt{2} & 1.1 \end{pmatrix}$$

if

$$G_o = \begin{pmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix}$$

has characteristic roots $\alpha_1 = 0$ and $\alpha_2 = 3$ and associated characteristic column vectors $\bar{u}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ \sqrt{2} \end{pmatrix}$ and $u_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix}$

6. Find the Jacobian of the transformation

$$y = Ax$$

7. Find the Jacobian of the transformation

$$Y = AXA',$$

X being symmetric.

8. Approximate

$$\log(x'Ax),$$

$$\text{with } A = \frac{1}{100}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$$

$$\text{for } x = \begin{pmatrix} 10.1 \\ 20.3 \end{pmatrix},$$

if the value for $x = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$ is 3.36730

9. Minimize $\operatorname{tr}|W|$ with respect to B,

where $W = E'E$

$E = Y - XB$

W positive definite

X has order (m,n) and rank n.

10. Expand $|I + tA|$

11. Minimize $x'Ax$ with respect to x, subject to $a'x = 1$.
A is positive definite and symmetric.

12. To find

$\min_{B} \operatorname{tr} \frac{AB}{n}$ with respect to B, subject to $|B| = 1$ and A positive definite,
A and B being symmetric,
n being being the order of A and B. [1])

13. Maximize

$$f = \log|C| + \operatorname{tr} AC^{-1} \quad \text{with respect to L and V,}$$

for $C = LL' + V$

V diagonal

C and A symmetric. [2])

14. Establish an approximation formula for

$\Delta|A|$, if A has order 4.

Realise that all $d^i|A| = 0$ for $i > 4$.

[1]) See Bellman (4), p.131, for a slightly different problem, requiring $B \neq 0$.
[2]) See Lawley (7), p. 11 et seq.

15. Approximate $\begin{vmatrix} 10.2 & 5 & 6.1 \\ 3.1 & 7.1 & 6.9 \\ 10.2 & 11.2 & 12.3 \end{vmatrix}$,

if $\begin{vmatrix} 10 & 5 & 6 \\ 3 & 7 & 7 \\ 10 & 11 & 12 \end{vmatrix} = 18$, and

$$\begin{pmatrix} 10 & 5 & 6 \\ 3 & 7 & 7 \\ 10 & 11 & 12 \end{pmatrix}^{-1} = \frac{1}{18}\begin{pmatrix} 7 & 6 & -7 \\ 34 & 60 & -52 \\ -37 & -60 & 55 \end{pmatrix},$$

up to second-order terms.

16. Prove that the quadratic form

$$f = x'Cx$$

has a minimum, if C is semi-positive definite.

17. Given $\frac{dX}{dt} = AX$, to determine $\frac{d|X|}{dt}$ [1])

Answer: $\frac{d|X|}{dt} = \operatorname{tr}\left(\frac{\partial|X|}{\partial X}\right)' \frac{dX}{dt} =$

$= \operatorname{tr}|X| X^{-1} \frac{dX}{dt} = |X| \operatorname{tr} X^{-1} AX = |X| \operatorname{tr} A$, because of theorems (81) and (78).

18. Solve the differential equation $\frac{d|X|}{dt}$, if $\frac{dX}{dt} = AX$ and $X_0 = I$. [2])

Answer: From ex. (17) we conclude:

$$|X| = ce^{t \operatorname{tr} A}.$$

Because $|X_0| = 1$, we finally get: $|X| = e^{t \operatorname{tr} A}$

19. Prove that the system of differential equations $\frac{dX}{dt} = AX$, with $X = C$ has solution $X = e^{tA}C$.

X, A and C are matrices, (A and C constants), t is a scalar parameter.

Proof: $\frac{dX}{dt} = \frac{d\, e^{tA}C}{dt} = \frac{d(I+tA+\frac{t^2}{2!}A^2+\ldots)C}{dt} =$

$= (A+tA^2+\frac{t^2}{2!}A^3+\ldots)C = A(I+tA+\frac{t^2}{2!}A^2+\ldots)C = Ae^{tA}C = AX$

20. To find $\frac{\partial|A-kI|}{\partial k}$ and $\frac{\partial^2|A-kI|}{k^2}$

[1]) See Bellman (4), p. 167, 168.
[2]) See Bellman (4), p. 168.

L I T E R A T U R E

1.Allen, R.G.D. (1953), Mathematical Analysis for Economists, Macmillan and Co, London.

2.Allen, R.G.D. (1956), Mathematical Economics, Macmillan and Co, London.

3.Anderson, T.W., (1958), An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York

4.Bellman, R. (1960), Introduction to Matrix Analysis, McGraw-Hill Book Company, New York

5.Bodewig, E. (1956), Matrix Calculus, North-Holland Publishing Company, Amsterdam

6.Burington, R.S. and C.C. Torrance (1939), Higher Mathematics, McGraw-Hill Book Company, New York

7.Chow, G.C. (1964), "A Comparison of Alternative Estimators for Simultaneous Equations", Econometrica, vol. 32, pp. 532 - 553

8.Debreu, G. (1952), "Definite and Semidefinite Quadratic Forms", Econometrica, vol. 20, pp. 295 - 300

9.Deemer, W.L. and I. Olkin (1951), "The Jacobians of Certain Matrix Transformations useful in Multivariate Analysis", Biometrika, vol. 38, pp. 345 - 367

10.Dwyer, P.S. and M.S. Macphail (1948), "Symbolic Matrix Derivatives", Annals of Mathematical Statistics, vol. 19, pp. 517 - 534

11.Enzer, H. (1966), "On a Useful Capital Growth Matrix", Econometrica, vol. 34, pp. 183 - 192.

12.Faddeev, D.K. and V.N. Faddeeva (1963), Computational Methods of Linear Algebra, W.H. Freeman and Company, San Francisco

13.Goldberger, A.S. (1964), Econometric Theory, John Wiley and Sons, New York

14.Hadley, G. (1961), Linear Algebra, Addison-Wesley Publishing Company, Reading, Mass.

15.Johnston, J. (1963), Econometric Methods, McGraw-Hill Book Company, New York

16.Kloek, T. (1966), Indexcijfers, enige methodologische aspecten, Ph.D. Thesis Netherlands School of Economics, Rotterdam

17.Lawley, D.N. and A.E. Maxwell (1963), Factor Analysis as a Statistical Method, Butterworth, London

18.Nagar, A.L. and N.C. Kakwani (1963), " Best Unbiased Prediction from a Simultaneous-Equation Model", Indian Economic Review, vol. 6, pp.

19. Nagar, A.L., (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", Econometrica, vol. 27, pp. 575 - 595

20. Neudecker, H. and C. van de Panne, (1966), "Note on the Asymptotic Standard Errors of Latent Roots of Econometric Equation Systems", Review of the International Statistical Institute, vol. 34, pp. 43 - 47

21. Neudecker, H. (1967), "On Matrix Procedures for Optimizing Differentiable Scalar Functions of Matrices", Statistica Neerlandica, vol. 21, pp. 101 - 107.

22. Olkin, I. (1953), "Note on 'The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis'", Biometrika vol. 40, pp. 43 - 46

23. Rao, C.R. (1966), "Generalized Inverse for Matrices and its Applications in Mathematical Statistics", Research Papers in Statistics, Festschrift for J. Neyman (F.N. David, Editor), John Wiley and Sons, London, pp. 263 - 279

24. Rothenberg, T.J. and C.T. Leenders (1964), "Efficient Estimation of Simultaneous Equation Systems", Econometrica, vol. 32, pp. 57 - 76

25. Theil, H. (1964), Optimal Decision Rules for Government and Industry, North-Holland Publishing Company, Amsterdam

26. Theil, H. (1965), "The Analysis of Disturbances in Regression Analysis", Journal of the American Statistical Association, vol. 60, pp. 1067 - 1079

27. Theil, H. and A.L.M. Schweitzer (1961), "The Best Quadratic Estimator of the Residual Variance in Regression Analysis", Statistica Neerlandica, vol. 15, pp. 19 - 24.

28. Wilks, S.S. (1962), Mathematical Statistics, John Wiley and Sons, New York

29. Theil, H., "Elements of Large-Sample Estimation Theory". Syllabus HT/104 of the Program of Mathematical and Managerial Economics, Netherlands School of Economics (1961).

I N D E X

adjoint matrix, 13
aggregation in input-output analysis, 49
Allen, R.G.D., i, 65
Anderson, T.W., i, 46
approximation, 63
asymptotic expectation, 58
asymptotic standard error, 58
asymptotic variance matrix, 58

Bellman, R. 71, 78, 79
bilinear form, 33, 57
Bodewig, E., i
Burington, R.S., 40

capital growth matrix, 60
Cauchy's inequality, 30
Cayley-Hamilton theorem, 27
characteristic equation, roots and vectors, 22, 56
Chow, G.C., 67
cofactor, 9
compatible matrix and vector norms, 42
convergence, 43
Cramer's rule, 14

Debreu, G.,29
Deemer, W.L., 63
definiteness, 28
derivatives matrix, 54
determinant, 9, 57
diag, 53
diagonalisation, 25
diagonal matrix, 8
difference, 63
difference equation, 38
difference operator, 27, 63
differential, 51
differential equation, 50, 79
Dwyer, P.S., 80

econometric equation system, 58
elementary column operations, 15
elementary row operations, 15
Enzer, H., 60
expectation, 34, 38, 39
expectation operator, 21

Faddeev, D.K., 42
Faddeeva, V.N., 42
fractional matrix powers, 27, 44

generalised inverse, 75
Goldberger, A.S., i, 58, 59

Hadley, G., i, iv
Hessian, 52

idempotent matrix, 8
input-output analysis, 49, 77
inverse matrix, 13
inversion, 17

Jacobian, 62
Johnston, J., i, 65

Kakwani, N.C., 71
Kloek, T., 64, 66
Kronecker matrix product, 4
characteristic roots and vectors, 24
determinant, 24
inverse, 14
product of Kronecker matrix products, 4, 5
rank, 32
trace, 12
transpose, 7

Lagrange multiplier, 65
Lawley, D.N., 78
least squares, 69
Leenders, C.T., 59
limit, 40
linear dependence, 16
linear programming, 68
lower bound, 49

Macphail, M.S., 80
marginal impact, 77
matrix,
addition, 2
exponential, 44
identity, 2
Kronecker multiplication, 4
multiplication, 3
norm, 42, 43
power series, 41
powers, 27
sequence, 40
series, 40
matrix function of a matrix, 54
minor, 9
modulus, 43

Nagar, A.L., 71
Neudecker, H., 59
nilpotent matrix, 8
norm, 42
null matrix, 8

Olkin, I., 63
optimal reaction function, 72
optimization, 64
order, 2
orthogonal, 23

Panne, C. van de, 59
partial derivatives matrix, 51
partitioning, 7
prediction error in input-output analysis, 49
principal components analysis, 66
principal minor, 29

quadratic form, 28, 57
product of quadratic forms, 31

rank, 31
Rao, C.R., 75
reduced form coefficients, 58
Rothenberg, T.J., 59

scalar, 1
scalar function of a matrix, 52

Schweitzer, A.L.M., 66
singular matrix, 9
skew-symmetric matrix, 9
statistical independence, 39
stochastic convergence, 46
submatrix, 7
subordinate matrix norm, 43
symmetric matrix, 9

Taylor's theorem, 63
Theil, H., 66, 70, 72
three-stage least squares, 67
Torrance, C.C., 40
trace, 56
transpose, 6
transposition, 6
triangular matrix, 62

unity matrix, 8
upper bound, 45

variance matrix, 21
vec, 13
vector, 1
    addition, 2
    identity, 2
    multiplication, 6
    norm, 42
    sequence, 40
    series, 40

Wilks, S.S., 46

On matrix procedures for optimizing differentiable scalar functions of matrices

by H. Neudecker)*

Introduction

In the statistical field of econometrics one often meets with optimization problems involving differentiable scalar functions of matrices.

Most frequently occurring scalar functions are *trace*, *latent roots* and *determinant* (or functions of these). Some econometric textbooks present short outlines on differential calculus in matrix notation, in which some building stones for matrix procedures are supplied. I refer to GOLDBERGER [2], JOHNSTON [3] and MALINVAUD [4]. I also want to refer to the articles by DEEMER and OLKIN [1] and OLKIN [6], which gave the most complete enumeration of building stones so far. DEEMER and OLKIN concentrated on determinants (Jacobians), and did not tackle matrix functions like latent roots and trace, which are frequently occurring in econometrics.

It is my contention that these building stones are not adequately used in econometrics. Most authors still resort to cumbersome methods consisting of writing out typical elements of the scalar function, and combining these later on into matrices, instead of using compact matrix methods.

Second-order conditions for optima are very seldom dealt with. If they are considered, very indistinct procedures are applied.

In this note I shall collect the relevant results for optimization, spread in literature. I shall apply these to the three subclasses of differentiable scalar functions (trace, latent root, determinant), and extend the exposition with some procedures, not to be found in the literature (as far as I know!), but being so simple that I don't want to claim any great originality! Only in the cases where I actually *prove* results, I think to say something new.

The basic elements

If f is a differentiable scalar function of the elements g_{ij} of a matrix G, then the first-order optimum condition is:

$$\mathrm{d}f = \sum^{ij} \frac{\partial f}{\partial g_{ij}} \mathrm{d}g_{ij} = 0,$$

the second-order optimum condition is:

$$\mathrm{d}^2 f = \sum^{ijkl} \frac{\partial^2 f}{\partial g_{ij} \partial g_{kl}} \mathrm{d}g_{ij} \mathrm{d}g_{kl} \begin{cases} > 0 \text{ (minimum)} \\ < 0 \text{ (maximum).} \end{cases}$$

*) University of Birmingham.

Sometimes we need an expression for the partial derivatives $\partial f/\partial g_{ij}$, in order to be able to proceed to df. In some cases we can dispense with them – unless we are specifically interested in them –; they are a by-product then of df. The $\partial f/\partial g_{ij}$ will be assembled in the matrix $\partial f/\partial G$:

$$\frac{\partial f}{\partial G} = \begin{pmatrix} \frac{\partial f}{\partial g_{11}} & \cdots & \frac{\partial f}{\partial g_{1m}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial g_{n1}} & \cdots & \frac{\partial f}{\partial g_{nm}} \end{pmatrix}$$

The matrix of second-order partial derivatives

$$\left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right)$$

is a by-product of d^2f; its definition is:

$$\left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right) = \begin{bmatrix} \frac{\partial^2 f}{\partial g_{11}^2} & \cdots & \frac{\partial^2 f}{\partial g_{11}\partial g_{1m}} & \cdots & \frac{\partial^2 f}{\partial g_{11}\partial g_{n1}} & \cdots & \frac{\partial^2 f}{\partial g_{11}\partial g_{nm}} \\ \vdots & & & & & & \vdots \\ \frac{\partial^2 f}{\partial g_{1m}\partial g_{11}} & \cdots & \frac{\partial^2 f}{\partial g_{1m}^2} & \cdots & \frac{\partial^2 f}{\partial g_{1m}\partial g_{n1}} & \cdots & \frac{\partial^2 f}{\partial g_{1m}\partial g_{nm}} \\ \vdots & & & & & & \vdots \\ \frac{\partial^2 f}{\partial g_{n1}\partial g_{11}} & \cdots & \frac{\partial^2 f}{\partial g_{n1}\partial g_{1m}} & \cdots & \frac{\partial^2 f}{\partial g_{n1}^2} & \cdots & \frac{\partial^2 f}{\partial g_{n1}\partial g_{nm}} \\ \vdots & & & & & & \vdots \\ \frac{\partial^2 f}{\partial g_{nm}\partial g_{11}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}\partial g_{1m}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}\partial g_{n1}} & \cdots & \frac{\partial^2 f}{\partial g_{nm}^2} \end{bmatrix}$$

In the following the basic properties of trace operations are supposed to be known. A standard result for matrices, indispensable in this field, is:

$$\mathrm{d}\,(GH) = (\mathrm{d}G)\,H + G\,(\mathrm{d}H). \tag{1}$$

It implies:

$$\mathrm{d}\,(G^{-1}) = -\,G^{-1}(\mathrm{d}G)\,G^{-1} \tag{1'},$$

because $\mathrm{d}(G^{-1}G) = \mathrm{d}I = 0$. Traces are part and parcel of the optimization process. They either appear already in f (if f is a trace), or turn up in df. There is a need for handling d tr G, therefore.

A simple result is:

$$\mathrm{d}\,\mathrm{tr}\,G = \mathrm{tr}\,\mathrm{d}G \tag{2}$$

Proof:

$$\mathrm{d}\,\mathrm{tr}\,G = \mathrm{d}\sum^{i} g_{ii} = \sum^{i} \mathrm{d}g_{ii} = \mathrm{tr}\,\mathrm{d}G.$$

Another basic result is:

$$\mathrm{d}f = \mathrm{tr}\,H\,(\mathrm{d}G)' \text{ implies } H = \frac{\partial f}{\partial G} \tag{3}$$

(the reverse also being true.)

Proof:

$$\mathrm{d}f = \sum^{ij} \frac{\partial f}{\partial g_{ij}}\,\mathrm{d}g_{ij} = \mathrm{tr}\left(\frac{\partial f}{\partial G}\right)'\mathrm{d}G = \mathrm{tr}\,\frac{\partial f}{\partial G}\,(\mathrm{d}G)'.$$

A theorem useful for identification of

$$\left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right) \text{ is:}$$

$$\mathrm{d}^2 f = \mathrm{tr}\,A\,(\mathrm{d}G)\,B\,(\mathrm{d}G)' \text{ implies } A' \otimes B = \left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right), \text{ if } A' \otimes B \text{ is symmetric} \tag{4}$$

Proof:

$$\mathrm{tr}\,A\,(\mathrm{d}G)\,B\,(\mathrm{d}G)' = \mathrm{tr}\,A \begin{bmatrix} (\mathrm{d}G)_{1.} \\ \vdots \\ (\mathrm{d}G)_{n.} \end{bmatrix} B \left[(\mathrm{d}G)'_{.1} \cdots (\mathrm{d}G)'_{.n}\right] =$$

$$= \mathrm{tr}\,A \begin{bmatrix} (\mathrm{d}G)_{1.}\,B\,(\mathrm{d}G)'_{.1} & \cdots & (\mathrm{d}G)_{1.}\,B\,(\mathrm{d}G)'_{.n} \\ \vdots & & \vdots \\ (\mathrm{d}G)_{n.}\,B\,(\mathrm{d}G)'_{.1} & \cdots & (\mathrm{d}G)_{n.}\,B\,(\mathrm{d}G)'_{.n} \end{bmatrix} =$$

$$= \sum^{ij} a_{ji}(\mathrm{d}G)_{i.}\,B\,(\mathrm{d}G)'_{.j} = \left[(\mathrm{d}G)_{1.} \cdots (\mathrm{d}G)_{n.}\right] \begin{pmatrix} a_{11}B & \cdots & a_{n1}B \\ \vdots & & \vdots \\ a_{1n}B & \cdots & a_{nn}B \end{pmatrix} \begin{bmatrix} (\mathrm{d}G)'_{.1} \\ \vdots \\ (\mathrm{d}G)'_{.n} \end{bmatrix} =$$

$$= \left[(\mathrm{d}G)_{1.} \cdots (\mathrm{d}G)_{n.}\right](A' \otimes B) \begin{bmatrix} (\mathrm{d}G)'_{.1} \\ \vdots \\ (\mathrm{d}G)'_{.n} \end{bmatrix}$$

On the other hand:

$$\mathrm{d}^2 f = \sum^{ijkl} \frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\,\mathrm{d}g_{ij}\mathrm{d}g_{kl} =$$

$$= \{(\mathrm{d}G)_{1.} \cdots (\mathrm{d}G)_{n.}\} \left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right) \begin{bmatrix} \{(\mathrm{d}G)_{1.}\}' \\ \vdots \\ \{(\mathrm{d}G)_{n.}\}' \end{bmatrix} =$$

$$= \{(\mathrm{d}G)_{1.} \cdots (\mathrm{d}G)_{n.}\} \left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right) \begin{bmatrix} (\mathrm{d}G)'_{.1} \\ \vdots \\ (\mathrm{d}G)'_{.n} \end{bmatrix}.$$

This proves the theorem.

$A' \otimes B$ is the kronecker product of A' and B.

$(\mathrm{d}G)_{i.}$ is the i^{th} row of $\mathrm{d}G$; $(\mathrm{d}G)_{.i}$ is the i^{th} column of $\mathrm{d}G$;

$(\mathrm{d}G)'_{.i}$ is the i^{th} column of $(\mathrm{d}G)'$.

