Discussion Paper No. 109

CANONICAL ANGLES AND COMPUTATIONS IN LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATION

by

Claude Cohen *

November 1974

* Graduate School of Management, Northwestern University

Canonical Angles and Computations in Limited Information Maximum Likelihood Estimation

Claude Cohen *
Northwestern University

Abstract

Limited information maximum likelihood (LIML) estimation is presented with an emphasis on the geometry of subspaces induced by the single equation in the structural system. This vector space approach reveals the structure of the LIML model in its canonical analysis form and Hooper's trace correlation coefficient. Two numerical procedures based on the Cholesky and singular value decompositions are described for an efficient solution of the generalized symmetric eigenvalue problem.

Key words

Cholesky decomposition; Householder decomposition; singular values; generalized inverse; canonical correlation; general symmetric eigenproblem; canonical angles; canonical vectors; dimension of inclination; coefficient of inclination; trace correlation; statistical computations.

* I wish to thank Professor Walter D. Fisher for his helpful comments.

1. INTRODUCTION

In this paper much of what is known about limited information maximum likelihood (LIML) is reworked and extended with a strong emphasis on geometrical thinking and efficient computational procedures. In addition to having pedagogical advantages, this vector space approach reveals the structure of the model in its canonical analysis form and permits a development which is tightly interwoven with the efficient solution of the generalized symmetric eigenvalue problem. All preliminaries are stated in Section 2, in particular Cholesky, Householder and the singular value decompositions. The geometry of LIML and canonical correlations is discussed in Section 3 while in Section 4 two numerical procedures are brought to the attention of the econometrician interested in statistical computing.

2. PRELIMINARIES

2.1 Notation and basic results about projectors and generalized inverses

$\lvert x \rvert$	the Euclidean norm of vector x, i.e. $\sqrt{\sum_i x_i^2}$
$M^{\perp}$	the orthogonal complement of subspace M
A' or A^t	the transpose of matrix A ($A: m \times n$)
$R(A)$	The range space of A
$tr(A)$	the trace of A
I_n	the identity matrix of order n
P_M	the orthogonal projector on the space M, i.e. $P_M^2 = P_M = P_M'$
$P_M^{\perp}$	the orthogonal projector on the space $M^{\perp}$, i.e.

$$P_M^{\perp} = I - P_M$$

A^+ the unique generalized increase (g.i.) of A satisfying the following equations:

(1) $$\begin{cases} AA^+A = A & A^+AA^+ = A^+ \\ (AA^+)' = AA^+ & (A^+A)' = A^+A \end{cases}$$ see, e.g. [16]

and for which we note that

(2) $$A^+A = P_{R(A')} \qquad AA^+ = P_{R(A)}$$

(2′) and if A is of full column rank, $A^+ = (A'A)^{-1}A'$

2.2 Cholesky Decomposition (CD)

The best way to compute the inverse of a symmetric positive definite (s.p.d.) matrix A is by Cholesky decomposition such that A = LL', L lower triangular. See e.g. [17]. The matrix A, in statistical work, is usually a variance-covariance or correlation matrix and if we denote its elements by a_{ij}, we have

$$\ell_{11} = \sqrt{a_{11}} \quad , \quad \ell_{i1} = a_{i1}/\ell_{11} \qquad i = 2,\ldots,p$$

then, for i = 2,...,p

$$\begin{cases} \ell_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1} \ell_{ij}^2} \\ \ell_{ij} = (a_{ij} - \sum_{k=1}^{j-1} \ell_{ik}\ell_{jk})/\ell_{ii} & j = i+1,\ldots,p \end{cases}$$

2.3 Householder's Decomposition (HD)

If X is an n x p matrix, $n \geq p$, rank (X) = p we can obtain a decomposition

(3) $$X = QR$$

where Q is orthogonal, i.e., $Q'Q = I_n$ and

$$R = \begin{pmatrix} \tilde{R} \\ ---- \\ 0 \end{pmatrix} , \quad \tilde{R} \text{ upper triangular} \tag{4}$$

Q is often obtained as the product of p Householder's transformations. A square matrix of the form $H = I - 2ww'$, where $w'w = 1$, is said to be a Householder transformation. It is easy to check that H is symmetric and orthogonal. The matrix X, in statistical work, is usually a data matrix. An efficient algorithm for solving the linear regression/least squares problem based on HD is given in [2].

Note also that Householder's decomposition of X yields the Cholesky decomposition of X'X, i.e., $L = \tilde{R}'$, since from $X = QR$, $X'X = R'Q'QR = R'R = \tilde{R}'\tilde{R}$.

2.4 The Singular Value Decomposition (SVD)

It is well known, see e.g. [7], that if M is any $p \times q$ real matrix, $p \leq q$, then there exist two orthogonal matrices S $p \times p$ and T $q \times q$ such that

$$S' M T = (D,0) \tag{5}$$

where

$$D = \text{diag}(d_i) \qquad i = 1,2,\ldots,p$$

$$0 \quad \text{is the } p \times (q-p) \text{ matrix}$$

$$d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$$

are the <u>singular values</u> of M (i.e., the non-negative square roots of the eigenvalues of MM'.)

Equivalently, $M = S(D,0)T'$, and if rank(M) = k $\quad d_{k+1} = \cdots = d_p = 0$.