Summing up: (3) gives us the possibility of switching from $\mathrm{d}f$ to $\partial f/\partial G$. (2) assisted by (1) gives the possibility of translating $\mathrm{d}f$ into $\mathrm{d}K$, if f is a function of K. (4) enables us to find

$$\left(\frac{\partial^2 f}{\partial g_{ij}\partial g_{kl}}\right) \text{ from } \mathrm{d}^2 f.$$

The case of a latent root

The fundamental *theorem* here gives an expression for $\partial\lambda_i/\partial A$, λ_i being a latent root of A. The expression is derived for nonrepeated roots only. In case of repeated roots the derivatives are either indeterminate (O/O) or $\pm\infty$. Let the equations for row and column characteristic vector be:

$$Au_i = \lambda_i u_i \text{ and } w_i'A = \lambda_i w_i', \text{ with } w_i'u_i = 1.$$

Then $(\mathrm{d}A)\,u_i + A\mathrm{d}u_i = (\mathrm{d}\lambda_i)\,u_i + \lambda_i \mathrm{d}u_i$,

which leads to $w_i'(\mathrm{d}A)\,u_i + w_i'A\mathrm{d}u_i = (\mathrm{d}\lambda_i)\,w_i'u_i + \lambda_i w_i'\mathrm{d}u_i$,

giving

$$\mathrm{d}\lambda_i = w_i'\,(\mathrm{d}A)\,u_i = \operatorname{tr} w_i u_i'\,(\mathrm{d}A)' \tag{5}$$

From this follows:

$$\frac{\partial\lambda_i}{\partial A} = w_i u_i', \tag{5'}$$

because of (3).

A *corollary* is:

$$\frac{\partial\,|A|}{\partial A} = \frac{\partial \overset{i}{\prod}\lambda_i}{\partial A} = \sum^{j} \frac{\partial\lambda_j}{\partial A} \overset{i\neq j}{\prod} \lambda_i =$$

$$= \sum^{j} w_j u_j' \overset{i\neq j}{\prod} \lambda_i = \sum^{j} \frac{1}{\lambda_j} w_j u_j' \overset{i}{\prod} \lambda_i = |A|\,(A')^{-1} \tag{5''}$$

(5″) is also true for repeated roots cases. It can be derived by putting instead of A, $A + \Delta$ and taking the limit $\lim_{\Delta\to 0} |A + \Delta|\,(A + \Delta)'^{-1}$

Use has been made of the property: $W'AU = \Lambda$,

where $W' = \begin{bmatrix} w_1' \\ \vdots \\ w_n' \end{bmatrix}$, $U = (u_1 \cdots u_n)$, $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

This implies:

$$(A')^{-1} = W \Lambda^{-1} U' = (w_1 \cdots w_n) \begin{pmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_n} \end{pmatrix} \begin{pmatrix} u_1' \\ \vdots \\ u_n' \end{pmatrix} =$$

$$= \sum^{j} \frac{1}{\lambda_j} w_j u_j',$$

applying the rules for partitioned multiplication.

Formula (5') is useful for the computation of the asymptotic standard errors of latent roots of econometric equation systems. See NEUDECKER and VAN DE PANNE [5].

The determinant case

The essential theorem, we saw is:

$$\frac{\partial \mid A \mid}{\partial A} = \mid A \mid (A')^{-1} \qquad (5'')$$

This has been established along other lines of reasoning by DEEMER and OLKIN, using co-factors.

The theorem can be extended to determinants of matrix functions:

$$\mid \Phi(A) \mid$$

(5'') leads then to $\dfrac{\partial \mid \Phi(A) \mid}{\partial \Phi(A)} = \mid \Phi(A) \mid [\{\Phi(A)\}']^{-1}$

In case $\Phi(A)$ is a higher-order function of A, e.g. $A'A$, $A'AA'$ and so on, (5'') will not help us very much, if we are in need of

$$\frac{\partial \mid \Phi(A) \mid}{\partial A}, \text{ or } \mathrm{d} \mid \Phi(A) \mid$$

expressed in $\mathrm{d}A$.

Application of (3) and (5'') will help us.

(3) and (5'') imply: $\mathrm{d} \mid \Phi(A) \mid = \mid \Phi(A) \mid \operatorname{tr} \{\Phi(A)\}^{-1} \mathrm{d}\{\Phi(A)\}'$.

(1) will enable us to break down $\mathrm{d}\{\Phi(A)\}$; then

$$\mathrm{d} \mid \Phi(A) \mid = \mid \Phi(A) \mid \operatorname{tr} [\{\Phi(A)\}']^{-1} Z,$$

where Z is an expression in $\mathrm{d}A$ and $(\mathrm{d}A)'$.

Reordering this expression will lead to:

$$\mathrm{d} \mid \Phi(A) \mid = \mid \Phi(A) \mid \operatorname{tr} \{\Phi(A)\}^{-1} S (\mathrm{d}A)',$$

with

$$\frac{\partial \mid \Phi(A) \mid}{\partial A} = \mid \Phi(A) \mid \{\Phi(A)\}^{-1} S \tag{6}$$

Application [1])

To minimize $f = \log \mid W \mid$ with respect to B, where $W = E'E$ and $E = Y - XB$. W is positive definite, so $\mid W \mid$ is positive. Necessary is: $\mathrm{d}f = 0$. According to (3) and (5″) $\mathrm{d}f = \operatorname{tr} W^{-1} \mathrm{d}W$.

Obviously, $\mathrm{d}W = \mathrm{d}(Y' - B'X')(Y - XB) =$

$= - Y'X(\mathrm{d}B) - (\mathrm{d}B)' X'Y + (\mathrm{d}B)'X'XB + B'X'X(\mathrm{d}B)$.

So $\mathrm{d} \log \mid W \mid = 2 \operatorname{tr} (X'XB - X'Y) W^{-1} (\mathrm{d}B)'$.

Necessary for the optimum is: $X'XB = X'Y$.

A by-product is:

$$\frac{\partial \log \mid W \mid}{\partial B} = 2 (X'XB - X'Y) W^{-1}.$$

We shall deal with the second-order condition further on.

The trace case

The essential result here is: $\mathrm{d} \operatorname{tr} G = \operatorname{tr} \mathrm{d} G$ (2)

Application [2])

To minimize $\operatorname{tr} A'AV$ with respect of A, under the condition $AX = I$. Of course A is nonsquare, ~~and square and diagonal~~ V is assumed to be positive definite, and symmetric.

The formulation of the problem is:

Minimize $f = \operatorname{tr} A'AV - \operatorname{tr} M'(AX - I)$, where M is a matrix of Lagrange multipliers.

It is easy to see that:

$\mathrm{d}f = \mathrm{d} \operatorname{tr} A'AV - \mathrm{d} \operatorname{tr} M'(AX - I) =$

$= \operatorname{tr} (AV + AV' - MX')(\mathrm{d}A)'$ because of (2)

Necessary for the optimum is: $AV + AV' = MX'$.

Next $\mathrm{d}^2 f$

From $\mathrm{d}f = \operatorname{tr} (AV + AV' - MX')(\mathrm{d}A)'$ it follows:

$\mathrm{d}^2 f = \mathrm{d} \operatorname{tr} (AV + AV' - MX')(\mathrm{d}A)' = 2 \operatorname{tr} (\mathrm{d}A) V (\mathrm{d}A)'$.

[1]) From: GOLDBERGER [2], p. 205 seq.

[2]) From: JOHNSTON [3], p. 180 seq.

This is a sum of quadratic forms, with matrix V, which is assumed to be positive definite.

Thus $d^2f > 0$, for $dA \neq 0$.

$$\left(\frac{\partial^2 f}{\partial a_{ij}\partial a_{kl}}\right) = 2I \otimes V \qquad \text{because of} \quad (4)$$

Application [3])

To establish the second-order conditions for application 1.

Clearly $d^2f = 2\,d\,\mathrm{tr}\,(X'XB - X'Y)\,W^{-1}\,(dB)' =$

$= 2\,\mathrm{tr}\,X'X\,(dB)\,W^{-1}\,(dB)' + 2\,\mathrm{tr}\,(X'XB - X'Y)\,(dW^{-1})\,(dB)' =$

$= 2\,\mathrm{tr}\,X'X\,(dB)\,W^{-1}\,(dB)'$ because $X'XB = X'Y$ (necessary condition).

We can transform further, and get then:

$$d^2f = 2\,\mathrm{tr}\,W^{-\frac{1}{2}}\,(dB)'X'X\,(dB)\,W^{-\frac{1}{2}} = 2\,\mathrm{tr}\,Z'Z.$$

We define $Z = X\,(dB)\,W^{-\frac{1}{2}}$, where $W^{-\frac{1}{2}}$ stands for any symmetric matrix K satisfying $K^2 = W^{-1}$.

Under convenient rank assumptions $\mathrm{tr}\,Z'Z > 0$ for $dB \neq 0$.

This establishes the minimum. A by-product is:

$$\left(\frac{\partial^2 f}{\partial b_{ij}\partial b_{kl}}\right) = 2X'X \otimes W^{-1} \qquad \text{in the minimum.}$$

References:

[1] DEEMER, W. L. and I. OLKIN, "The Jacobians of certain matrix transformations useful in multivariate analysis," Biometrika Vol. **38** (1951).

[2] GOLDBERGER, A. S., "Econometric Theory", New York: John Wiley & Sons, 1964.

[3] JOHNSTON, J., "Econometric Methods", New York: McGraw-Hill, 1963.

[4] MALINVAUD, E., "Méthodes statistiques de l'économétrie", Paris: Dunod, 1964.

[5] NEUDECKER, H. and C. VAN DE PANNE, "Note on the asymptotic standard errors of latent roots of econometric equation systems," Discussion Paper Series A, No. 54, Faculty of Commerce and Social Science, University of Birmingham.

[6] OLKIN, I., Note on "The Jacobians of certain matrix transformations useful in multivariate analysis". Biometrika Vol. **40** (1953).

[3]) From: GOLDBERGER [2], p. 205 seq.

NOTE ON THE ASYMPTOTIC STANDARD ERRORS OF LATENT ROOTS OF ECONOMETRIC EQUATION SYSTEMS

by

H. Neudecker and C. van de Panne
University of Birmingham

REVIEW OF THE INTERNATIONAL
STATISTICAL INSTITUTE
Volume 34 : 1, 1966

NOTE ON THE ASYMPTOTIC STANDARD ERRORS OF LATENT ROOTS OF ECONOMETRIC EQUATION SYSTEMS

by

H. Neudecker and C. van de Panne [1]
University of Birmingham

1. INTRODUCTION

One of the interesting properties of econometric equation systems is the stability of these systems. This stability depends on the latent roots of the system. Since the estimates of the coefficients of an econometric equation system have standard errors, the roots which depend on these estimates will also have standard errors. In certain cases it is important to know the standard error of the absolute value of the latent roots.

Systems having roots with absolute values smaller than one are stable, whereas systems having at least one root with an absolute value greater than or equal to one are unstable. However, it might be that in some cases in which the estimates of the latent roots indicate that the system is stable, the standard error of the estimates is so large that the absolute value of some of the roots does not differ significantly from one; in such cases the system does not possess a significant stability. On the other hand, in some system which has a latent root with an absolute value larger than one, the instability may not be significant, so that also in this case it cannot be decided whether the system is stable or not. Hence an estimate of the standard error of the latent roots of a system is useful in determining whether a system is significantly stable or unstable.[2] In general, only an estimate of the asymptotic standard deviation will be available.

Theil and Boot [4] have given a procedure for the derivation of the asymptotic standard error of the absolute value of a complex latent root, but it has the disadvantage of being impracticable for larger systems since it uses cofactors in its computations which are computationally inefficient except in very small systems.

This note offers an alternative procedure which does not have this drawback. At the same time this procedure will be slightly more general in considering also the asymptotic standard error of a real root and that of the argument of a complex root.

2. THE CASE OF A REAL LATENT ROOT

Any linear econometric equation system can be written as a system of difference equations of a certain order in the endogenous variables. By a redefinition of variables this system can be reformulated into a system of difference equations of the first order as follows[3]

$$y(t) = Ay(t-1) + Bx(t), \tag{2.1}$$

where $y(t)$ is a vector of, say, n endogenous variables and $x(t)$ a vector of m exogenous variables, all referring to period t. A and B are then $n \times n$ and $n \times m$ matrices, respectively.

[1] Now at the State University of Groningen, Holland.
[2] For a measure of the degree of stability of a system, see Smyth [3] and van de Panne [2].
[3] See e.g. Theil and Boot [4].

The n latent roots of A are then the roots of the econometric equation system. They can be found in the usual way by solving the characteristic equation

$$| A - \lambda I | = 0, \tag{2.2}$$

or by any other method for finding the latent root of a matrix. Since A is not symmetric, the roots are not necessarily real. The characteristic columns and rows are determined by the following equations

$$(A - \lambda_j I)\, v_j = 0, \quad j = 1, \ldots, n, \tag{2.3}$$

$$w_j' (A - \lambda_j I) = 0, \quad j = 1, \ldots, n, \tag{2.4}$$

where v_j and w_j are column vectors of n elements; if there are no multiple latent roots, they are uniquely determined. Furthermore, we have, after normalization

$$\begin{aligned} w_k' v_j &= 1 \text{ for } k = j \quad k, j = 1, \ldots, n. \\ &= 0 \text{ for } k \neq j\ . \end{aligned} \tag{2.5}$$

We want to find the asymptotic standard error of the latent roots of A from the asymptotic covariance matrix of the elements of A, of which, we assume, a suitable estimate is available. The latent roots are differentiable scalar functions of the elements of A, see (2.2). Here a result can be used which can be found, e.g. in Goldberger [1, p. 122-125, especially (6.32)]; this result relates the asymptotic variance of a scalar function $y = f(x)$, where x is a vector of random variables, to the asymptotic covariance matrix of x.

According to this result we have the following. Let us write the columns of the matrix A as a vector a' in the following manner:

$$a' = [a_{11} \ldots a_{n1} \ldots a_{1n} \ldots a_{nn}]. \tag{2.6}$$

The matrix A will in general be an estimate based on a sample size T. We assume that this estimate is asymptotically unbiased, so that the asymptotic expectation of the vector a is α:

$$\operatorname*{Lim}_{T \to \infty} E\, a = \alpha \tag{2.7}$$

The asymptotic covariance matrix of the vector a is assumed to be

$$\frac{1}{T} \operatorname*{Lim}_{T \to \infty} \left\{ E\, T (a - Ea)\, (a - Ea)' \right\} = V = \frac{1}{T}\, W, \tag{2.8}$$

where W is a matrix of fixed constants.

Let us write for the root λ_j

$$\lambda_j = f(a)\ . \tag{2.9}$$

Then, according to the results we referred to, we have for the asymptotic variance of λ_j:

$$\frac{1}{T} \operatorname*{Lim}_{T \to \infty} E\, \{T (\lambda_j - E\lambda_j)^2\} = \left(\frac{\partial \lambda_j}{\partial a}\right)' V \left(\frac{\partial \lambda_j}{\partial a}\right), \tag{2.10}$$

where $\left(\frac{\partial \lambda_j}{\partial a}\right)$ is the column vector of partial derivatives of λ_j with respect to the elements of a.

This vector of partial derivatives is found as follows. Let us differentiate (2.3):

$$\{dA - (d\lambda_j)\, I\}\, v_j + (A - \lambda_j I)\, dv_j = 0. \tag{2.11}$$

After premultiplication by w_j' we find

$$w_j' \{dA - (d\lambda_j)\, I\}\, v_j + w_j' (A - \lambda_j I)\, dv_j = 0. \tag{2.12}$$

The last term is zero according to (2.4). On multiplying out the remainder we find

$$w_j' (dA)\, v_j - (d\lambda_j)\, w_j' v_j = 0. \tag{2.13}$$

Since according to (2.5) we have $w_j' v_j = 1$, we can write (2.13) as

$$d\lambda_j = w_j' (dA)\, v_j. \tag{2.14}$$

Let us also write all columns of dA in the form of a row vector $(da)'$:

$$(da)' = [da_{11} \ldots da_{n1} \ldots da_{1n} \ldots da_{nn}]. \tag{2.15}$$

(2.14) can then be written as

$$d\lambda_j = [v_{j1} w_j' \;\; v_{j2} w_j' \ldots v_{jn} w_j']\, da = s_j' da, \tag{2.16}$$

where s_j' stands for the row vector between the square brackets. From this we find

$$\left(\frac{\partial \lambda_j}{\partial a}\right) = s_j. \tag{2.17}$$

Hence we have for the asymptotic variance of λ_j

$$\frac{1}{T} \operatorname*{Lim}_{T \to \infty} E\{T(\lambda_j - E\lambda_j)^2 = s_j'\, V\, s_j, \tag{2.18}$$

in case λ_j is real.

3. THE CASE OF COMPLEX ROOTS

If λ_j is complex, its standard error will in general be complex as well. In this case, however, we are interested in the standard error of the absolute value and the argument of the root. The absolute values of the roots determine the stability of the system, while the arguments are connected with the periods of the fluctuations of the system.

Let the root be

$$\lambda_j = p_j + iq_j \tag{3.1}$$

where p_j and q_j are real. Differentiating the absolute value of λ_j we find

$$d\,|\,\lambda_j\,| = d\,(p_j^2 + q_j^2)^{1/2} = (p_j dp_j + q_j dq_j)\,(p_j^2 + q_j^2)^{-1/2} \tag{3.2}$$

dp_j and dq_j can be found from (2.9) as follows

$$d\lambda_j = dp_j + i dq_j = w_j' (dA)\, v_j = s_j' da. \tag{3.3}$$

If λ_j is complex, then the corresponding characteristic vectors w_j and v_j will also be complex and hence also the vector s_j, which can be separated into a real part $\bar{s}_j$ and a complex part $i\bar{\bar{s}}_j$ with $\bar{\bar{s}}_j'$ a vector of real elements. We find therefore

$$dp_j = \bar{s}_j'\, da, \quad dq_j = \bar{\bar{s}}_j'\, da. \tag{3.4}$$

Substituting this into (3.2) we find

$$d \mid \lambda_j \mid = (p_j^2 + q_j^2)^{-1/2} (p_j \bar{s}_j' + q_j \bar{\bar{s}}_j') \, da. \tag{3.5}$$

From this we find

$$\left(\frac{\partial \mid \lambda_j \mid}{\partial a}\right) = (p_j^2 + q_j^2)^{-1/2} (p_j \bar{s}_j' + q_j \bar{\bar{s}}_j'). \tag{3.6}$$

The asymptotic standard error of $\mid \lambda_j \mid$ is therefore equal to

$$\{(p_j^2 + q_j^2)^{-1} (p_j \bar{s}_j' + q_j \bar{\bar{s}}_j') \, V (p_j \bar{s}_j + q_j \bar{\bar{s}}_j)\}^{1/2}, \tag{3.7}$$

which, if an estimate of V is available, can be computed. The argument of a complex number $p_j + iq_j$ is defined by

$$\tan \theta_j = \frac{q_j}{p_j}. \tag{3.8}$$

Differentiating, we find

$$d(\tan \theta_j) = \sec^2 \theta_j d\, \theta_j = -q_j p_j^{-2} \, dp_j + p_j^{-1} \, dq_j, \tag{3.9}$$

from which we obtain

$$\begin{aligned} d\, \theta_j &= (\cos^2 \theta_j) (-q_j p_j^{-2} \, dp_j + p_j^{-1} \, dq_j) = \\ &(\cos^2 \theta_j) (-q_j p_j^{-2} \bar{s}_j' + p_j^{-1} \bar{\bar{s}}_j') \, da; \end{aligned} \tag{3.10}$$

the last equality is found using (3.4).

From this we find

$$\left(\frac{\partial \theta_j}{\partial a}\right) = (\cos^2 \theta_j) (-q_j p_j^{-2} \bar{s}_j' + p_j^{-1} \bar{\bar{s}}_j'). \tag{3.11}$$

From this follows the asymptotic standard error of θ_j:

$$\{[\cos^4 \theta_j] [- q_j p_j^{-2} \bar{s}_j + p_j^{-1} \bar{\bar{s}}_j'] \, V \, [- q_j p_j^{-2} \bar{s}_j + p_j^{-1} \bar{\bar{s}}_j]\}^{1/2}$$

In some cases it is desirable to know the standard error of the period of fluctuation rather than the argument. This period can be expressed in terms of the argument as follows

$$T_j = 2\pi \theta_j^{-1}, \tag{3.12}$$

from which we deduce

$$\frac{dT_j}{d\theta_j} = -2\pi \theta_j^{-2}, \tag{3.13}$$

so that

$$\left(\frac{\partial T_j}{\partial a}\right) = -2\pi \theta_j^{-2} \left(\frac{\partial \theta_j}{\partial a}\right). \tag{3.14}$$

Hence the corresponding asymptotic standard error is equal to

$$2\pi \theta_j^{-2} (\cos^2 \theta_j) \{(- q_j p_j^{-2} \bar{s}_j' + p_j^{-1} \bar{\bar{s}}_j') \, V (- q_j p_j^{-2} \bar{s}_j + p_j^{-1} \bar{\bar{s}}_j)\}^{1/2} \tag{3.15}$$

The formulas derived above can also be used to asses the impact of changes of

structural coefficients on the stability of the system. This may be of importance for tax parameters. If, for example, we have a tax equation of the following kind

$$T = t_1 + t_2 Y + t_3 Y_{-1}, \tag{3.16}$$

we may consider separate or simultaneous changes in the coefficients, dt_1, dt_2, dt_3. An expression for the impact of these changes on the elements of A can be given, after which the impact on the roots or their absolute values and arguments can be found.

REFERENCES

[1] Goldberger, A. S. (1964) *Econometric Theory*. New York, John Wiley.

[2] van de Panne, C. Variances and Stability in Linear Dynamic Models. Discussion Paper Series A, No. 56, Faculty of Commerce and Social Science, University of Birmingham.

[3] Smyth, D. J. Built-In Flexibility and Automatic Stabilisation. Discussion Paper Series A, No. 37, Faculty of Commerce and Social Science, University of Birmingham.

[4] Theil, H. and Boot, J. C. G. (1962) The Final Form of Econometric Equation Systems. *Review of the International Statistical Institute*, 30, p. 136–152.

RESUME

L'article expose une méthode générale qui permet d'obtenir l'expression asymptotique de l'écart-type des valeurs propres des systèmes d'équations économétriques.
Lorsque les valeurs propres sont des nombres complexes on considère les valeurs asymptotiques de l'écart-type du module et de l'argument.

METHODOLOGY OF SPATIAL EQUILIBRIUM MODELS OF THE RICE ECONOMY OF SOUTH INDIA

by

M. G. KANBUR & H. NEUDECKER

Reprinted from :

'Artha Vijnana', Vol. 8 No. 1, March 1966

GOKHALE INSTITUTE OF POLITICS AND ECONOMICS, POONA

METHODOLOGY OF SPATIAL EQUILIBRIUM MODELS OF THE RICE ECONOMY OF SOUTH INDIA*

M. G. KANBUR** and H. NEUDECKER

THE DECADE OF THE FIFTIES witnessed the development of spatial equilibrium models from theoretical exposition to empirical application. Linear programming methods were then introduced into spatial equilibrium model building.

Spatial equilibrium models seek to determine an equilibrium consumption pattern and an optimum trade pattern for a product between three or more market places within perfectly competitive market area (or, what is the same, within a perfectly regulated market area). In this paper we have made an attempt to illustrate the methodology of spatial equilibrium models by an empirical application to the rice economy of South India.

Spatial equilibrium analysis is the outcome of two well-known schools of thought in economics. The first one believed in classical, neoclassical and modern trade theory whereas the second one stemmed from location theory. The classical writers were mainly concerned with the welfare aspects of trade. The comparative cost theory of the classical economists assumed that trade takes place when there are international differences existing between real unit cost ratios. According to them, some trade is better than no trade. Traditional trade theory was criticized as it ignored completely the existence of space aspects of economic activity. The economists belonging to the school of location theory observed this gap in the traditional trade theory and Von Thunen constructed a generalized location model with the help of which it was possible to explain the effect of transportation on the kind of production to be found at given distances from the market centre. With the absence of spatial variables in the traditional trade theory, it is difficult to explain intra- and inter-regional deve-

* A paper presented at the Fifth Indian Econometrics Conference, held in Delhi on 20-22 December, 1965.

** The author is thankful to Prof. A. A. Walters for having suggested the topic. He is grateful to Faculty of Commerce and Social Sciences for the financial assistance.

lopment of resources and other problems of social-economic adjustment in a spatially differentiated world. The aim of the location theorists was to explain what the classical economists omitted, namely, social and economic activity as influenced by space. It may be noted here that several contributions to date have revealed the generalization of spatial analysis and none of them provided a theory in the Walrasian tradition.

A stage has now been set when we are in a position to define a spatial equilibrium model. Such a model may be broadly defined as any theoretical formulation with 'space' as one component. Thus the spatial equilibrium models formulate the equilibrium situation of the economy in its totality and study the interdependence and interrelatedness of economic elements. They also characterize the several economic activities namly, the regional location and production, the regional level of consumption, and the relative level of regional prices.

Studies to date have revealed that the first attempt to combine general equilibrium analysis with location theory with a view to creating a general system is that of August Losch. He emphasized the general inter-relation of spatially separated economic units and analysed the choice of location in terms of spatial interdependence. It was Enke, Samuelson, Beckmann and Baumol who set out the problem of inter-connected competitive markets in a new form and also opened a new approach to spatial pricing systems and competitive locational equilibrium.

The general problem of equilibrium among spatially separated markets has been stated in one of its simplest forms by Enke. His formulation proceeds as follows :

> "There are three (or more) regions trading a homogeneous good. Each region constitutes a single and distinct market. The regions of each possible pair of regions are separated — but not isolated — by a transportation cost per physical unit which is independent of volume. There are no legal restrictions to limit the actions of the profit-seeking traders in each region. For each region, the functions which relate local production and local use to local price are known and consequently, the magnitude of the difference which will be exported or imported at each local price is known. Given these trade functions and transportation costs, we wish to ascertain (i) the

net price in each region, (2) the quantity of exports or imports for each region (3) which regions import, export or do neither (4) the aggregate trade in each commodity (5) the volume and direction of trade between each possible pair of regions. . . ."

Samuelson worked on the Enke problem and showed how it contains the minimum transportation cost problem of Koopman-Hitchcock :

"A specified number of ships is to be sent out from each of a number of ports. They are to be allocated among a number of other receiving ports with the total sent in to each such port being specified. If we are given the unit costs of shipments between every two ports, how can we minimise the total cost of the programme."

From our empirical standpoint, the following studies may be mentioned. Fox and Judge adapted Enke-Samuelson-Beckmann formulations and developed spatial price equilibrium models of the feed-livestock and egg sectors of the economy respectively. Fox and Taeuber analysed the feed-livestock economy and determined equilibrium prices and shipments for the factor (feed concentrates) and for the product (livestock). The model was solved using two linear programmes, one for feed and the other for livestock. Snodgrass and French and Henry and Bishop used the transportation model as formulated by Koopmans to derive optimum geographical flows for the dairy and broiler sector of the economy. These studies suggest the practicability of research relating to spatial pricing systems and competitive locational equilibrium. Karl Fox in 1954 also showed how this approach might be used in studying the effect on the pattern and volume of trade among several countries or groups of countries of changes in tariffs and export-subsidies.

Although most of the interesting empirical applications are found in the field of agriculture, there are some efforts to develop spatial models in the field of industry as well. To quote, Thomas A. Marschak formulates a spatial model of U.S. Petroleum Refining.

Herbert Gulicher has applied the transportation model to the analysis of oligopolistic elements in the West German petrol transportation system.

The spatial equilibrium model will be more accurate in the short run. The individual producer and the policy maker can make consider-

able use of consumption estimates for predictions. Efforts have to be made to examine how this model will behave in predicting long-run regional production. These are some of the components which can be predicted if not accurately at east nearer to accuracy. (1) regional price formation, (2) regional consumption, and (3) regional storage demands. Of the three, price forecasts will be the most useful to consumers, producers and most service industries. The models are useful tools for economic analysis and play an important role in the analysis of industry-wide policies, regional price determination and inter-regional production adjustment and resource allocation.

A spatial equilibrium analysis of this nature requires specification of three types of functional relationships; namely demand functions, transport cost model and the supply functions. The demand function is based on the relation between per capita consumption of the commodity, per capita income of the region, the retail price of the commodity and the retail price of the substitutes. The transportation cost model has to be developed from mileage-rate data of transporting a unit commodity from one region to another region. This has to be built up for different types of transportation, namely, rail, road and waterways; for each route we choose that type which involves least cost of transportation. The supply function has to be handled in two ways. First, product supply may be assumed to be predetermined. The second approach may be to develop production cost and minimize the total costs (transportation cost *plus* production cost) in meeting regional demand.

The first point to be noted is the partition of the area (space) under study into reasonably meaningful and realistic geographical units or regions. Although the area should be divided into homogeneous units, the final demarcation will sometimes involve subjective considerations because of the restrictions on the extent of the availability of data.

The following assumptions are made to reduce the model to a simplified version of reality : (1) perfect competition and full employment; (2) regional demands are represented by linear demand functions; (3) divisibility in quantity of product flow; (4) imports or exports in a region are independent of imports or exports in other regions; (5) of all the possible regions in which a space might be divided, only n regions are considered among which k regions will export or $n - k$ regions will import; (6) within regions, spatial homo-

geneity is assumed; (7) freight charges are the same both ways and include all such charges incurred in transporting a unit commodity from a seller in one region to a buyer in another region; (8) the shipment of commodity from one region to another does not affect the per capita disposable income of the consumers in either of the regions; and (9) production, distribution, and consumption of the commodity take place at the point. These assumptions are basically given by (or used by) Judge & Wallace. They, however, did not formulate all of these explicitly (namely 5 and 8).

Mathematical Model

Let X_i denote the quantity of the commodity consumed in the i^{th} region. Then, the demand function for the i^{th} region may be expressed as follows :

$$X_i = f(P_i, \overset{*}{I_i}, \overset{*}{N_i}, P_R)$$

Where

P_i = Retail price of the Commodity in region i
I_i = per capita income in region i
N_i = population in region i
P_R = Retail price of the substitutes.

The supply of the commodity in the i^{th} region is denoted by $S_i^* = K_i$

The equilibrium condition to be satisfied is that $\sum_i X_i = \sum_i \overset{*}{S_i}$.

Given observations relating to per capita income, population, production, and the demand function of each region, it is possible to determine a structure of regional prices using approximate price differentials. If P_1 is the price prevailing in region 1 and P_2 is the price prevailing in region 2 then in case of shipment (under Perfect Market conditions) $P_1 - P_2 = t_{12}$ if t_{12} is the transport cost of shipping a unit commodity from region 1 to region 2. This is true according to perfect market concept. In order to obtain a structure of geographical prices that are tied together by transport costs, the procedure followed is to choose a base region and then approximate a set of price differentials to that base. To start with a set of price differentials on first approximation will be the transport costs of shipping a commodity to the respective regions from the base region

* Variables are assumed to be pre-determined.

chosen for the purpose. Provided with a set of approximate regional price differentials, the next task is to work out the equilibrium price in the base region. As the variables of the demand relation are assumed to be linear, an easy procedure can be followed for arriving at the equilibrium set of geographical prices, using the equilibrium conditions that supply (predetermined) equals demand. Equilibrium price in any other region is obtained by equating it to the equilibrium price in the base region plus the price differential. Incidentally the price differential may be positive, negative or zero. After the determination of equilibrium prices in the regions, the demand relations are utilized to arrive at the total consumption in each region. The regional estimates of consumption are compared with the regional supply (predetermined) to find out whether the region is surplus or deficit and also the margin of surplus and deficit. The next task is to check on the authenticity of the regional price differentials used on first approximation. To do this is nothing but determining minimum cost flows by linear programming methods. The following tableau is presented to make the problem more clear. Let t_{ij} denote the cost of transporting a commodity from the i^{th} region to the j^{th} region. Let X_{ij} represent the amount of product transported from the i^{th} surplus region to the j^{th} deficit region.

				Deficit Regions	*Totals*
	i^j	(1)	(2)	.....(m)	s_i
	(1)	t_{11}	t_{12}	..... t_{1m}	s_1
	(2)	t_{21}	t_{22}	..... t_{2m}	s_2
Surplus	(3)	t_{31}	t_{32}	..... t_{3m}	s_3
Regions	⋮				⋮
	(n)	t_{n1}	t_{n2}	..... t_{nm}	s_n
Totals	d_j	d_1	d_2	..... d_m	$\sum_{i=1}^{n} s_i = \sum_{j=1}^{m} d_j$.

From the analysis described in previous paragraphs, s_i and d_j are known. The problem is to find a set of X_{ij} such that

$$\sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}\ t_{ij} \text{ is minimum.}$$

$$\text{Subject to } \sum_{j=1}^{m} X_{ij} = s_i \qquad i = 1, 2 \ldots\ldots n$$

$$\sum_{i=1}^{n} X_{ij} = d_j \qquad j = 1, 2 \ldots\ldots m$$

$$\sum_{i=1}^{n} s_i = \sum_{j=1}^{m} d_j.$$

and $X_{ij} \geq 0$ for all i and j

With the aid of the duality theorem of linear programming, a unique set of price differentials may be derived corresponding to the equilibrium set of flows; subsequently initial price differentials are checked in case they are different, the new set of price differentials is used, and the entire procedure of determining regional surpluses and deficits is repeated, and the transportation model is again solved. If the price differentials arrived at any stage are the same as that of the previous stage, then it can be concluded that the optimum shipment pattern has reached.

The Empirical approach : The Rice Economy of South India (1960-61 conditions).

This section of the paper illustrates the application of the spatial equilibrium models to the rice economy of the South Zone which consists of the states of Andhra Pradesh, Madras, Mysore and Kerala. Rice is an important crop of South India and as such justifies the development of rather sophisticated models which might assist in taking policy decisions. Another factor which favoured us to undertake the analysis was the homogeneity of the South Zone. There is not much disparity among the different states of the South Zone so far as the living standards are concerned.

For purposes of our analysis, the South Zone has been divided into 11 regions. Each region consists of a group of districts. Each region has to be represented by a point, we noted. Here, regional centres which are important markets are chosen as such points.

The totality of regional centres has been chosen in such a way that the Southern Zone is adequately "spanned up."

Secondly, we have not disaggregated below the district level. Data

on production, population and other elements are mostly available at district levels.

Market-Demand Relationship

To put the model to operational use requires the necessity of regional market-demand relationships for rice. But certain difficulties in the form of non-availability of adequate data prevented us from building up demand functions at regional level. To overcome this difficulty an effort has been made to build up the demand function for the South Zone as a whole. The arithmetic linear demand function was postulated as

$$x_1 = a_1 + b_1 x_2 + c_1 x_3.$$

where x_1 is the per capita consumption of rice.

x_2 is the per capita real income.

x_3 is the retail price of rice.

Data applying to the 1960-61 observed values of the variables were then employed along with the income- and price-elasticity estimates to derive the desired linear coefficients. Thus the linear market-demand relationship for the South Zone is given by

$$x_1 = 212{\cdot}09 + 0{\cdot}23x_2 - 0{\cdot}71x_3$$

Transport model

We have assumed that the largest parts of shipment of rice is done through railways. Rail rates for transportation of rice and wheat have been used to construct a model which is of the form

$$C_{ij} = A_1 + A_2 M_{ij} + A_3 M_{ij}^{\frac{1}{2}}$$

where C_{ij} denotes the cost of transporting a unit quantity of rice from point i to point j by rail. M_{ij} is the rail mileage between i and j, A_1, A_2 and A_3 are parameters to be estimated. The method of least squares was used to estimate the unknown parameters. The results are

$$C_{ij} = -26{\cdot}23 + 0{\cdot}062\, M_{ij} + 7{\cdot}52\, M_{ij}^{\frac{1}{2}}$$
$$R^2 = {\cdot}908$$

The point that is worth mentioning is that we have used smooth functions to generate transport costs instead of using actual rates existing between two points.

Equilibrium solutions

An equilibrium solution for the economy as a whole involves the use of a definite structure of regional prices that are connected by means of transport costs. In the present analysis, we have chosen Madras as the base region and so the approximate regional price differentials will therefore be the respective unit transportation costs from Madras. The final spatial equilibrium solutions, regional equilibrium prices, regional consumption, surplus and deficits and optimum flows for rice — all referring to the conditions prevailing in the year 1960-61 are summarized in Table 1.

The results in Table 1 are only to be taken as illustrative but a few comments on them may not be out of place. The equilibrium-prices as shown in it are rather high and do not show much variation. The per capita equilibrium consumption figures also vary between 190 and 200 pounds only. Since an identical demand function is assumed to hold for all the regions and since the variations in the prices are also small, the variations in the equilibrium-consumption would mostly be due to variations in per capita incomes, which may also not be large. Kakinada the largest surplus region has to send its surplus to the largest deficit region namely, Hubli. The rest of the deficit in the Hubli region is made up from shipments from the neighbouring surplus regions of Vijayawada and Shimoga. The surplus regions on the whole send their surplus to neighbouring deficit regions as one would normally expect. If some of the assumptions made here, namely those of fixed supply, identical demand functions for the different regions, absence of imports and exports from outside the overall region and of unrestrained capacity for transportation (of rice) by rail, are relaxed, one may expect to get better results. Work on these lines is now in progress.

Conclusions

Knowledge gained from the regional equilibrium prices and shipment pattern has immediate practical importance, in framing various policy-decisions in the public and private sectors of the country's economy. It is also possible to examine how outcomes change under different initial conditions. It will be instructive to ascertain how changes in the regional supplies of the commodity will change the equilibrium solution. Given the spatial equilibrium model, it is pos-

Table 1 : REGIONAL EQUILIBRIUM PRICES, CONSUMPTION, SURPLUSES AND OPTIMUM FLOWS (1960-61) (FINAL)

Regions	*Equilibrium price*	*Equilibrium consumption (per capita)*	*Surplus*	*Destinations and quantities of shipments (1000 lb.)*					*Price differential from Madras*
				Nellore	*Nizamabad*	*Ernakulam*	*Hubli*	*Mysore*	
	Paisse/lb.	*lb.*	1000 *lb.*	2	4	5	9	11	
1 Kakinada	125·17	191·31	1040413				1040413		—0·51
2 Nellore	125·80	190·86	—398358						0·12
3 Vijayawada	125·36	191·18	960461	398358	530998		31105		—0·32
4 Nizamabad	126·09	190·66	—530998						0·41
5 Ernakulam	126·57	194·45	—826604						0·89
6 Madras	125·68	199·68	281347					281347	0·00
7 Kumbhakonam	125·34	199·92	1012102			656596		255506	—0·34
8 Tirunelvelli	126·03	199·57	170008			170008			+0·35
9 Hubli	126·28	194·19	—1290508						0·60
10 Shimoga	125·75	194·57	331261				218990	112271	+0·07
11 Mysore	126·40	194·11	—749124						0·72

Total shipments (1000 lb.) =3795592; Total costs = Rupees 3,414,246

sible to find answers to the following problems : (i) Effect of the increase in the supply of the commodity on the regional price and the prices elsewhere; (ii) effect on aggregate exports and imports; and (iii) changes in the optimum pattern of shipment as a result of changes in unit transport costs of shipment. Lastly it is important to know that this competitive model illustrates what could happen under given assumptions and does not necessarily say what will actually happen. Further work on these lines is in progress.

APPENDIX

Problems of Stability of the optimum solution

In Table 1 certain regions are exporters, others are importers and there is a certain pattern of shipments between these regions. In short, there is a definite functional pattern. A relevant problem is : How stable is the functional pattern displayed by Table 1 ?

A preliminary question is : Stability with respect to which factors? There are different levels of active factors. Firstly, there are the supplies; secondly, there are the demand functions; thirdly, there are the transport costs. The model apart from its parametric specification is a maintained hypothesis.

From a short-run-policy point of view, the transport costs are the relevant factors. We shall try to answer the question how stable the functional pattern displayed by Table 1 is. (It will turn out that we can only give some sufficiency conditions easily; this will in most cases be enough).

Suppose we have mis-estimated the transport costs by a factor ε ($\varepsilon > 0$). So instead of the cost matrix C we would have used εC. Which values of ε will preserve the functional pattern of Table 1 ?

It is obvious that the optium solution of Table 1 is a feasible solution in the case of a cost matrix εC. There will be an adjustment now in the gross demands, because the price differentials (to be denoted by d_i) will be multiplied by ε as well. Consequently surpluses and deficits will change. So a new iterative process will start. How will gross demands be in the first iteration round ?

There is an extremely easy formula for this, which will be proved :

$$V_i^x = \varepsilon V_i + \frac{1}{n}(1-\varepsilon)S \qquad 1.$$

Here V_i is the gross demand of the i^{th} region in the optimum solution (of Table 1); V_i^x is the gross demand we are after; S is total supply (stable), n is the total number of regions.

Proof :

Let the demand equation be denoted by

$$V_i = a + b(\bar{p} + d_i) \qquad 2.$$

Here $\bar{p}$ is the equilibrium price, and d_i is the price differential of the i^{th} region. In the first iteration round there will be price differentials εd_i and an equilibrium price $\bar{p}^x$. This $\bar{p}^x$ will be expressed first.

Equation 2 gives us ;

$$V_i^x = a + b(\bar{p}^x + \varepsilon d_i),$$

$$S = \sum_i V_i^x = na + nb\bar{p}^x + b\varepsilon \sum_i d_i,$$

$$\text{so} \quad \bar{p}^x = \frac{S - na - b\varepsilon \sum_i d_i}{nb}$$

Adding up equation 2 gives us:

$$S = \sum_i V_i = na + nb\bar{p} + b \sum_i d_i,$$

$$\text{so } b \sum_i d_i = S - na - nb\bar{p}$$

Substituting this expression for $b \sum_i d_i$ in 4 leads to

$$\bar{p}^x = \frac{S - na - \varepsilon S + \varepsilon na + \varepsilon nb\bar{p}}{nb}$$

Substituting this expression for $\bar{p}^x$ in 3 gives then :

$$V_i^x = a + \frac{S - na - \varepsilon S + \varepsilon na + \varepsilon nb\bar{p}}{n} + b\varepsilon d_i$$

$$= \varepsilon a + \frac{1}{n}(1 - \varepsilon) S + \varepsilon b\bar{p} + \varepsilon b d_i =$$

$$= \varepsilon V_i + \frac{1}{n}(1 - \varepsilon) S$$

If all transportation costs are multiplied by ε, the gross demands will change into V_i^x, this being equal to a linear combination of V_i and average supply (not a convex one, because ε can be greater than 1!).

Formula (1) will help us in finding (some) stability conditions.

With the help of the $\overset{x}{V}_i$ and the S_i (supplies), the revised margin of surplus for a surplus region (in Table 1) will be

$$S_i - \overset{x}{V}_i \qquad (i = 1,2.....6)$$

and the revised margin of deficit for a deficit region (in Table 1) will be

$$\overset{x}{V}_j - S_j$$

Given these $S_i - \overset{x}{V}_i$ ($i = 1,...,6$ in our case) and $\overset{x}{V}_j - S_j$ ($j = 7, .-11$) we can construct a parametric transportation scheme, having the pattern of Table 1. It is displayed by Table 2.

Table 2 : A PARAMETRIC TRANSPORTATION SCHEME

Origin \ DESTINATIONS	7	8	9	10	11	Maximal ~~Total~~ supply (*lakh lb.*)
1					8785+1619ε	8785+1619ε
2	6178—2195ε	8418—3109ε			6883ε—6572	8024+1579ε
3				18816ε—17693	11313—9123ε	9693ε—6380
4				5094ε—2281		5094ε—2281
5			10874ε—4309	29468—25912ε		25159—15038ε
6			2167ε—466			2167ε—466
~~Total~~ Maximal demand (*lakh lb.*)	6178—2195ε	8418—3109ε	13041ε—4775	9494—2002ε	13526—621ε	5114ε+32841

From these parametric transport costs follow requirements for ε. ε should need 10 requirements to make the above scheme feasible. The 10 requirements are :

$$6178 - 2195\varepsilon > 0$$
$$8418 - 3109\varepsilon > 0$$
$$8785 + 1619\varepsilon > 0$$
$$6883\varepsilon - 6572 > 0$$
$$11313 - 9123\varepsilon > 0$$

$$
\begin{aligned}
18816\varepsilon - 17693 &> 0 \\
5094\varepsilon - 2281 &> 0 \\
10874\varepsilon - 4309 &> 0 \\
2167\varepsilon - 466 &> 0 \\
29468 - 25912\varepsilon &> 0
\end{aligned}
$$

The totality of these leads to the following condition.

$$0{\cdot}22 < \varepsilon < 1{\cdot}12$$

It implies that if we use the transportation cost matrix εC instead of C, then for each ε in the mentioned interval will give rise to the same functional pattern which is displayed in Table 1.

This does not imply that each ε outside this interval will lead to a different functional pattern! It is not easy to give complete sufficiency conditions.

If an $\varepsilon\,(> 1.12)$ is selected, there will be first of all a likely shift in the surplus (deficit) areas; some deficit regions will become surplus in the first stage of the new iteration, some surplus regions will become deficit area. This does not imply that this new transformation is stable; switches back are possible. More research has to be done in this respect. Anyhow, an underestimation of less than 78 per cent, and an overestimation of less than 17 per cent of the transport costs, would not have affected the functional pattern. The pattern as we have established it, makes quite a solid impression therefore!

It is difficult to deal with the cost of different ε's for different routes. Research is currently undertaken in this respect.

We can interpret the result in another way : a uniform change of transport costs between – 78 per cent and + 12 per cent, will not affect the functional pattern.

A similar analysis could be set up for the coefficients of the demand equation.

References :

(1) BAUMOL, W. J., 'Spatial Equilibrium with Supply Points separated from Markets with Supplies Predetermined'.
Ditto Report, Bureau of Agricultural Economics, U.S. Department of Agriculture, Washington D.C., 1952.

(2) BECKMANN, M., 'A continuous Model of Transportation', *Econometrica*, Vol. 20, 1952.

(3) DORFMAN, R., P. A. SAMUELSON and R. M. SOLOW, *Linear Programming and Economic Analysis*, Mcgraw Hill Book Co., New York, 1958.
(4) ENKE, S., 'Equilibrium Among Spatially Separated Markets : Solution by Electric Analogue', *Econometrica*, Vol. 19, 1951.
(5) FOX, K. A., 'A Spatial Equilibrium Model of the Livestock feed Economy', *Econometrica*, Vol. 21, 1952.
(6) FOX, K. A. and R. C. TAEUBER, 'Spatial Equilibrium Models of the Livestock-Feed Economy', *American Economic Review*, Vol. 45, 1955.
(7) HENRY, W. R. C. E. BISHOP, 'North Carolina Broilers in Inter-regional regional Competition', *Agricultural Economic Information Series No. 56*, North Carolina State College, 1957.
(8) JUDGE, G. G. and T. D. WALLACE, 'A Spatial Price Equilibrium Analysis of the Livestock Economy', TB-78, Department of Agricultural Economics Oklahoma State University, 1959.
(9) LOSCH, A., *The Economies of Location* (English Translation), Yale University Press, New Haven, Connecticut, 1954.
(10) SAMUELSON, PAUL A., 'Spatial Price Equilibrium and Linear Programming', *American Economic Review*, Vol. 42, 1952.
(11) WALTER ISARD, *Location and Space-Economy*, Technology Press of Massachusetts Institute of Technology, and John Wiley and Sons Inc.; New York, 1956.

Summary

In this paper a spatial equilibrium model has been applied to the rice-economy of South India. After a brief review of spatial equilibrium models considered earlier, the spatial equilibrium model considered here is specified. South India is divided into 11 relatively homogeneous regions, each region being a group of some districts. On the assumptions of a closed overall region (i.e. the region of South India), fixed regional supplies of rice, identical demand functions for rice in the 11 regions and an unrestricted capacity for transportation of rice by rail, equilibrium solutions are obtained. Some of the assumptions made are not very realistic. They need to be dropped and equilibrium solutions with the relaxed assumptions need to be obtained. The relevance of such spatial models for the formulation of economic policy has also been briefly discussed. Finally, in the appendix, a set of sufficient conditions for stability of the functional pattern displayed by the equilibrium solution with reference to variations or measurement errors in transportation-costs have been obtained.

सारांश

दक्षिण भारत की चावल अर्थरचना के अवकाशिक[1] संतुलन प्रतिदर्शों का रीतिविधान[2]

एम्. जी. कणबूर और एच्. न्यूडेक्कर

दक्षिण भारत की चावल अर्थरचना के लिए एक अवकाशिक संतुलन प्रतिदर्श इस लेख में प्रयुक्त किया गया है. इसके पहले अवकाशिक संतुलन प्रतिदर्शों के बारे में किये गये विचार का संक्षेप में पुनर्विलोकन करने के उपरांत यहां जिस अवकाशिक संतुलन प्रतिदर्श का विचार किया गया है उसका निर्देश प्रस्तुत है. दक्षिण भारत को सापेक्ष रूप से समरूप ग्यारह विभागों में बांटा गया. हरएक विभाग कुछ ज़िलों का समूह रहा. एक संवृत समावेशक विभाग (अर्थात् दक्षिण भारत विभाग), चावल का नियत विभागीय संभरण[3] ग्यारह विभागों में चावल की एक-समान मांग का कार्य, और रेलद्वारा चावल के परिवहन की अनिर्बंधित क्षमता,—इन सारी मान्यताओं पर संतुलन के हल प्राप्त किये गये, इनमें से कुछ मान्यताएं बहुत यथार्थवत् नहीं हैं. उन्हें छोड़ ही देना और शिथिलीकृत[4] मान्यताओं से संतुलन हलों को प्राप्त करना आवश्यक है. आर्थिक नीति के निर्धारण के लिए ऐसे अवकाशिक प्रतिदर्शों की संगतता की भी संक्षेप में यहां चर्चा की गयी है. लेख के अंत में परिशिष्ट के अंतर्गत, विचरणों या परिवहन-लागतों के बारे में मापन त्रुटियों संबंधी संतुलन हल द्वारा दिग्दर्शित फलनात्मक संरचना की स्थिरता के लिए आवश्यक स्थितियां पायी गयी हैं.

१. अवकाशिक – spatial
२. रीतिविधान – methodology
३. संभरण – supply
४. शिथिलीकृत – relaxed

ECONOMICS OF PLANNING
Vol. 6, No. 1, 1966
Printed in Norway

Montias' procedure for balancing organizations' outputs, goods outputs and final goods demand

Heinz Neudecker
Department of Econometrics, University of Birmingham

PRESENT LIMITATIONS AND POSSIBLE IMPROVEMENTS

In his book, *Central Planning in Poland* (pp. 345 onward), J. M. Montias [2] defines an m-dimensional vector r of values of gross outputs of m different organizations (e.g. ministries, central boards or multiplant enterprises), an n-dimensional vector y of the given bill of net outputs (final demand) and an n-dimensional vector x of total gross outputs of the n goods.[1]

In general, there will be more goods than organizations producing them, so $m<n$.

The vectors r, y and x are related by the equations

$$x=Br+y \tag{1}$$

– total output equals the intermediate use *plus* final deliveries –

$$x=Dr \tag{2}$$

– total output equals the output of the organizations –
where B and D are $n \times m$ matrices, the matrix B having the coefficients

$$b_{is}\ (i=1...n;\ s=1...m)$$

which denote the input of good i per value unit of output of organization s, the matrix D having the coefficients d_{is} which denote the amount of output of good i produced per value unit of output of organization s.

It is assumed therefore that "each organization produces its different products in fixed proportions".

The author sets out "to find the vector r capable of sustaining the given bill of net outputs y"; he intends to show that "it is possible for the planners to frame an internally consistent program without direct

[1] In his book, Montias used an o-dimensional r (o being a letter, not a number, of course). For reasons of clearness, I substituted m for o, changing quoted texts accordingly. My B and D matrices are not identical with Montias'. I used $\bar{B}$ and $\bar{D}$ for logical reasons, where he used B and D. If necessary I changed quotations accordingly.

knowledge of the technical coefficients", the input-output coefficients a_{ij} which would enable us to apply the renowned relationship $r=(I-A)^{-1}y$. The basic difficulty he meets with in solving this problem is the order of the matrices B and D. If B and D were square, the solution of the problem would be:

$$r=(D-B)^{-1}y \tag{3}$$

One would reach this solution by subtracting equations (1) and (2), and solving for r subsequently. (Of course, this solution presupposes that $(D-B)^{-1}$ exists, and that it is nonnegative.)

In reality this case of square B and D will seldom arise.

Montias gives a procedure for finding r (and x), in case $m<n$. "We may circumvent this obstacle ($m<n$, *HN*) by choosing m goods out of the total n and by working exclusively with these goods for the first part of the solution. It is assumed that it is possible to pick out these goods in such a way that none of them will be produced by more than one organization. (This should not be difficult in practice.) The choice should fall mainly on the more important goods, from the viewpoint of their priority standing and from their usual scale of output", he says.

Thus, he selects corresponding square fragments of the order m from D and B (let us call them $\bar{D}$ and $\bar{B}$), such that $\bar{D}$ is a matrix with only one element per row. Implicitly, he assumes that none of the organizations is producing nothing of these goods, so $\bar{D}$ also has only one element per *column*. "The matrix $\bar{D}$ is made up of a single element d_{ij} per row and per column". Montias postulates that "every good x_i can sustain some net output y_i after satisfying the intermediate requirements

$$\sum_{s=1}^{m} b_{is}r_s". \qquad (i=1 \ldots\ldots m).$$

At the same time, the author curtails y correspondingly.

By renumbering organizations (or goods) he transforms $\bar{D}$ into a diagonal matrix, and correctly says that $(\bar{D}-\bar{B})^{-1}$ exists and has nonnegative elements. As far as I can see, there is no need, however, to assume that each organization produces only one of these m commodities and has a *net* output for this. It is possible to work with a system that is much looser; each organization should have a *net* output for one product, and can have negative or zero outputs for all others. We require then that the system should be able to produce at least one positive net output vector, in other words to be 'productive'.

In symbols: the character of $\bar{D}-\bar{B}$ should be such that

$$\bar{d}_{ij}-\bar{b}_{ij} \leq 0 \text{ for } i \neq j$$
$$\bar{d}_{ii}-\bar{b}_{ii} > 0;$$

moreover, there should exist a nonnegative $\bar{r}$ such that $\bar{D}\bar{r} > \bar{B}\bar{r}$, in other words $y=(\bar{D}-\bar{B})\bar{r}>0$. It is easy to prove now that $(\bar{D}-\bar{B})^{-1}$ exists and is positive. We apply theorem 9.1 and its corollary, in Gale (1), p. 296 seq.[1] In effect, some redefinitions are necessary to make the application of the theorem obvious. Let us first redefine: $\bar{D}-\bar{B}=A^*-A^{**}$ such that

$$a_{ij}^{*} = \delta_{ij}(\bar{d}_{ij}-\bar{b}_{ij})$$
$$a_{ij}^{**} = (\delta_{ij}-1)(\bar{d}_{ij}-\bar{b}_{ij})$$
$$-\delta_{ij} \text{ is the kronecker delta} -;$$

A^* and A^{**} are nonnegative matrices, A^* being diagonal, A^{**} being hollow. In this new notation we require that there exists a nonnegative $\bar{r}$ such that $A^*\bar{r} > A^{**}\bar{r}$, or $\bar{r} > A^{*-1}A^{**}\bar{r}$.

Gale has proved that $(I-A^{*-1}A^{**})^{-1}$ exists and is positive. In other words, $(A^*-A^{**})^{-1}A^*$ exists and is positive. It is clear then that $(A^*-A^{**})^{-1}$ exists, and is also positive.

In this article, we will therefore not confine ourselves to $\bar{D}-\bar{B}$ matrices, whose $\bar{D}$ is diagonal (or reducible to a diagonal matrix). There is no point in doing so.

Let us follow Montias' procedure, but now for a more general $\bar{D}-\bar{B}$. He restricts the system of equations (1) and (2) to:

$$\bar{x}=\bar{B}r+\bar{y} \tag{1'}$$
$$\bar{x}=\bar{D}r \tag{2'}$$

and derives a solution

$$r=(\bar{D}-\bar{B})^{-1}\bar{y} \tag{3'}$$

The r he finds, Montias inserts in equation (2) and gets then:

$$x=D(\bar{D}-\bar{B})^{-1}\bar{y} \tag{4}$$

This x is used to (re)find y by means of equation (1):

$$y=x-Br=D(\bar{D}-\bar{B})^{-1}\bar{y}-B(\bar{D}-\bar{B})^{-1}\bar{y}=(D-B)(\bar{D}-\bar{B})^{-1}\bar{y} \tag{5}$$

[1] I am grateful to J. Butterworth for having applied the Gale theorem to this case. The proof is entirely his. The whole paper, in fact, has benefited from his comments.

It must be clear at once that this y will *generally* be different from the y the author started with, still worse, it can have negative elements.

We shall show this. Select $\bar{y}$. Partition then

$$D-B=\begin{bmatrix}\bar{D}-\bar{B}\\ \cdots\cdots\\ \tilde{D}-\tilde{B}\end{bmatrix};$$

this gives

$$(D-B)(\bar{D}-\bar{B})^{-1}\bar{y}=\begin{bmatrix}\bar{D}-\bar{B}\\ \cdots\cdots\\ \tilde{D}-\tilde{B}\end{bmatrix}(\bar{D}-\bar{B})^{-1}\bar{y}=$$

$$\begin{bmatrix}(\bar{D}-\bar{B})(\bar{D}-\bar{B})^{-1}\bar{y}\\ \cdots\cdots\cdots\cdots\\ (\tilde{D}-\tilde{B})(\bar{D}-\bar{B})^{-1}\bar{y}\end{bmatrix}=\begin{bmatrix}\bar{y}\\ \cdots\cdots\cdots\cdots\\ (\tilde{D}-\tilde{B})(\bar{D}-\bar{B})^{-1}\bar{y}\end{bmatrix}$$

Although $(\bar{D}-\bar{B})^{-1}\bar{y}\geqq 0$, there is no need that $\tilde{y}=(\tilde{D}-\tilde{B})(\bar{D}-\bar{B})^{-1}\bar{y}$ is nonnegative; because $(\tilde{D}-\tilde{B})$ can have negative elements.

Montias sees this problem, but confines himself to saying: "If in a given case the resulting bill of net outputs y_{m+1} to y_n were not satisfactory, then net outputs y_1 to y_m would have to be adjusted in order to release resources for the remaining sectors. The procedure would have to be repeated until the net output proportions in vector y for all n goods were deemed satisfactory."

Thus, he suggests manipulating the $\bar{y}$ vector in such a way that the last $n-m$ elements of the y vector emerging from the computation process (denoted by $\tilde{y}$) are acceptable, i.e. nonnegative and then to adopt the vector $\begin{bmatrix}\bar{y}\\ \tilde{y}\end{bmatrix}$ as y vector. This is, of course in flagrant contradiction with the conditions of the original problem, where y was *given* and not *manipulable!* Alas.

I want to conclude from this that it is, in general, impossible to select y arbitrarily when there are no stocks brought into the picture, to make up for inadequate components of $\tilde{y}$, or to absorb oversatisfactory components. If there is no possibility of introducing stocks, the set of y vectors will have to be *restricted.* It is not difficult to see that acceptable nonnegative y vectors have to satisfy the relation.

$$y = Hy$$

where

$$H = \begin{bmatrix} I & \vdots & O \\ \cdots\cdots\cdots\cdots & \vdots & \cdots\cdots \\ (\tilde{D}-\tilde{B})(D-B)^{-1} & \vdots & O \end{bmatrix}^{1}$$

It is obvious that if there exist more than one nonnegative characteristic vectors corresponding to characteristic value 1, all positive linear combinations of these vectors will also satisfy the relation $y=Hy$. The set is infinite, then. In both cases (unrestricted and restricted y) there arises a selection problem, which y to choose!

THE CHOICE PROBLEM; SOME POSSIBLE FORMULATIONS

The choice problem can be tackled in two different ways.

Firstly, by admitting only *implicit* valuations in the choice made by the policymakers and implemented by the planners. Secondly, by formulating *explicit* valuations (preference functions) for the planners, and leaving the choice to them. In both procedures the planners have a mere technical role. In the first, the policy-makers choose a y (on basis of an implicit valuation), submit it to the planners, who test it for feasibility; if it is feasible, they compute x and r. If it is not feasible, they communicate this to the policy-makers and invite them to make a second choice, and so on.

Both the original and the amended Montias procedure belong to the first class. Because the feasible y (satisfying: $y=Hy$) probably form an infinite set, there is no a priori practical delimitation of the y set possible; so there is no difference for the policy-makers. In both cases they choose in vacuo. This is hardly acceptable.

An alternative approach is that of *explicit* valuation. Here preferences are made objective and discussable.

We shall tentatively formulate some approaches in this spirit.

First of all, we shall stick to the *original* formulation given by Montias, in which y is predetermined. If y does not belong to the set ($y=Hy$), the following approach is useful.

If we inspect the equations (1) and (2) again, we see that equation (1)

[1] It is not the only matrix representing this idempotent transformation. We can arbitrarily choose $\bar{D}-\bar{B}$ that meets the requirements and a corresponding $\tilde{D}-\tilde{B}$ from $D-B$.

is an input equation and equation (2) is an output equation. Equation (1) determines x from an input point of view: input into the organizations and the final consumption sectors. Equation (2) establishes an identity relation between x and r: the output vector in terms of goods is directly related to the output vector of the organizations. Equation (1) gives a possibility of introducing a disturbance. So we may rewrite the equations (1) and (2) in this way:

$$x = Br + y + v \qquad (1'')$$
$$x = Dr$$

where v is the difference between Dr and $Br+y$:

$$v = (D-B)r - y \qquad (6)$$

v can be interpreted as an *increase* in stocks.

An acceptable solution for x and r has to optimize stocks-behaviour, we could say. For some goods, there will be stocks available from which these could be withdrawn. For other goods, there will be no stocks available, so that withdrawals are inconceivable.

Let us write

$$v = z_1 - z_2$$

where z_1 is the vector of increases in stocks and z_2 is the vector of decreases, so that v is the vector of *net* increases in stocks. Suppose further that both increases and decreases are bounded:

$$o \leqq z_1 \leqq w$$
$$o \leqq z_2 \leqq u,$$

i.e. the increase is maximally w, and the decrease is maximally u.

A reformulation of Montias' problem could then be:

Minimize $\quad z_1' A^{(1)} z_1 + z_2' A^{(2)} z_2$

subject to
$$z_1 \leqq w$$
$$z_2 \leqq u$$
$$(D-B)r = y + z_1 - z_2$$
$$r \geqq 0$$
$$z_1 \geqq 0$$
$$z_2 \geqq 0$$

The matrices $A^{(1)}$ and $A^{(2)}$ are in the simplest case diagonal, with *positive* elements $a_{ii}^{(1)}$ and $a_{ii}^{(2)}$, and these indicate the desirability of increases and decreases.

The more desired an increase in stocks is, the smaller will be the corresponding element in $A^{(1)}$; the more desired a decrease is, the smaller will be the corresponding element in $A^{(2)}$. If an increase for good i is less desired than a decrease, $a_{ii}^{(1)}$ will be greater than $a_{ii}^{(2)}$.

What we do here is essentially minimizing the weighted sum of squares of increases and decreases of stocks. We can allow for cross-effects as well. Then $A^{(1)}$ or $A^{(2)}$ are no longer diagonal. We have to impose the condition that they are symmetric positive semi-definite. The problem as we stated it now, is a quadratic programming problem. A general solution method can be found in Van de Panne and Whinston [3].

In case y belongs to the set ($y=Hy$), this optimization procedure produces the result:

$$z_1=z_2=0.$$

If y is subject to choice, and no stocks enter into the picture, a choice procedure could be:

$$\text{Maximize } y'Ay \equiv r'\,(D-B)'\,A(D-B)r$$
$$\text{subject to}$$
$$0 \leqq r \leqq a$$
$$0 \leqq (D-B)r$$

or:

$$\text{Maximize } y'Ay \equiv r'\,(D-B)'A(D-B)r$$
$$\text{subject to}$$
$$x \equiv Dr \leqq b$$
$$0 \leqq r$$
$$0 \leqq (D-B)r$$

In these preference functions A should be symmetric negative semi definite. In the first optimization procedure we introduced capacity constraints for the organizations. These are necessarily in financial terms.

In the second procedure there are physical capacity constraints for the products. These approaches have not exhausted all possibilites of formulating a choice procedure! Linear combinations of y could be added.

REFERENCES

[1] Gale, D., *The Theory of Linear Economic Models,* 1960.

[2] Montias, J. M., *Central Planning in Poland,* 1962.

[3] Van de Panne, C. and Whinston, A., The Simplex and the Dual Method for Quadratic Programming (*Operational Research Quarterly,* Vol. 15).

Interne leveringen in de inzet-afzetanalyse

door H. Neudecker *)

UDC 512.8:339.21

Summary

Input-output analysis can be set up in gross or net terms. In the first case the internal deliveries of a sector are included in the table, in the other case they are neglected. These two different approaches give rise to two different leontief matrices: A and D.

Stone determined a relation between the two matrix multipliers $(I - A)^{-1}$ *and* $(I - D)^{-1}$.

By applying this relation all results in the two definition systems can be translated into each other. In this note we will analyze some relationships between increments in gross and net production.

It turns out that in case the external delivery structure changes (the non-diagonal coefficients of A), relative increments in gross and net production are identical.

A further result is that neither relative nor absolute increments in gross and net production are identical in case of changes in the internal delivery structure (the diagonal coefficients of A).

The third result is that relative increments in gross and net production are identical in case of a change in final demand.

0.1. Inleiding

Het wordt in de inzet-afzetanalyse niet essentieel geacht of men met netto- of met brutoproduktie rekent [1]). Men kan de interne leveringen al dan niet bij de analyse betrekken [2]).

Stone bepaalde in zijn werk "*Input-output and national accounts*" het verband tussen de matrixmultiplicatoren die gebaseerd zijn op brutoproduktie en nettoproduktie respectievelijk [3]). Hieruit blijkt dat het niet van belang is in welk definitie-systeem – hetzij dat van de brutoproduktie, hetzij dat van de netto-

*) Landbouw-Economisch Instituut, 's-Gravenhage. De schrijver is dank verschuldigd aan G. HAMMING en C. L. SCHEFFER voor hun kritische opmerkingen.

[1]) In navolging van P. VENEKAMP spreken wij van inzet-afzetanalyse in plaats van input-outputanalyse. Wij beschouwen een open stelsel.

[2]) Onder „interne leveringen" verstaan wij de leveringen van een produktiesector aan zich zelf. Onder „externe leveringen" verstaan wij de leveringen van een produktiesector aan een andere.

[3]) De matrixmultiplicatoren transformeren de uiteindelijke consumptie door de consumenten (ook wel „autonome vraag" genoemd) in sectorproduktie (bruto of netto).

produktie – men inzet-afzetanalyse bedrijft. Het is eenvoudig bruto- en nettoproduktie in elkaar om te rekenen.

In deze notitie interesseert ons welk verband er is tussen de *mutaties* in bruto- en nettoproduktie die terug te voeren zijn op veranderingen in de inzet-afzetstructuur en de autonome vraag. Onder „inzet-afzetstructuur" verstaan wij dan de leontief matrix die mede gebaseerd is op interne leveringen.

Het gaat er dan om aan te geven het verband tussen *mutaties* in brutoproduktie en nettoproduktie in het geval van veranderingen in

1. de niet-diagonale inzetcoëfficiënten van de op *bruto*produktie berustende leontief matrix;
2. de diagonale inzetcoëfficiënten van de op *bruto*produktie berustende leontief matrix;
3. de autonome vraag. [1]).

Dit is het vraagstuk waarmee we ons hier bezig zullen houden.

0.2. Inleiding

In het geval van interne leveringen is de inzet-afzetrelatie

$$p = (I - A)^{-1}c \tag{1}$$

Hierin is A de leontief matrix, gebaseerd op brutoproduktie

p de brutoproduktie-vector
c de vector van de autonome vraag.

(Voor de elementen a_{ii} van A geldt: $0 \leqslant a_{ii} < 1$).
Als de interne leveringen verwaarloosd worden, is de inzet-afzetrelatie:

$$q = (I - D)^{-1}c. \tag{2}$$

Hierin is D de leontief matrix, gebaseerd op nettoproduktie

q de nettoproduktie-vector
c de vector van de autonome vraag.

(De elementen d_{ii} van D zijn $= 0$).
Stone geeft in zijn boek (VIII,4) als relatie tussen de matrix-multiplicatoren $(I - A)^{-1}$ en $(I - D)^{-1}$:

$$(I - D)^{-1} = QP^{-1}(I - A)^{-1} \tag{3}$$

waar Q en P de in diagonaalmatrixvorm genoteerde q en p zijn. De juistheid van de formules (1), (2) en (3) is gemakkelijk in te zien. Het betoog verloopt als volgt:

[1]) 1, 2 en 3 hebben betrekking op resp. de externe leveringen, de interne leveringen en de uiteindelijke consumptie.

Als Z de matrix van interne en externe leveringen is en Q_0 de in diagonaalmatrixvorm genoteerde nog te definiëren q_0 en i de sommatievector is (met alle elementen = 1), dan definiëren we de volgende relaties

$$q_0 = p - q \quad \text{(definitie van } q_0\text{)}$$
$$Q_0 = Z - Z^x \quad \text{(definitie van } Z^x\text{)}$$

waarna volgt:

$$A = ZP^{-1} \therefore API = Zi \therefore Ap = p - c, \text{ waaruit (1) volgt}$$
$$D = Z^xQ^{-1} \therefore DQi = Z^xi$$
$$\therefore Dq = q - c, \text{ waaruit (2) volgt.}$$
$$I - D = I - Z^xQ^{-1} = I - (Z - Q_0)Q^{-1} = I - (Z - P + Q)Q^{-1}$$
$$(I - D)Q = Q - (Z - P + Q) = P - Z = P - AP = (I - A)P$$
$$I - D = (I - A)PQ^{-1}$$
$$\therefore (I - D)^{-1} = QP^{-1}(I - A)^{-1}$$

Hiermee is (3) dus bewezen.

0.3. Inleiding

Wegens $Q^{-1}q = P^{-1}p$, geldt:

$$q = QP^{-1}p. \tag{4}$$

Omdat $q_i = (1 - a_{ii})\,p_i$, kunnen we ook schrijven:

$$q = (I - B)\,p \tag{4'}$$

waarbij B een diagonale matrix is met als elementen de diagonale elementen van A. Het is duidelijk dat

$$QP^{-1} = I - B \tag{4''}$$

omdat [illegible] $q_i = (1 - a_{ii})\,p_i$

Wij postuleren dat elke sector nettoproduktie heeft, m.a.w. dat Q^{-1} bestaat. Dan bestaat ook $(I - B)^{-1}$.

Wij kunnen nu de verbanden tussen de mutaties in brutoproduktie en nettoproduktie op eenvoudige wijze aangeven.

1.1. Verband tussen de mutaties in brutoproduktie en nettoproduktie als slechts de niet-diagonale inzetcoëfficiënten veranderen

In dit geval zijn dus de a_{ij}'s constant. (Ook c is constant maar van deze eigenschap behoeft geen gebruik te worden gemaakt.) Het betoog verloopt als volgt:

Uit (4') volgt $\Delta q = (I - B)\Delta p$, omdat immers $\Delta B = 0$. Noodzakelijk en voldoende voor $\Delta q_i = \Delta p_i$ is $a_{ii}\Delta p_i = 0$. In het geval een sector geen interne

leveringen heeft ($a_{ii} = 0$) geldt: $\Delta p_i = \Delta q_i$. Heeft een sector wel interne leveringen, dan is in het algemeen $\Delta p_i \neq \Delta q_i$.

Δq en Δp vallen derhalve in het algemeen niet samen. Wel geldt:

$$Q^{-1}\Delta q = P^{-1}\Delta p, \text{ want}$$
$$Q^{-1}\Delta q = Q^{-1}(I-B)\Delta p = P^{-1}\Delta p$$
wegens (4') en (4'').

De *relatieve* mutaties in brutoproduktie en nettoproduktie zijn dus gelijk, als slechts de niet-diagonale inzetcoëfficiënten veranderen.

Hieruit volgt dat de interne leveringen, zolang de interne leveringsverhoudingen niet veranderen, buiten de analyse kunnen worden gehouden.

1.2. Verband tussen de mutaties in brutoproduktie en nettoproduktie als slechts de diagonale inzetcoëfficiënten veranderen

In dit geval veranderen dus alleen de a_{ii}'s. De finale vraag c verandert niet.

Vonden wij in I.1. dat het voor de *relatieve* produktieverandering irrelevant is of we interne leveringen verwaarlozen, zolang de interne leveringsverhoudingen niet veranderen, nu zullen we zien dat als aan deze voorwaarde niet is voldaan en als de finale vraag niet verandert, er divergenties optreden in de mutaties.

Uit (4') volgt:

$$\Delta q = -(\Delta B)p + (I-B)\Delta p - (\Delta B)\Delta p = -(\Delta A)p + (I-B)\Delta p - (\Delta A)\Delta p.$$

Omdat $\Delta c = 0$, geldt:

zodat:
$$(\Delta A)p = (I-A)\Delta p - (\Delta A)\Delta p,$$
$$\Delta q = -(I-A)\Delta p - (I-B)\Delta p = (A-B)\Delta p. \tag{5}$$

De relatieve mutaties zijn in het algemeen niet aan elkaar gelijk.

Immers: $Q^{-1}\Delta q = Q^{-1}(A-B)\Delta p$ krachtens (5).
Verder: $Q^{-1}(A-B)\Delta p = P^{-1}(I-B)^{-1}(A-B)\Delta p$ wegens (4'').
Gelijkheid van $Q^{-1}\Delta q$ en $P^{-1}\Delta p$ impliceert dan

$$P^{-1}[I-(I-B)^{-1}(A-B)]\Delta p = 0.$$

Als niet-triviale noodzakelijke voorwaarde voor deze gelijkheid hebben we

$$|I-(I-B)^{-1}(A-B)| = 0.$$

Echter, $|I-(I-B)^{-1}(A-B)| = 0$ impliceert $|I-B|\,|I-(I-B)^{-1}(A-B)| = |I-B-(A-B)| = |I-B-A+B| = |I-A| = 0$.

Als niet-triviale noodzakelijke voorwaarde voor de gelijkheid van $Q^{-1}\Delta q$ en $P^{-1}\Delta p$ vinden we dus het niet-bestaan van $(I-A)^{-1}$.[1]

[1]) Voor de inverteerbaarheidsvoorwaarden van $I-A$ wordt naar de literatuur verwezen.

Uit (5) volgt als niet-triviale noodzakelijke voorwaarde voor $\Delta q = \Delta p$ $|I - A + B| = 0$. Het is eenvoudig in te zien dat $(I - A + B)^{-1}$ in het algemeen bestaat, zodat $|I - A + B| \neq 0$.

Bij het bewijs hiervan zal een beroep worden gedaan op de theorie van de genormeerde lineaire ruimten. Volgens de definitie van een genormeerde ruimte is de produktieruimte waar wij het over hebben genormeerd als met elk element x een niet-negatief getal $\| x \|$ correspondeert, de „norm" van x, zodanig dat

1) $\| x \| = 0$ als en slechts als $x = 0$
2) $\| ax \| = |a| \cdot \| x \|$
3) $\| x + y \| \leqslant \| x \| + \| y \|$

Het is duidelijk dat de lengte van een vector, d.w.z. voor x: $(\sum x_i^2)^{\frac{1}{2}}$ aan deze normdefinitie voldoet.

De produktieruimte is dus het positieve deel van een (eindig dimensionale) genormeerde lineaire ruimte $\mathscr{B}$. De transformatie, waarvan Z de matrix is, heet begrensd als er een niet-negatief getal m bestaat zodat voor alle x in $\mathscr{B}$

$$\| Zx \| \leqq m \| x \| \tag{6}$$

geldt. Het kleinste getal m met de eigenschap dat (6) nog geldt heet de norm van Z:

$$\| Z \| = \max_{x} \frac{\| Zx \|}{\| x \|} \tag{7}$$

Men kan nog opmerken dat, als Z een positieve transformatie voorstelt (d.w.z. de eigenschap heeft positieve vectoren over te voeren in positieve; dit is dus zeker waar als Z uitsluitend niet negatieve elementen heeft), men het max in (7) al kan berekenen door x alleen positieve vectoren te laten doorlopen.

Als de norm van Z kleiner dan 1 is, heeft $I - Z$ een inverse. (Op deze eigenschap wordt hier niet nader ingegaan; verwezen zij naar de desbetreffende literatuur).

In ons geval gaat het er nu om te bewijzen dat $\| A - B \|$ kleiner dan 1 is. Voor dit bewijs is het inzicht nodig dat $\| y \| < \| x \|$ als $y \leqq x$ (voor $x, y \geqslant 0$; $x \neq y$) Het is duidelijk dat de lengte van een vector deze eigenschap heeft.

Krachtens de relaties (1) en (4') en de definitie van q_0 geldt

$$(A - B)\, p = p - c - q_0.$$

Zolang c en q_0 niet beide gelijk nul zijn, geldt dus

$$(A - B)\, p \leqq p.$$

(Noodzakelijk voor $c \neq 0$ is dat $(I - A)^{-1}$ bestaat. Noodzakelijk en voldoende voor $q_0 \neq 0$ is dat minstens één sector interne leveringen heeft).

In overeenstemming met de boven verkregen resultaten hebben we dan voor $p \geqq 0$, dus (zie de opmerking hierboven) voor alle p in $\mathcal{B}$

$$\| (A - B)\, p \| < \| p \|,$$

maar volgens de definitie van de norm van $A - B$ geldt tevens

$$\nu = \max_{p} \frac{\| (A-B)p \|}{\| p \|} \qquad \| (A-B)p \| \leq \nu \| p \|,$$

waarbij ν de norm van $A - B$ is.

Hieruit concluderen we dat $\| A - B \| < 1$ is.[1])

Omdat de norm van $A - B$ kleiner dan 1 is, heeft $I - A + B$ een inverse.

De gevolgtrekking hieruit is dat in het algemeen $\Delta q \neq \Delta p$.

1.3. Verband tussen de mutaties in brutoproduktie en nettoproduktie, als slechts de autonome vraag gevarieerd wordt

Als slechts de autonome vraag verandert, geldt krachtens (4′):

$$\Delta q = (I - B)\Delta p$$

Noodzakelijk en voldoende voor $\Delta p = \Delta q$ is $a_{ii}\Delta p_i = 0$. In het geval dat een sector geen interne leveringen heeft ($a_{ii} = 0$) geldt $\Delta p_i = \Delta q_i$. Heeft een sector wel interne leveringen, dan is in het algemeen $\Delta p_i \neq \Delta q_i$.

$Q^{-1}\Delta q$ en $P^{-1}\Delta p$ zijn identiek. Dit ziet men gemakkelijk in:

uit $\Delta q = (I - B)\Delta p$ volgt wegens (4″):

$$Q^{-1}\Delta q = Q^{-1}(I - B)\Delta p = P^{-1}\Delta p \qquad \text{q.e.d.}$$

Dit wil zeggen dat de relatieve mutaties in bruto- en nettoproduktie identiek zijn in het geval slechts de autonome vraag verandert.

2. Naschrift

Uit (3) en (4″) volgt een alternatieve formulering van STONE's relatie, namelijk

$$(I - D)^{-1} = (I - B)\,(I - A)^{-1}$$

Deze heeft het voordeel geheel in termen van inzetcoëfficiënten te luiden.

Literatuurverwijzingen

[1] BEAR, D. V. T., D. JORGENSON en H. M. WAGNER, "*Elementary proofs of propositions on leontief-minkowski matrices*" (Metro-economica, volume 14).

[2] KOLMOGOROV, A. N. en S. V. FOMIN, "*Elements of the theory of functions and functional analysis*", volume 1.

[3] STONE, R., "*Input-output and national accounts*".

[4] ZAANEN, A. C., "*Linear analysis*".

[1]) Noot van de redactie: deze conclusie is juist omdat het hier een e i n d i g-dimensionale genormeerde ruimte betreft.

METROECONOMICA

RIVISTA INTERNAZIONALE DI ECONOMICA
RÉVUE INTERNATIONALE D'ÉCONOMIQUE
INTERNATIONAL REVIEW OF ECONOMICS

VOLUME XV AGOSTO-DICEMBRE 1963 FASCICOLO II-III

COMITATO DIRETTIVO
COMITÉ de DIRECTION *BOARD of EDITORS*

C. BRESCIANI TURRONI
Università, Milano

LÉON DUPRIEZ
Università, Lovanio

ERALDO FOSSATI †
Università, Genova

RAGNAR FRISCH
Università, Oslo

WASSILY LEONTIEF
Harvard University
Cambridge, Mass. (U.S.A.)

GÉORGES LUTFALLA
Ec. nat. org. ec. et. soc.
Parigi

ALEXANDER MAHR
Università, Vienna

J. MARSCHAK
Cowles Foundation
Yale University, New-Haven
(U.S.A.)

G. UGO PAPI
Università, Roma

FRANÇOIS PERROUX
Professore al College de France

R. W. PFOUTS
Università della North Carolina
Chapel-Hill (U.S.A.)

MANLIO RESTA
Università, Trieste
Managing Editor

RENÉ ROY
Éc. Nat. Ponts et Chaussées
Parigi

ERICH SCHNEIDER
Università, Kiel

G. L. S. SHACKLE
Università, Liverpool

RICHARD STONE
Università, Cambridge

J. TINBERGEN
Scuola Sup. di Economia
Rotterdam

GERHARD TINTNER
University of Southern California
Department of Economics
University Park
Los Angeles 7, California

A NOTE ON INPUT-OUTPUT ANALYSIS

by H Neudecker

EDITORE EDITEUR PUBLISHER
LICINIO CAPPELLI - BOLOGNA-TRIESTE

Depositi - Dépôts - Agents

G. E. STECHERT-HAFNER Inc.
31 East 10th Street
NEW-YORK, 3 N. Y. (U.S.A.)

CENTRO IMPORTADO LIBRO ITALIANO
Railén, 159, esa. Rosellon
BARCELLONA

EINAR MUNKSGAARD
Nörregade 6 COPENHAGEN

LIBRAIRIE ENCYCLOPEDIQUE
7 Rue du Luxembourg
BRUXELLES IV

LIBRAIRIE PAYOT
LAUSANNE

SPRINGER VERLAG
Mölkerbastei WIEN 1

INDICE

TABLE DES MATIÈRES — CONTENTS

BALL-BODKIN - Income, the price level, and generalized multipliers in Keynesian economics 59

RONALD E. MILLER - Comments on the «General Equilibrium» model of Professor Moses 82

K. MIYAZAWA - S. MASEGI - Interindustry analysis and the structure of Income-Distribution 89

HEINZ STÖWE - Einige bemerkungen zur geschichte der Parameterschätzung interdependenter ökonomischer Systeme 104

J. KMENTA - Estimates of the Cobb-Douglas production function: a reappraisal 117

C. TISDELL - Notes upon some of the present theories of choice under uncertainty 125

S. N. AFRIAT - An identity concerning the relation between the Paasche and Laspeyre indices 136

C. C. MORRISON - A note on estimating the concentration ratio with reference to Pareto distribution 141

L. W. McKENZIE - The turnpike theorem of Morishima 146

Z. S. ZANNETOS - Markov chain depreciation model and survivor curve depreciation 155

NEUDECKER - A note on input-output analysis 167

Books reviewed - Analisi di opere 168

Books received - Libri ricevuti 172

La rivista appare ogni quattro mesi. Manoscritti libri e periodici per recensione o annuncio vanno indirizzati al Prof. MANLIO RESTA, direttore dell'Istituto di Economia, dell'Università di Trieste. Le comunicazioni relative all'amministrazione e agli abbonamenti vanno indirizzate a Libreria Licinio Cappelli, Corso Italia 12, Trieste.

La rivista declina ogni responsabilità in rapporto ad opinioni o a fatti espressi dai collaboratori ed in rapporto ai manoscritti affidatile. Ogni collaboratore riceverà 20 estratti dello studio pubblicato.

* * *

La revue parait chaque quattre mois. Les manuscrits, les livres pour comptes-rendus, les periodiques doivent être adréssés au Prof. MANLIO RESTA, direttore dell'Istituto di Economia dell'Università di Trieste. Les communications rélatives à l'administration et aux abonnements doivent être adressées à Libreria Licinio Cappelli, Corso Italia 12, Trieste.

La révue décline toute responsabilité au sujet des opinions et des faits exprimés par ses collaborateurs et au sujet des manuscrits, qui lui sont confiés. La révue fournit gratuitement à ses collaborateurs 20 extraits de leurs articles publiés.

* * *

The review is published every four months. Communications for the editors, manucripts, books and periodicals to be addressed, Prof. MANLIO RESTA, direttore dell'Istituto di Economia dell'Università di Trieste. Business letters to be addressed: Libreria Licinio Cappelli, Corso Italia 12, Trieste.

Neither the editor nor the publisher take any responsability for opinions or facts stated by contributors and for the manuscripts sent them. Every contributor will receive 20 offprints of his published essay.

CONDIZIONI DI ABBONAMENTO
CONDITIONS D'ABONNEMENT
ANNUAL SUBSCRIPTION PRICE

ITALIA: tre numeri annuali ..... **L. 2.800**

FRANCIA: trois numéros annuels ..... **N. F. 36.00**

INGHILTERRA: three numbers a year ..... **Lg. 2.10.0**

USA and others countries: three numbers **a year** ..... **$ 7.00**

(two years **$ 12**)

SVIZZERA: trois numéros annuels ..... **Frs. Sv. 32**

A NOTE ON INPUT-OUTPUT ANALYSIS

by H. Neudecker

Let $x = (I - A) - {}^{\prime}c$ be the familiar equation of input-output analysis, where x is the total production vector, c the vector of final demand and A the matrix of technical coefficients.

Let $x_{ij} = a_{ij}x_j$ be the relationship between

x_j total production of the i^{th} sector

x_{ij} delivery from the i^{th} sector to the j^{th} sector

a_{ij} the ij^{th} element of A $\quad (a_{ii} = 0)$.

Let then the deliveries from all sectors to the j^{th} sector $(j = 1 \ldots J)$ change such that the total production of that sector has a constant value added proportion.

This implies

$$\frac{x_j^* - \overset{i}{\Sigma} x_{ij}^*}{x_j^*} = \frac{x_j - \overset{i}{\Sigma} x_{ij}}{x_j} \text{ or } \frac{\overset{i}{\Sigma} x_{ij}^*}{x_j^*} = \frac{\overset{i}{\Sigma} x_{ij}}{x_j} \text{ or } \overset{i}{\Sigma} a_{ij}^* = \overset{i}{\Sigma} a_{ij} \qquad (1)$$

By asterisks we denote the new entities.

We then prove the following theorem:

« The relative redistribution of deliveries from all sectors to the j^{th} sector implies that the production of *at least* one sector is constant or diminishing. (We assume constant final demand.) ».

Proof: $x = (I - A) - {}^{\prime}c$

$$\Delta x = (I - A) - {}^{\prime}\Delta A \, (I - A - \Delta A) - {}^{\prime}c$$

$$(I - A) \, \Delta x = \Delta A \, (I - A - \Delta A) - {}^{\prime}c$$

$$\pi' \, (I - A) \, \Delta x = \pi' \Delta A \, (I - A - \Delta A) - {}^{\prime}c = 0 \text{ , so the}$$

theorem has been proved because of (1) [1].

That $\pi' \, (I - A)$ consists of elements ≥ 0, at least one of which is > 0, one can easily find from the fundamental property of input-output analysis that at least one sector produces value. (This property is also used in inverting $I - A$).

---

[1] π' is the summation vector $(1 \ldots 1)$.

METRIKA, Vol. 2 No. 1 (1959)

Micro-Economic Business Test Data Compared With Traditional Statistics[1)]

By D. B. Jochems and H. Neudecker, Rotterdam[2)]

1. Introduction

This article is concerned with data from surveys which have in recent years been set up in the Netherlands and in several other countries according to the prototype of the Munich Business Test. These Business Tendency Surveys are based upon monthly questionnaires in which entrepreneurs are asked to state the actual direction of change of certain micro-economic variables (production, prices, etc.) in the preceding month (derived by comparing the level of the variable in that month with that of the previous month), as well as the anticipated direction of change in the next month[3)]. Hence, for each firm, each variable and each month, there are three possible answers in the sphere of realizations, viz., increase (+), no change (=) and decrease (−); and the same applies to expectations and plans[4)]. For a more detailed description in English we refer to Theil's expository article in Econometrica [17] and to a Committee Report on Economic Survey Data [22]. Here, we confine ourselves to observing that the method considered, after having been set up by the IFO-Institut für Wirtschaftsforschung at Munich, Germany, less than ten years ago, has been

1) The authors have highly valued the stimulating criticism of Professor Theil, who suggested this problem to them and was always willing to participate in the discussions.

2) D. B. Jochems and H. Neudecker, Econometric Institute of the Netherlands School of Economics. Rotterdam, Pieter de Hoochweg 122.

3) Most questions in the Dutch surveys refer to a time period of one month, while in Business Tendency Surveys of other countries, time periods of two, three and six months are not unusual. Moreover, there are various bases of comparison of the direction of change, e.g., the same period one year ago instead of one month ago.

4) The notation (+), (=), (−) is not fully satisfactory, but we adopt it here because it has become a tradition. More satisfactory notations are (+), (0), (−), indicating the sign of the first difference, or (>), (=), (<) indicating whether this month's level exceeds, is equal to, or is below last month's level.

taken over by agencies in several other countries, such as Austria, Belgium, Denmark, Japan, the Netherlands, the Union of South Africa, and Sweden[5].

It goes without saying that there are numerous problems connected with these surveys and their results. First, there is the problem of a rational statistical design in relation to the selection of participants among the firms of the various branches of manufacturing and commercial activity, while the non-response is another statistical problem. A recent meeting, held at Munich in July 1957, and devoted to Business Test problems, revealed that in most cases the statistical design is rather unsophisticated, the Konjunkturinstitutet at Stockholm, Sweden, being in some aspects a favourable exception. It seems clear that more substantial efforts in this direction are worth-while.

Quite another set of problems is that of the analysis of the results of these surveys. It is useful to make a distinction according to two types, viz., the analysis of the interrelationships of Business Test data themselves, and that of the relation between these data on the one hand and "traditional" (i.e., numerical) statistical data on the other hand. Examples of the first type of analysis are: the accuracy of the predictions of the directions of change for the next month (the plans and expectations mentioned earlier), measured by comparing these predictions with the actual direction of change as reported at the end of the month; the relationship between expected buying prices and planned selling prices, between actual buying and actual selling prices in successive stages of industry; the effects of changes in certain determinant variables on the plans of the entrepreneurs, etc. Considerable efforts in these directions have been made by Anderson and his collaborators [2, 3, 4, 5, 6] in Germany and by Theil [19, 20, 21] in the Netherlands, both on German Business Test data, while Hastay did similar work on Dun & Bradstreet data in the United States [7]. The second type of analysis (that of the relationship with conventional data) can be divided into two categories according to the macro- or the micro-level upon which the comparison of the two types of data is carried out. First, there is the macro-economic approach, which implies that an aggregative time series based on Business Test data is constructed (e.g., for the production of shoe manufacturers or for the selling prices of shoe wholesalers), after which this series is compared with a time series based on corresponding numerical data. Anderson [1, 2] was the first to apply this idea; he found a rather satisfactory similarity. In a recent study Jochems and De Wit [10] arrived at similar encouraging conclusions. No doubt, these results are of considerable importance; in the first place, because

[5] In France the "Institut National de la Statistique et des Etudes Economiques" (I.N.S.E.E.) has developed independently a similar Business Tendency Survey, which is held each half year but differs in some aspects substantially from the Ifo-Survey. For a detailed description of these differences, see Stutz [14].

reliable Business Test data may facilitate the task of business forecasters, as these data are published in most cases one or two months earlier than traditional statistics. In the second place, because Business Test data cover many more variables than familiar statistics do, such as plans and expectations, so that the inclusion of Business Test time series in regressions of "conventional" variables can be a new and important possibility for widening our horizon. On the other hand, it is not at all self-evident that B.T. results are indeed comparable with statistical data, because it is conceivable that entrepreneurs, when answering questions in the survey, have things in mind which are different from those covered by the definitions of traditional statistics.

By the alternative, micro-economic, approach it is possible to shed some light on the problem of comparability at the level of the individual firm. For instance, we can observe to what extent individual firms reporting an increase, say, have actually experienced an increase during that time period according to the numerical statistical data. This micro-economic approach is the primary object of this study and of an earlier report of the Econometric Institute [8], which dealt with the so-called indifference interval of entrepreneurs. The possibility of such a micro-economic study is in general hampered by the fact that it requires monthly statistical data for individual firms. These data are collected by official statistical bureaux, which are unable to disclose data on individual firms. In the Netherlands, fortunately, the B.T. Survey too is conducted by the Central Bureau of Statistics, so that this Bureau was in a position to supply us with corresponding numerical and survey data and still protects its secrecy about its informants[6].

A description of the available material is given in Section 2 below, after which two approaches for establishing the relationship between survey results and traditional figures are considered, viz., in Section 3 (based on indifference intervals), and in Sections 4 and 5 (based on so-called response functions). Section 6 deals with an alternative interpretation of the concept of corresponding traditional statistics, and the conclusions follow in Section 7.

2. Description of the Material

The analysis is based upon B.T. data supplied by 40 firms in the Dutch Shoe Industry for two variables: daily production and stocks of unsold ready-made shoes. All firms supplied information on the former variable, the non-response

[6] The authors are greatly indebted to Mr. H.G.C. Nanninga, Mr. K.W. van Vugt and Mr. A.A. Bouwman, division chief and staff members of the Central Bureau of Statistics, for their cooperation in having the material for this study arranged and for their kind readiness to supply additional information.

rate being zero. Data on stocks of unsold ready-made footwear are only available for 21 of these firms, viz., those which maintain relatively sizable stocks because they produce primarily in anticipation of demand; the non-response rate for this variable being 5 per cent, which is due to 6 of the 21 firms reporting irregularly. The period covered is the 35-month period from February 1954 until December 1956. An over-all picture of the Test reports given is presented in Table 1; the large fraction of no-change reports for production is striking.

Table 1: The distribution of test reports for production and stocks of the Dutch Shoe Industry, 1954 - 1956

Variable	Total number of reports	Relative frequencies: (+)	(=)	(−)
Daily production	1400	.17	.76	.07
Stocks of unsold ready-made shoes	701	.15	.36	.49

The corresponding numerical variables are daily production [7] and stocks of unsold ready-made shoes, at the end of the month [8]; both are obtained by adding pairs of shoes. In order to conceal the identity of the firms, the numerical data are of the form of percentage changes [9] from the preceding month.

3. Indifference Intervals

It seems self-evident that, in general, no entrepreneur will regard a small positive change (e.g., a production increase of 1 per cent) as an increase. Instead, he will consider it as a case of no change, and the same applies to small decreases. The simplest way to take account of this is to distinguish between three intervals in the distribution of numerical changes, viz., one interval of the larger positive changes which are indeed reported as increases, one interval of the larger negative changes, reported as decreases, and one interval of the changes in between, reported as cases of no change. This third interval is known as the indifference interval, a concept which has been introduced by Theil and Cramer [20] and which has been the object of two unpublished studies,

[7] Daily production is derived by dividing monthly production by the number of working days. This derivation could not take account of the lost working days due to holidays and special occasions. Particularly the figures for daily production in July and August are seriously affected by these disturbances.

[8] The numerical data on stocks are supplied by some firms at the end of four- or five-week periods. These data have been adjusted to a monthly basis by graphical interpolation under the assumption of linearity of inventory fluctuations.

[9] Here as well in the remainder of this report the terms "changes", "relative changes", and "percentage changes" have identical meaning.

one by Lönnqvist [13] referring to the Swedish Textile and Clothing Industry and the Swedish Engineering Industry, and the other by Jochems [8] for the Dutch Shoe Industry. The paper of Lönnqvist deals i.a. with a special survey of the Konjunkturinstitutet among entrepreneurs asking them to state the smallest percentage change which makes them report (+) or (−) instead of (=), values which can be regarded as the limits of the indifference interval. It was found that the width of these intervals for the different variables was varying from about 6 per cent in the case of selling prices to more than 11 per cent for investment in machinery, while also the different branches of industries and large and small firms differed in this respect. It seems plausible that this variability is positively related to the degree to which the variables themselves are fluctuating: variables which are normally subject to minor quarter-to-quarter changes (like selling prices) are characterised by small indifference intervals, and those which are subject to sizable fluctuations (like inventories of small firms) by large intervals.

The second paper is not based on a special survey, but, instead, on the data which are the object of the present paper. The indifference interval for each month was estimated on the basis of the assumption that the firms which report increases correspond with the largest numerical increases according to the statistical data and, similarly, that the firms reporting decreases correspond to the largest numerical decreases. This leads to indifference intervals which are three or four times larger than those of the Swedish entrepreneurs. This difference can be partly explained by the fact that the Swedish surveys refer to changes over longer periods (quarters) than the Dutch surveys, so that the monthly changes in production and inventories will presumably be much larger on the average than the corresponding Swedish changes. It is possible that also the formulation of the Swedish questionnaire is a partial explanation of this difference; for if it is asked "What is the smallest change which you report as (+) or (−) ?", and if this is interpreted as "What is the smallest change which you ever reported as (+) or (−) ?", then there may easily be a bias in downward direction.

4. Response Functions

There can be no doubt that the above-mentioned procedure for the estimation of indifference intervals is based on an unrealistic assumption. It is simply not true that the firms reporting increases were actually confronted with the largest numerical increases according to the corresponding traditional data. As a matter of fact, when we analyse the correspondence between Business Test data and traditional statistical data, we should expect sizable deviations (for several reasons, to be explained below), in precisely the same way as we obtain low correlations in cross-section analyses of the consumption pattern of individual consumers and the investment pattern of individual firms.

A much less restrictive theory is that of the response functions. This approach, due to Pfanzagl [14] and further developed by Theil [21], introduces the probability p that an entrepreneur reports an increase when he is confronted with a certain change δ, and a probability q that he reports a decrease in that case. It is assumed that p and $-q$ are monotonically nondecreasing functions of δ between 0 and +1 and between −1 and 0, respectively; further, that these functions are identical for all firms and all periods.

An attempt to estimate p and q as a function of δ is made in Table 2. This table contains, for various intervals of δ [10], the relative frequency of the cases in which the firms experiencing certain numerical changes report an increase or a decrease. For production a distinction is made between 14 larger and 26 smaller firms [11]; for stocks this is impossible because of the small number of cases for which relative frequencies would have to be derived. An inspection of the table shows that its only encouraging feature is that the relative frequencies of (+) reports tend to increase from left to right and that those of (−) reports tend to decrease; apart, of course, from certain irregularities, which are at least partly due to small numbers. In no case does a response function approximate the unit value for substantial numerical changes in either direction, while large numerical decreases (increases) are sometimes, though not very frequently, reported as increases (decreases).

The (−) reports for stocks occur with rather large frequencies, especially when the numerical change is negative, but also when it is positive. Clearly, there is a tendency of the firms to report inventory changes on the negative side.

Various experiments have been made in order to improve these results for production, but they did not result in any substantial improvement [12]. On the other hand, a series of Wilcoxon's tests applied in all 35 months to the numerical changes corresponding to the three types of reports revealed that the changes belonging to a reported increase tend to be stochastically larger than those belonging to reported no change, and that the latter changes tend to be stochasti-

[10] The intervals of numerical changes (δ) in Table 2 and some of the later tables of this paper are based upon the logarithmic transformation $\log\left(1+\frac{\delta}{100}\right)$ in order to eliminate the effect of the asymmetric character of the distribution of percentage changes; cf. Kaptein and Jochems [12] and Kaptein [11].

[11] For other purposes weights have been given to the various firms, such that 26 firms have unity weight and 14 firms have weights ranging from 2 to 7. This enables us to make this distinction between small and large firms.

[12] First, the months July, August and September were omitted (because of the holiday-effect; cf. footnote 7). Second, seven firms were omitted which reported in all months no change in their production. Neither of these eliminations (combined or separately) led to p- or q-values of 0.50 or larger.

Table 2: Relative frequencies of (+) and (−) reports for various intervals of numerical changes in production and stocks of the Dutch Shoe Industry, 1954-1956

(+) Reports

	all δ	δ: ≤ -30	$-29/-10$	$-9/0$	$+1/+5$	$+6/+10$	$+11/+20$	$+21/+40$	$+41/+80$	$\geq +81$
Produktion:										
All firms	.17	.09	.14	.12	.17	.22	.21	.28	.22	.27
Large firms	.14	.11	.10	.10	.14	.18	.20	.17	.25	.50
Small firms	.19	.09	.16	.14	.20	.25	.22	.33	.22	.19
Stocks:										
All firms	.15	.00	.03	.09	.14	.16	.33	.34	.32	.34

(−) Reports

	all δ	δ: ≤ -46	$-45/-30$	$-29/-18$	$-17/-10$	$-9/-6$	$-5/-1$	$0/+10$	$+11/+40$	$\geq +41$
Produktion:										
All firms	.07	.35	.13	.11	.16	.13	.04	.03	.04	.04
Large firms	.06	.25	.14	.12	.10	.05	.03	.04	.03	.17
Small firms	.08	.37	.12	.10	.20	.18	.04	.02	.05	.06
Stocks:										
All firms	.49	.64	.67	.65	.70	.47	.47	.41	.35	.34

cally larger than those which belong to a reported decrease [13]. Thus, a stochastic relationship between Business Test data and conventional statistics seems well-established; but it is highly disturbed.

Table 3: Relative frequencies of (+) and (−) reports on production and stocks for various intervals of numerical changes, δ, in separate months: Dutch Shoe Industry, 1954-1956

Month	all δ	A	B	C	D	all δ	A	B	C	D
					Production					
	(+) Reports					(−) Reports				
J	.16	.04	.31	.12	.23	.04	.13	.00	.00	.00
F	.21	.25	.12	.20	.29	.04	.08	.00	.00	.06
M	.17	.08	.08	.15	.34	.06	.17	.12	.02	.03
A	.09	.07	.06	.11	.21	.19	.39	.24	.05	.07
M	.08	.08	.05	.06	.17	.15	.31	.13	.00	.22
J	.11	.11	.08	.12	.16	.15	.28	.11	.08	.11
J	.17	.16	.22	.20	.00	.07	.07	.00	.20	.14
A	.28	.24	.11	.17	.34	.01	.00	.00	.00	.01
S	.26	.10	.19	.29	.28	.01	.00	.00	.04	.00
O	.23	.09	.19	.42	.24	.04	.06	.03	.03	.05
N	.17	.22	.15	.23	.08	.04	.13	.00	.00	.04
D	.12	.10	.15	.20	.03	.07	.12	.07	.05	.03
Total	.17	.13	.13	.19	.24	.07	.15	.08	.03	.05
					Stocks					
	(+) Reports					(−) Reports				
J	.27	.00	.00	.33	.53	.35	.45	.25	.17	.40
F	.23	.00	.07	.33	.47	.38	.69	.21	.25	.32
M	.26	.10	.23	.40	.55	.35	.48	.31	.20	.18
A	.08	.00	.17	.00	.31	.53	.69	.50	.44	.31
M	.05	.00	.00	.00	.18	.66	.81	.81	.57	.41
J	.16	.00	.17	.14	.21	.46	.62	.50	.57	.38
J	.20	.05	.00	.20	.36	.54	.68	.50	.40	.48
A	.02	.00	.10	.00	.00	.60	.76	.30	.40	.43
S	.07	.00	.00	.30	.09	.56	.73	.30	.50	.36
O	.02	.00	.00	.00	.17	.64	.63	.73	.44	.83
N	.12	.00	.00	.00	.26	.53	.71	.60	.62	.37
D	.40	.00	.00	.00	.47	.22	.50	.50	.33	.17
Total	.15	.02	.07	.15	.33	.49	.67	.47	.43	.34

[13] Cf. Jochems and Neudecker [9].

5. Response Functions for Separate Months

In the preceding section we dealt with the response functions for all data taken together, in this section we will consider monthly response functions, which may shed some light on possible differences over time. In order to have a sufficient number of observations of (+) and (−) reports, we have aggregated the data of corresponding months, assuming that the seasonal pattern, if any, is not changing over time. For the same reason we have computed the relative frequencies for four intervals only, viz., interval $A: \delta \leqslant -10$, $B: -9 \leqslant \delta < 0$, $C: 0 < \delta \leqslant +10$, and $D: \delta \geqslant +11$, each interval containing approximately one quarter of all numerical observations.

Table 3 shows that, in the case of stocks, the p-values are in general largest in D and smallest in A, the B- and C-frequencies being usually in between, and mutatis mutandis for the response functions q of (−) reports, which is as we could expect. For production, however, the picture is more irregular, in spite of the fact that the number of underlying data is about twice as large.

It is of course conceivable that entrepreneurs adjust their reports to the normal seasonal fluctuations in their variables in the sense that they report an increase or a decrease only when the numerical figures are sufficiently larger or smaller than the "normal" change according to the seasonal pattern. In order to eliminate the seasonal disturbances on the monthly response functions for production, we considered an alternative distribution of the numerical changes into the four quartiles instead of into four fixed intervals. This led to some minor improvements in certain q-values; in general, however, the improvements and deteriorations were balanced.

6. An Alternative Interpretation of Corresponding Traditional Statistics on Production

Up till now, our considerations have been based on the assumption that the survey data and the numerical data refer essentially to the same variables. In this section we shall make certain alternative assumptions. We shall confine ourselves to production, for which the results of the preceding sections were somewhat less promising than for stocks; apart, of course, from the negative bias of the inventory reports.

Production as measured by the numerical data is simply output. Now it is conceivable that the entrepreneur, when answering the questions of the Business Test questionnaire, is more inclined to report on "productive activity" than on output. Let us then assume that this productive activity – being the variable on which the entrepreneur really reports – in month t can be approximated by the mean of output in t and $(t+1)$:

(6.1) $$P_t = \frac{1}{2}(x_t + x_{t+1}),$$

x_t being the output in t. This seems reasonable because the activity in t may result in output that is partly measured in that month and partly in month $(t+1)$.

First, the changes in productive activity in t as compared with $(t-1)$ have been derived by comparing the activity in successive months:

(6.2) $$\eta_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\frac{1}{2}(x_t + x_{t+1}) - \frac{1}{2}(x_{t-1} + x_t)}{\frac{1}{2}(x_{t-1} + x_t)} = \frac{\frac{x_{t+1}}{x_{t-1}} - 1}{\frac{x_t}{x_{t-1}} + 1}.$$

In the second place we have considered the possibility that the entrepreneur, when reporting on production, compares the present activity in his firm with the productive activity in a time period somewhat longer ago, viz., he takes the activity of two months ago as a basis for comparison. The change in productive activity in such "alternating" months is then defined as:

(6.3) $$\zeta_t = \frac{P_t - P_{t-2}}{P_{t-2}} = \frac{\frac{1}{2}(x_t + x_{t+1}) - \frac{1}{2}(x_{t-2} + x_{t-1})}{\frac{1}{2}(x_{t-2} + x_{t-1})} = \frac{\frac{x_{t+1}}{x_{t-2}} + \frac{x_t}{x_{t-2}}}{\frac{x_{t-1}}{x_{t-2}} + 1} - 1.$$

Some results are shown in Table 4, the arrangement of which is identical with that of Table 2. Similarly, Table 5 contains the η- and ζ-results for separate months which correspond with the δ-results of Table 3. A quick glance at these tables suggests that there is not much difference compared with the findings of the preceding sections[14]. A closer inspection reveals a slightly more substantial difference, especially for the (+) reports. Upon comparing the Tables 3 and 5, we find that, for the material as a whole, the relative frequency of (+) reports

14) The correspondences between the two types of data seem to be the least satisfactory in all tables in the months June until August and November until January. It is interesting to mention in this connection that in these periods new orders are received and production is reinstalled for summer or winter production. The change-over from positive to negative seasonal deviations and reversely takes also place in these periods, as can be seen from the following seasonal pattern (average seasonal deviations), derived from the index numbers for production in the Dutch Shoe Industry [15]: J, 0; F, +9; M, +39; A, +12; M, +8; J, +9; J, −29; A, −31; S, −4; O, −5; N, −8; D, 0. These substantial seasonal fluctuations may be caused partly by the fact that some entrepreneurs disclose numerical data of a mixed nature to the statistical bureaux, mixed in the sense that the production index may reflect in some months the level of sales as well as of production.

Table 4: Relative frequencies of (+) and (−) reports on production for various intervals of numerical changes δ, η and ζ: Dutch Shoe Industry, 1954-1956

(+) Reports

Response functions	all	δ, η, ζ: $\leqslant -30$	$-29/-10$	$-9/0$	$+1/+5$	$+6/+10$	$+11/+20$	$+21/+40$	$+41/+80$	$\geqslant +81$
$p(\delta)$	.17	.09	.14	.12	.17	.22	.21	.28	.22	.27
$p(\eta)$	.17	.00	.09	.12	.22	.22	.27	.25	.21	.25
$p(\zeta)$	.17	.08	.08	.14	.18	.21	.24	.25	.37	.29

(−) Reports

Response functions	all	δ, η, ζ: $\leqslant -46$	$-45/-30$	$-29/-18$	$-17/-10$	$-9/-6$	$-5/-1$	$0/+10$	$+11/+40$	$\geqslant +41$
$q(\delta)$	.07	.35	.13	.11	.16	.13	.04	.03	.04	.04
$q(\eta)$	.07	.29	.44	.21	.08	.16	.05	.04	.05	.06
$q(\zeta)$	.07	.33	.28	.15	.12	.06	.04	.03	.04	.05

for interval D (increases of 11 per cent and larger) is about twice as large as that for A (decreases of 10 per cent and larger) if we take the ordinary numerical changes (δ), but that it is about three times larger if we take η or ζ. The lat-

Table 5: Relative frequencies of (+) and (−) reports on production for various intervals of numerical changes, η and ζ, in separate months: Dutch Shoe Industry, 1954-1956

Month	all η	A	B	C	D	all ζ	A	B	C	D
					(+) Reports					
J	.16	.00	.10	.19	.22	.16	.00	.27	.09	.25
F	.21	.00	.10	.20	.34	.21	.00	.06	.25	.33
M	.17	.07	.11	.21	.38	.17	.06	.11	.11	.33
A	.09	.10	.02	.13	.22	.09	.00	.15	.11	.13
M	.08	.07	.07	.12	.00	.08	.08	.02	.15	.14
J	.11	.07	.10	.36	.14	.11	.08	.12	.33	.00
J	.17	.10	.10	.38	.18	.17	.13	.25	.29	.30
A	.28	.33	.17	.14	.33	.28	.22	.11	.27	.36
S	.26	.00	.18	.32	.22	.26	.00	.29	.16	.28
O	.23	.05	.19	.42	.22	.23	.00	.15	.29	.32
N	.17	.12	.27	.10	.19	.17	.12	.18	.29	.13
D	.07	.08	.08	.10	.00	.07	.05	.09	.13	.05
Total	.17	.08	.12	.29	.26	.17	.09	.13	.19	.27
					(−) Reports					
J	.04	.00	.10	.03	.00	.04	.11	.00	.04	.00
F	.02	.00	.00	.04	.03	.02	.00	.00	.00	.07
M	.06	.07	.14	.02	.00	.06	.20	.11	.00	.02
A	.19	.45	.23	.05	.22	.19	.45	.22	.05	.00
M	.15	.33	.09	.07	.33	.15	.36	.05	.04	.29
J	.15	.17	.10	.09	.29	.15	.19	.00	.11	.29
J	.07	.07	.03	.12	.06	.07	.07	.08	.00	.20
A	.01	.00	.00	.00	.01	.01	.00	.00	.00	.02
S	.01	.00	.00	.02	.00	.01	.00	.00	.00	.01
O	.04	.05	.05	.00	.11	.04	.08	.00	.03	.07
N	.04	.04	.02	.03	.06	.04	.06	.03	.00	.04
D	.07	.08	.08	.05	.14	.07	.19	.05	.00	.05
Total	.07	.15	.08	.05	.05	.07	.17	.05	.02	.05

ter result is obviously more favourable; it suggests that entrepreneurs are indeed inclined to report on productive activity of the moving-average type rather than on output, at least as far as (+) reports are concerned. No such effect can be detected for (−) reports in this connection. An inspection of the results of the application of a series of Wilcoxon's tests to the numerical changes δ, η and ζ corresponding to the three types of reports reveals[15] that in all cases the changes belonging to reported increase tend to be stochastically larger than those belonging to reported no change, and that the latter tend to be stochastically larger than those which belong to a reported decrease. However, the former is found most frequently for η and ζ and the latter most frequently for δ. Perhaps these observed differences are not unnatural; it would imply that entrepreneurs are immediate reporters of bad events, but that they report a favourable development only when time has shown them that it has a solid basis.

7. Conclusion and Appraisal

(i) There exists a positive relation between Business Test reports and numerical statistical data in the sense that the fraction of (+) reports is an increasing function of the corresponding numerical change and similarly, the fraction of (−) reports is a decreasing function of the numerical change. This applies both to production and to stocks. However, these functions are highly disturbed; in particular, there is no evidence for assuming that the fractions of (+) reports and of (−) tend to unity for sufficiently large numerical increases and decreases, respectively. In principle there are three sources of deviation, viz.:

1. the variable which is measured according to the numerical statistical data differs conceptually from the variable which the entrepreneur has in mind when answering the questions of the survey[16];

2. the survey questions are answered in a careless manner;

3. the numerical statistical information supplied by the firm to the statistical bureau is inaccurate.

(ii) As to the first possibility, there may be in some cases a difference between stocks according to the numerical data and stocks according to the survey in that sold stocks are covered as well in one or the other. It is conceivable that the negative bias of the survey reports compared with the numerical data is due

[15] Cf. Jochems and Neudecker [9].

[16] Cf. footnote 14.

to the feature that they have been accounted for in the numerical data, but no additional information supporting this view is available. As to production, it seems plausible that (+) reports refer to a moving average of monthly output rather than to output itself. This statement does not apply to (−) reports.

(iii) The contribution of inaccurate numerical data to the deviations observed should not be underestimated. In particular, there is the number of vacation days and the fact that some firms have a four-week accounting period[17]. Although a correction was made whenever this was known, this correction is not necessarily perfect; and it was not always known.

References

[1] Anderson jr., O.: "The Business Test of the IFO-Institute for Economic Research, Munich, and its Theoretical Model", Review of the International Statistical Institute, Vol. 20 (1952), pp. 1-17.

[2] Anderson jr., O., Bauer, R. K. and Giehl, R.: "Zur Theorie des Konjunkturtestes", IFO-Studien, Vol. 1 (1955), pp. 159-204.

[3] Anderson jr., O., Bauer, R. K., Führer, H., Petersen, J. P. and Wolfsteiner, M.: "Short-Term Entrepreneurial Reaction Patterns", Paper presented at the 17th European Meeting of the Econometric Society, Kiel, 1955.

[4] Anderson jr., O., Fürst, H. and Schulte, W.: "Zur Analyse der unternehmerischen Reaktionsweise", IFO-Studien, Vol. 2 (1956), pp. 129-154.

[5] Anderson jr., O., Bauer, R. K. and Fels, E.: "Zur Treffsicherheit kurzfristiger Unternehmererwartungen", IFO-Institute, Munich, 1954.

[17] Cf. footnote 8.

[6] Anderson jr., O.: "Das Konjunkturtest-Verfahren und sein Beitrag zur Empirischen Untersuchung der ex ante – ex post – Problematik", München 1957.

[7] Hastay, M.: "The Role of Business Expectations in the Operating Behavior of the Firm", Paper presented at the Conference of the Quality and Economic Significance of Anticipations Data, Princeton, 1957.

[8] Jochems, D.B.: "A Micro-Analysis of the Indifference Intervals of Entrepreneurs Participating in Business Tendency Surveys", Report 5815 of the Econometric Institute (1958).

[9] Jochems, D.B. and Neudecker, H.: "An Application of Wilcoxon's Test to Numerical and Survey Data of the Dutch Shoe Industry", Report 5822 of the Econometric Institute (1958).

[10] Jochems, D.B. and de Wit, G.M.: "The Macro-economic Relationship between Business Tendency Surveys and Numerical Data", Report 5904 of the Econometric Institute (1959).

[11] Kaptein, E.: "Fluctuations in Inventories of Dutch Shoe Factories, 1954-1956", Report 5814 of the Econometric Institute (1958).

[12] Kaptein, E. and Jochems, D.B.: "The Analysis of Frequency Distributions of Changes in Economic Microvariables Applied to the Dutch Shoe Industry, 1954-1956", Report 5808 of the Econometric Institute, Paper presented at the 31th Session of the International Statistical Institute at Brussels, 1958.

[13] Lönnqvist, Åke: "On the Relation Between Ex Ante und Ex Post Data in the Swedish Business Tendency Surveys", Paper presented at the 3rd International Conference on Business Tendency Surveys, Munich, 1957 (forthcoming in IFO-Studien, 4. Jg. (1958): "Beziehungen zwischen ex ante – und ex post-Daten im Swedischen Konjunkturtest.")

[14] Pfanzagl, J.: "Zur Methodik des Konjunkturtest-Verfahrens", Statistische Vierteljahresschrift, Vol. 52 (1952), pp. 161-173.

[15] "Statistiek voor de Nijverheid", 1952-1956, Table 2, Centraal Bureau voor de Statistiek.

[16] Stutz, F.N.: Der Konjunkturtest. Neue Wege der Konjunkturdiagnose und -prognose, Zürich and St. Gallen 1957.

[17] Theil, H.: "Recent Experiences with the Munich Business Test", Econometrica, Vol. 23 (1955), pp. 184-192.

[18] Theil, H.: "On the Time Shape of Economic Microvariables and the Munich Business Test", Review of the International Statistical Institute, Vol. 20 (1952), pp. 105-120.

[19] Theil, H.: "Measuring the Accuracy of Entrepreneurial Anticipations", Paper presented at the 17th European Meeting of the Econometric Society, Kiel, 1955.

[20] Theil, H. and Cramer, J.S.: "On the Utilization of a New Source of Economic Information: An Econometric Analysis of the Munich Business Test", Paper presented at the 16th European Meeting of the Econometric Society, Uppsala, 1954.

[21] Theil, H.: Economic Forecasts and Policy, Amsterdam 1958.

[22] United Nations: "Short-Term Business Indicators in Western Europe", Economic Bulletin for Europe, Vol. 7 (1955), No. 3, pp. 34-78.

REVIEW OF ECONOMIC STUDIES, Vol. XXV, No. 2 (1958)

Substitution, Complementarity, and the Residual Variation around Engel Curves

1. Introduction

It is well-known that the empirical determination of substitution and complementarity relationships among pairs of consumption goods is hampered by serious difficulties. The time series analysis which is necessary for the estimation of the underlying price and income elasticities is characterised by the problems of small samples, multicollinearity among the explanatory variables, etc., so that rather large standard errors of the resulting estimates is a normal result. It is understandable, therefore, that alternative procedures have been devised in order to determine substitution and complementarity relations. The present paper is concerned with one of these procedures, which is based on budget data. The number of observations in budget enquiries is usually very much larger than in the case of time series analysis, but there is of course the disadvantage that no adequate price variation among the observational units exists.[1] The procedure followed implies then that, for any pair of commodities for which one wants to establish a substitution or complementarity relationship, the two Engel curves are derived, after which the two sets of residuals around these curves are correlated with each other. A positive residual correlation is then regarded as a *prima facie* indication of complementarity, a negative correlation as an indication of a substitution relationship. Allen and Bowley [1] (to whom, as far as the present authors are aware, this method is due) computed thirty-eight residual correlations for pairs of expenditure groups and arrived at figures ranging from—0·94 (for butter and margarine) to 0·73 (for fresh fruit and vegetables), most of them being negative. Following their lead, one of the authors [5] computed eighteen residual correlations belonging to an Engel curve for the total quantity bought of a particular commodity group and to the corresponding curve for the average price paid for this same group.[2] These correlations, which are concerned with substitution relationships between " quality " and " quantity " of single commodity groups, turned out to be all negative. This dominance of negative values is in accordance with the theoretically established dominance of substitution over complementarity relationships.

It seems obvious that this approach is of interest for the empirical determination of substitution relationships, provided it is indeed true that the residual correlations measure what they are supposed to measure. For the only thing which one has to do is to divide the consumer's budget into an appropriate number (say n) of categories, after which the

[1] There is of course price variation in the sense that different families pay different prices for the " same " commodity because of differences in quality ; a reference to this phenomenon is made in a later part of this section. This kind of price variation is not suitable for the determination of price elasticities.

[2] Allen's and Bowley's analysis is confined to food items, and the explanatory variable in the Engel curves is total food consumption per equivalent adult (not income). The dependent variables of their regressions are the expenditures per equivalent adult. In the quantity-quality analyses no division took place by the number of equivalent adults, but this number was introduced as additional explanatory variable besides income.

$\binom{n}{2}$ residual correlations indicate, for each pair of categories, whether there is substitution or complementarity, and perhaps also the strength of this relationship. However, the proviso mentioned above is not of a simple kind, as Allen and Bowley observed already in their work. Later research by Prais [3] did not result in particularly hopeful suggestions.[1] Since it is unclear at present what is really measured by the residual correlations, it seems worth-while to explore this matter in more detail. This is the primary purpose of this paper ; a second purpose is that of the economic interpretation of the " disturbances " of Engel curves—a subject which is largely, but undeservedly, neglected in the analysis of these and most other econometric relations.

The main result of the analysis which follows is that, on the one hand, it is possible (under certain conditions) to express the residual variances and covariances in terms of the classical substitution terms of consumer's analysis, whereas, on the other hand, it appears that the covariance for one pair of commodities is not solely determined by the substitution term of this pair, but by substitution terms of other pairs as well. This is explained in the next six sections, after which the results are summarised in more detail in Section 8. This concluding section contains also a suggestion on a method of deriving substitution terms from the matrix of all residual variances and covariances.

2. Substitution and Complementarity Re-Stated

We shall base ourselves on the classical theory of consumer's demand, for which reference can be made to, e.g., Hicks [2]. The present section is devoted to the definition of symbols.

First, we have the utility function

$$u = u(x_1, \ldots, x_n) = u(x), \tag{2.1}$$

$x_1, x_2, \ldots$ being the quantities bought of the various commodities (or x as a column vector). This function is to be maximised subject to the budget constraint :

$$\sum_i p_i x_i = M, \tag{2.2}$$

the p's being prices and M income.

Let us write :

$$u_{ij} = \frac{\partial^2 u}{\partial x_i \, \partial x_j} \tag{2.3}$$

for the derivatives of the utility function of the second order. An essential role in the second-order maximum condition is played by the matrix of these derivatives, bordered by (negative) prices :

$$U = \begin{bmatrix} u_{11} & \ldots & u_{1n} & -p_1 \\ \cdot & \cdot & \cdot & \cdot \\ u_{n1} & \ldots & u_{nn} & -p_n \\ -p_1 & \ldots & -p_n & 0 \end{bmatrix} \tag{2.4}$$

[1] Also, cf. Prais and Houthakker (4), p. 102. Prais did not compute residual correlations, but he introduced the consumption of the competing commodity as an additional explanatory variable in the Engel regression. Essentially, this procedure is not different from computing residual correlations.

The inverse of this matrix is written as

$$(2.5) \qquad U^{-1} = \begin{bmatrix} c_{11} & \cdots & c_{1n} & c_1 \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & \cdots & c_{nn} & c_n \\ c_1 & \cdots & c_n & c_0 \end{bmatrix}$$

or in partitioned form :

$$(2.6) \qquad U^{-1} = \begin{bmatrix} C & c \\ c' & c_0 \end{bmatrix},$$

C being a symmetric and negative semi-definite matrix (its rank being $n - 1$), c a column vector, c' its transpose, and c_0 a scalar. The matrix C is called the *substitution matrix*. Two commodities i and j are said to be each other's substitutes from the point of view of the consumer if their *substitution term* $C_{ij} > 0$; complementary if $C_{ij} < 0$.

3. The Relevance of Quadratic Utility Assumptions

Our starting point will be the assumption of a quadratic utility function :

Assumption I. *The consumer's utility function* (2.1) *has the specification*

$$(3.1) \qquad u(x) = a'x + \tfrac{1}{2}x'Bx = \\ = \Sigma\, a_i x_i + \tfrac{1}{2} \Sigma \Sigma\, b_{ij} x_i x_j,$$

the a's and b's being coefficients which are independent of the quantities bought.

The relevance of this assumption for our present purposes is easily seen. For suppose that the utility function is not quadratic. In that case the second-order derivatives of u are no longer constants, as they are in the quadratic case ($u_{ij} = b_{ij}$). Since the substitution terms are derived by inverting the bordered matrix of these derivatives, these terms are then dependent on the quantities bought and hence on prices and income. The dependence on prices is not serious if we make the conventional (though not necessarily realistic) assumption that they are constant during the budget period and the same for all consumers of the budget survey. The dependence on income is more serious, however. The point is not so much that the income of a particular family may change during the budget survey ; but it is that, if the utility function is not quadratic, two commodities may be substitutes or complementary depending on income, so that for two families with different income in the same budget survey the question of substitution or complementarity for this pair of goods may be answered differently. Clearly, a residual correlation cannot take account of such divergencies, because this is a quantity which is derived for a budget survey as a whole. To give the residual correlation as a measure of substitution or complementarity any chance at all, we should exclude such possibilities by assuming that the matrix B is the same for all families of the survey, so that they all have identical substitution matrices.[1]

Assumption II. *Assumption I applies to all families of the budget survey ; and their matrices B, as well as the prices, are identical for all of them.*

[1] If one prefers expressions that are invariant against monotonic transformations of the utility function, we should say that the B-matrices are identical except possibly for a multiplicative scalar.

This assumption implies that the preferences of all consumers of the survey are supposed to be *perfectly homogeneous* as regards the coefficients of the quadratic terms of their preference functions. We shall not suppose that this also applies to the coefficients of the linear terms ; if we would do so, the Engel curves would have no residual variation at all.

4. The Engel Curves Under the Assumptions of Section 3

The first-order maximum conditions of consumer's equilibrium are derived by differentiating

$$a'x + \tfrac{1}{2}x'Bx - \lambda(p'x - M),$$

λ being a Lagrange multiplier and p the price vector. The result is

$$Bx - p\lambda = -a, \tag{4.1}$$

which, when combined with the budget constraint,

$$-p'x = -M \tag{4.2}$$

gives the following solution for x and λ :

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = U^{-1} \begin{bmatrix} -a \\ -M \end{bmatrix} \tag{4.3}$$

Using (2.5), we find :

$$x_i = -\sum_h c_{ih}a_h - c_iM, \tag{4.4}$$

or in matrix form :

$$x = -Ca - cM. \tag{4.5}$$

The relations (4.4) and (4.5) can be regarded as Engel curves. The coefficients c_i and c_{ih} are the same for all families, as follows from Assumption II, so that the slopes of these curves ($-c_i$) are identical for all of them. The constant terms, $-\Sigma c_{ih}a_h$, are not the same, because the a's are assumed to differ from family to family. Let us then write α_h for the mean value of the a_h's of all families in the survey and write (4.4) as :

$$x_i = -\sum_h c_{ih}\alpha_h - c_iM - \sum_h c_{ih}(a_h - \alpha_h). \tag{4.6}$$

The first two terms in the right-hand side are the same for all consumers (except of course for the values taken by the variable M) ; and the third, viz. :

$$v_i = -\sum_h c_{ih}(a_h - \alpha_h), \tag{4.7}$$

can be regarded as a disturbance. It is seen that this disturbance contains two types of factors, viz., substitution terms and the deviations of the linear preference coefficients from their average. The occurrence of substitution terms points indeed to a relationship between the residual variation around Engel curves and substitution and complementarity phenomena.

Suppose then that the following assumption is satisfied :

Assumption IIIa. *For each family participating in the budget survey, the deviation of its linear preference coefficient* a_h *from the group average* α_h *can be regarded as a random drawing from a universe with zero mean and a finite covariance matrix which is common to all families ; and no correlation exists between this deviation and income over the set of all families, for* $h = 1, \ldots, n$.

If this assumption is indeed satisfied, then the disturbances v_i of the Engel curves are subject to a probability distribution, and they are uncorrelated with income. The latter condition will be weakened in Section 7. The first part of Assumption IIIa implies that the families are supposed to be *stochastically homogeneous* with respect to their linear preference coefficients.

5. The Residual Variation in the Case of Uncorrelated Linear Preference Coefficients

Let us write :

$$\text{var } v_i = Ev_i^2 \tag{5.1}$$

for the variance of the disturbances of the Engel curve for x_i over the set of families participating in the survey. The mean value of v_i is of course zero because of Assumption IIIa.

Applying (4.7), we find :

$$\text{var } v_i = \sum_h \sum_k c_{ih}c_{ik} \text{ cov } (a_h, a_k), \tag{5.2}$$

where

$$\text{cov } (a_h, a_k) = E[(a_h - \alpha_h)(a_k - \alpha_k)]$$

is the covariance of the linear preference coefficients a_h and a_k, which is, according to assumption IIIa, the same for all families. To make further progress, it is necessary to make an assumption concerning these covariances. The simplest is :

Assumption IVa. *The linear preference coefficients $a_1, \ldots, a_n$ are all uncorrelated*:

$$\text{cov}(a_h, a_k) = 0 \quad \text{if } h \neq k. \tag{5.3}$$

In that case (5.2) reduces to :

$$\text{var } v_i = \sum_h c_{ih}^2 \sigma_h^2 \tag{5.4}$$

where

$$\sigma_h^2 = \text{var } a_h. \tag{5.5}$$

Similarly, we find for the covariance of the disturbances v_i and v_j corresponding to the Engel curves for x_i and x_j :

$$\text{cov } (v_i, v_j) = \sum_h \sum_k c_{ih}c_{jk} \text{ cov } (a_h, a_k) \tag{5.6}$$

which reduces under Assumption IVa to :

$$\text{cov } (v_i, v_j) = \sum_h c_{ih}c_{jh}\sigma_h^2. \tag{5.7}$$

The residual correlation is then found by dividing (5.7) by the geometric average of (5.4) and the similar expression for var v_j. If our interest is restricted to the qualitative distinctions of positive and negative correlation and of substitution and complementarity, we can confine ourselves to the covariance (5.7). It is seen that the sign of this covariance is not determined by the substitution term c_{ij} (which ought to be the case if the conjecture of the first paragraph of Section 1 were correct), but that it is determined by all substitution terms in which one of the commodities i and j enters. More precisely, this covariance is obtained by considering all commodities $h = 1, \ldots, n$ of the entire budget, multiplying the substitution terms of i and j with such a commodity with each other, weighting the product with the variance σ_h^2, and adding. In other words, a large positive covarianc indicates that the two commodities i and j " agree " with each other by being both substitutes of or both complementary with most of the commodities of the budget ; and this

does not necessarily imply that i and j are complementary. Conversely, a large negative covariance indicates that they disagree in the sense that, if one of them is a substitute of h, the other is usually complementary with h ; which does not necessarily imply that i and j are substitutes. It is to be noted, however, that the two terms corresponding to $h = i$ and j in the summation contribute to the expected sign : both $c_{ii}c_{ij}\sigma_i^2$ and $c_{ij}c_{jj}\sigma_j^2$ have signs opposite to $c_{ij} = c_{ji}$, so that their contribution to the covariance is negative when i and j are substitutes, positive when they are complementary. Furthermore, it will be noted that the weighting by σ_h^2 implies that those substitution terms c_{ih}, c_{jh} are heavily weighted which belong to a commodity h for which the linear preference coefficients a_t have large variance.

The following alternative interpretation is also instructive. Suppose we write :

$$\sigma_h^2 = 1 \quad \text{for } h = 1, \ldots, n. \tag{5.8}$$

This always possible when suitable units for the quantities of the various commodities are chosen ; in the case of a large variance of certain linear preference coefficients a_h it implies that the corresponding quantities x_h are measured in relatively small units. Combining (5.8) and (5.4), we find :

$$\text{var } v_t = \sum_h c_{ih}^2, \tag{5.9}$$

which is a diagonal element of the product matrix CC' ; or, given the symmetry of the substitution matrix, of C^2. Similarly, combining (5.8) and (5.7) :

$$\text{cov } (v_i, v_j) = \sum_h c_{ih}c_{jh}, \tag{5.10}$$

which is the element in the i-th row and j-th column of C^2.[1] Hence we have found that, given the assumptions made and given an adequate choice of units, the covariance matrix of the disturbances around the n Engel curves equals the square of the substitution matrix. The presumption of the first paragraph of Section 1 was different : from the remarks made there we might have expected that—given an adequate choice of units, etc.—this covariance matrix eqauls $-C$, substitution corresponding to negative and complementarity to positive residual correlations.

6. The Residual Variation in a Case of Correlated Linear Preference Coefficients

It can be maintained that Assumption IVa is not very realistic. For, if a consumer's linear preference coefficient a_h exceeds the group average α_h, this can be regarded as an indication that this consumer values h more than the average consumer does ; and, if h and k are substitutes, it seems likely that he will value k less than the average consumer does. This argument suggests that a_h and a_k are negatively correlated when h and k are substitutes, and positively correlated when they are complementary. In that case we should replace (5.4) by :

$$\text{var } v_t = \sum_h \sum_k c_{ih}c_{ik}\sigma_h\sigma_k\rho_{hk}, \tag{6.1}$$

where the correlation :

$$\rho_{hk} = \frac{\text{cov}(a_h, a_k)}{\sigma_h\sigma_k} \tag{6.2}$$

[1] $x'CCx = y'y > 0$; because C is singular, it is possible to find $x \neq 0$, satisfying $Cx = y = 0$, so that C^2 is positive semi-definite.

does no longer vanish for $h \neq k$, but is positive if h and k are complementary ($c_{hk} < 0$) and negative if they are substitutes ($c_{hk} > 0$). To make further progress, it is necessary to make more specific assumptions of the following kind. The correlation ρ_{hk} is of course confined to the interval (—1, 1) ; similarly, the substitution term c_{hk} is smaller in absolute value than $\sqrt{c_{hh}\ c_{kk}}$, as follows from the semi-definiteness of the substitution matrix. Since it is plausible that ρ_{hk} is an algebraically decreasing function of c_{hk}, it seems worthwhile to analyse the implications of :

ASSUMPTION IVb. *The correlations between the linear preference coefficients, as defined in* (6.2)', *are given by*

$$\rho_{hk} = \frac{-c_{hk}}{\sqrt{c_{hh}c_{kk}}} \tag{6.3}$$

Combining (6.3) and (6.1), we find :

$$\text{var } v_i = \sum_h \sum_k c_{ih}\, c_{ik}\, \sigma_h\, \sigma_k\, \frac{-c_{hk}}{\sqrt{c_{hh}c_{kk}}} \tag{6.4}$$

This expression can be simplified if we write :

$$\sigma_h/\sqrt{(-c_{hh})} = 1 \qquad \text{for } h = 1, \ldots, n, \tag{6.5}$$

which is possible by the choice of adequate units. Then (6.4) reduces to :

$$\text{var } v_i = -\sum_h \sum_k c_{ih}\, c_{ik}\, c_{hk}, \tag{6.6}$$

which is a diagonal element of the matrix $-C^3$. Similarly, we find for the covariance :

$$\text{cov}\,(v_i, v_j) = \sum_h \sum_k c_{ih}\, c_{jk}\, \sigma_h\, \sigma_k\, \frac{-c_{hk}}{\sqrt{c_{hh}\, c_{kk}}} \tag{6.7}$$

or, using (6.5),

$$\text{cov}(v_i, v_j) = -\sum_h \sum_k c_{ih}\, c_{jk}\, c_{hk}, \tag{6.8}$$

which is the (i, j)-th element of $-C^3$. So we have found that, under the present assumptions, the covariance matrix of the disturbances around the n Engel curves is neither $-C$, nor C^2, but $-C^3$.[1]

The economic interpretation of this result is more complicated than that of Section 5. The sign of the covariance is no longer determined by the degree to which i and j agree or disagree by being both substitutes of or complementary with all commodities of the budget. Instead, we should consider the substitution terms in which i enters together with all commodities h, and also the terms in which j enters with all k. In Section 5 h and k were identical commodities, but this is no longer the case. The substitution terms c_{ih} and c_{jk} are linked together by the third term c_{hk}, and the resulting product $c_{ih}\, c_{jk}\, c_{hk}$ is to be summed over all h and k. What we can say is this : if, for those pairs of commodities h, k which are substitutes, i and j are both usually complementary with or substitutes of h and k, respectively ; and if, for those pairs h and k which are complementary, i is usually a substitute of h and j complementary with k or *vice versa* ; then the covariance of v_i and v_j tends to be negative. A similar statement can be given for the case of a positive covariance. Comparing this with the result of Section 5, we must say that the distance between this covariance and the simple substitution term c_{ij} has become larger.

[1] $-C^3$ is positive semi-definite, because it is possible to write $-x'C^3x$ as $-z'Cz$, which is positive semi-definite.

7. The Case of Correlation Between Linear Preference Coefficients and Income

In this section we shall remove the assumption that no correlation exists beteen the linear preference coefficients $a_1, a_2, \ldots$ and income. The simplest alternative assumption is that of linear dependence :

$$a_h - \alpha_h = q_h(M - \bar{M}) + w_h, \tag{7.1}$$

$\bar{M}$ being the average income of the families participating in the budget survey and w_h a disturbance. More precisely ;

Assumption IIIb. *For each family participating in the budget survey, the deviation of its linear preference coefficient a_h from the group average α_h can be described as a stochastic linear function of its income in the bugdet period in accordance with* (7.1), *where q_h is a coefficient independent of the particular family and w_h a random disturbance with zero mean and a finite covariance matrix which is common to all families ; and no correlation exists between w_h and income over the set of all families, for $h = 1, \ldots, n$.*

Applying this to (4.4), we find :

$$x_i = -\sum_h c_{ih}(\alpha_h - q_h\bar{M}) - (c_i + \sum_h c_{ih}q_h)M - \sum_h c_{ih}w_h, \tag{7.2}$$

where $-(c_i + \Sigma c_{ih}q_h)$ is the new slope of the Engel curve for the set of all families, and

$$v_i^* = -\sum_h c_{ih}w_h \tag{7.3}$$

the new disturbance. It is seen that the new slope differs from the value $-c_i$ which describes the rate of change of consumption with respect to income when the income of the individual consumer varies. We can say that the new slope is biased relative to $-c_i$, the bias being due to the substitution terms and the q's in $-\Sigma c_{ih}q_h$.

The variance of the disturbance v_i^* can be derived as follows :

$$\begin{aligned}
\text{var } v_i^* &= \sum_h \sum_k c_{ih}c_{ik}E(w_hw_k) \\
&= \sum_h \sum_k c_{ih}c_{ik}E\Big[\big\{a_h - \alpha_h - q_h(M - \bar{M})\big\} \times \\
&\qquad \big\{a_k - \alpha_k - q_k(M - \bar{M})\big\}\Big] \\
&= \sum_h \sum_k c_{ih}c_{ik}\big\{\operatorname{cov}(a_h, a_k) + q_hq_k \operatorname{var} M - \\
&\qquad q_h \operatorname{cov}(M, a_k) - q_k \operatorname{cov}(M, a_h)\big\} = \\
&= \sum_h \sum_k c_{ih}c_{ik}\big\{\operatorname{cov}(a_h, a_k) - q_hq_k\big\}\operatorname{var} M \ .
\end{aligned} \tag{7.4}$$

because

$$q_h \operatorname{cov}(M, a_h) = q_k \operatorname{cov}(M, a_h) = q_hq_k \operatorname{var} M.$$

In the same way :

$$\operatorname{cov}(v_i^*, v_j^*) = \sum_h \sum_k c_{ih}c_{jk}\big\{\operatorname{cov}(a_h, a_k) - q_hq_k\big\}\operatorname{var} M \ , \tag{7.5}$$

Evidently, if the linear preference coefficients cease to be uncorrelated with income, two effects are of importance for the sign of the residual correlations. First, the variance of the income distribution plays a determining role. Secondly, just because of the presence of the term $-q_hq_k$ var M, it is no longer true that the terms $c_{ih}c_{jk}$ for $h \neq k$ are absent in the case of uncorrelated linear preference coefficients, cov $(a_h, a_k) = 0$. More specific results require additional assumptions about the q's, which will not be pursued here.

8. Conclusion and Summary

The result of the analysis is negative in the sense that no conditions have been found under which the residual correlation of the Engel curves of two commodities, or at least the sign of this correlation, is determined by the substitution term corresponding to this pair of commodities. Indeed, it is plausible that no simple conditions of this kind exist. The result is positive, however, in the sense that the residual correlation has been expressed in the theoretical concepts of consumption analysis ; at least under certain conditions, the most important of which is that the utility function be quadratic, the coefficients of the quadratic terms being the same for all consumers. Under these conditions the residuals around Engel curves can be said to be determined by substitution terms and by the coefficients of the linear part of the utility functions, provided that these coefficients are not correlated with the consumers' incomes. If this proviso is not fulfilled, the following effects occur : first, the slopes of the Engel curves are biased by the presence of substitution terms ; second, the variance of the income distribution plays a role in the residual variances and covariances of the Engel curves ; third, the relatively simple result which occurs in the special case of uncorrelated linear preference coefficients (to be considered below) is then no longer applicable.

If the above proviso is fulfilled, a distinction should be made between the case in which the coefficients of the linear part of the preference functions are correlated over the set of all families participating in the budget survey, and that in which they are not correlated. In the latter case the result is, in qualitative terms the following : the sign of the residual correlation for any pair of Engel curves is not determined by the question whether the two commodities are substitutes or complementary, but by the degree to which these commodities agree by being both substitutes of or complementary with all commodities of the budget, including these two. However, it can be shown that the substitution or complementarity relationship between the two commodities themselves contribute to a correctness of the presumption of Section 1 in the sense that they tend to make the residual correlation negative in the case of substitution, positive in the case of complementarity. Hence, if there is only little substitution or complementarity between either of these commodities and all other commodities, this tendency may be sufficient to determine the sign of the residual correlation.

The results are more complicated when the linear preference coefficients are correlated. For the special case considered in Section 6 it was found that the sign of the residual correlation for two commodities i and j is not uniquely determined by their substitution term c_{ij} (which is the presumption of Section 1), nor even by the terms c_{ih} and c_{jh} in which at least one of the two commodities i and j participate (which is the case of uncorrelated linear preference coefficients), but that a determining role is played by the substitution terms c_{hk} of all pairs of commodities h, k of the entire budget. More precisely, whereas the presumption of Section 1 was that—given a choice of appropriate units for the quantities of the various commodities—the covariance matrix of the disturbances around the set of all Engel curves equals $-C$, C being the substitution matrix, we found that this matrix is C^2 for the uncorrelated case of the preceding paragraph and $-C^3$ for the correlated case considered here.

It is conceivable that this result is used to determine $-C$ itself. A theorem in matrix algebra[1] states that, if V is a real symmetric and positive semi-definite matrix, there is a

[1] See e.g. Zurmühl (6), p. 166, for a detailed treatment.

unique positive semi-definite symmetric matrix W, for each positive integer m, such that :

$$W^m = V.$$

On the basis of this theorem it may be possible to determine $-C$, either starting from C^2 or from $-C^3$. A comparison of the alternative outcomes is obviously interesting, but this is largely an empirical subject which will not be pursued here.

Rotterdam.

H. THEIL
H. NEUDECKER.

REFERENCES

[1] Allen, R. G. D., and Bowley, A. L. : *Family Expenditure*, London, 1935.
[2] Hicks, J. R. : *Value and Capital*, Oxford, second edition, 1946.
[3] Prais, S. J. : "Non-Linear Estimates of the Engel Curves," *The Review of Economic Studies*, Vol. XX, No. 2, 1953, p. 98.
[4] Prais, S. J., and Houthakker, H. S. : *The Analysis of Family Budgets*, Cambridge University Press, 1955.
[5] Theil, H. : "Quantities, Prices and Budget Enquiries," *The Review of Economic Studies*, Vol. XIX, No. 50, 1952-53, p. 146.
[6] Zurmühl, R. : *Matrizen*, Springer-Verlag, Berlin/Göttingen/Heildelberg, 1950.