If we denote the columns of S and T by S_i and T_i, respectively, then T_i's are the eigenvectors of MM' corresponding to d_i^2 and $T_i = \frac{1}{d_i} M'S_i$, $i = 1,\ldots,\text{rank}(M)$. T_j, for $j = \text{rank}(M) + 1,\ldots,q$ can be chosen so that T is orthogonal. An efficient algorithm for SVD and its application to least squares can be found in [10].

2.5 Generalized Symmetric Eigenproblem (GSE)

It is well known that, if A and B are symmetric matrices

$$\min_{|x|=1} \frac{x'Ax}{x'Bx}$$

corresponds to the smallest eigenvalue of $(A - \lambda B)x = 0$. If B is positive definite the solution of the generalized symmetric eigenproblem (GSE)

$$(6) \qquad Ax = \lambda Bx$$

can be reduced to the standard symmetric eigenproblem

$$(7) \qquad Cy = \lambda y$$

where $C = L^{-1}AL^{-t}$ and $y = L'x$ and $LL' = B$ is the Cholesky decomposition of B into upper/lower triangular matrices. Note that the eigenvalues of (6) and (7) are the same. An efficient algorithm for solving (6) can be found in [15]. In statistical work, the solution of (6) is often obtained from

$$(8) \qquad B^{-1}Ax = \lambda x$$

which is, in general, a non-symmetric eigenproblem requiring more computational work.

2.6 Canonical Analysis and Inclination of Subspaces

Canonical correlation theory reduces the study of $p \times q$ correlations between two sets of variates $\{X_1,X_2,\ldots,X_p\}$ and $\{Y_1,Y_2,\ldots,Y_q\}$, $p \leq q$, (assumed without loss of generality--to have zero means) to the study of p canonical correlations λ_i, $i = 1,\ldots,p$, between two sets of canonical variates $\{U_1,U_2,\ldots,U_p\}$ and

$\{V_1, V_2, \ldots, V_q\}$ which are normalized linear combinations of the X_i's and the Y_j's respectively.

If X and Y are jointly distributed with nonsingular sample variance-covariance matrix S partitioned as

$$S = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$$

the problem is then to find matrices $A:p \times p$ and $B:q \times q$ where

$$\text{(9)} \qquad \text{Var-Cov}\begin{pmatrix} U \\ V \end{pmatrix} = \left(\begin{array}{c|c|c} I_p & \Lambda & 0 \\ \hline \Lambda & \multicolumn{2}{c}{} \\ \cline{1-1} 0 & \multicolumn{2}{c}{I_q} \end{array}\right)$$

for $U = AX$ and $V = BY$, $\Lambda = \text{diag}(\lambda_i)$ $i=1,\ldots,p$, with $1 \geq \lambda_1 \geq \cdots \geq \lambda_p \geq 0$.

The sought $\{\lambda_i, A_i, \; i=1,\ldots,p\}$ are the solutions of the generalized eigenproblem

$$\text{(10)} \qquad S_{xy} S_{yy}^{-1} S_{yx} A_i = \lambda_i^2 S_{xx} A_i$$

and

$$\text{(11)} \qquad B_i = \frac{1}{\lambda_i} S_{yy}^{-1} S_{yx} A_i \qquad i = 1,2,\ldots,p$$

To solve the canonical correlation problem, two numerical algorithms have appeared in [4] and [9]. The former based on GSE performs Cholesky decomposition of S_{xx} and S_{yy} while checking for ill-conditioning (i.e., checking for the near dependence of the variables X_i's and Y_j's). The latter is more stable numerically, and is based on SVD; S_{xx}, S_{yy} are not computed, instead, the X and Y data matrices are decomposed by Householder's transformations such that

(12) $$X = Q_x\begin{pmatrix}\tilde{R}_x\\0\end{pmatrix} \qquad Y = Q_y\begin{pmatrix}\tilde{R}_y\\0\end{pmatrix}$$

The canonical correlations are then the singular values of $Q_x' Q_y$.

In its geometrical context, the canonical analysis model can be described as follows:

Let X and Y be full column rank data matrices of dimension $n \times p$ and $n \times q$, respectively, $p \leq q \leq n$. Let L_x be the supspace of dimension p generated by the columns of X and L_y of dimension q be the subspace generated by the columns of Y. The smallest angle $\theta_1 \in [0, \frac{\Pi}{2}]$ between L_x and L_y is defined by

$$\cos\theta_1 = \max_{U\in L_x} \max_{V\in L_y} U'V, \qquad |U| = |V| = 1 \text{ being column vectors}$$

Assume that the maximum is attained for $U = U_1$ and $V = V_1$, then θ_2 is defined as the smallest angle between the orthogonal complement of L_x with respect to U_1 and that of L_y with respect to V_1. Continuing in this way until one of the subspaces is empty, we obtain a set of p canonical <u>angles of inclination</u> $\{\theta_i\}$ between the subspaces L_x and L_y

(13) $$\cos\theta_i = \max_{U\in L_x} \max_{V\in L_y} U'V = U_i' V_i$$

subject to $|U| = |V| = 1$

$$U_i' U_j = U_i' V_j = V_i' V_j = V_i' U_j = 0$$

$$i = 1,\dots,p \quad j = 1,\dots,i-1.$$

Note that the <u>canonical vectors</u> $\{U_i, i = 1,\dots p\}$, $\{V_j, j = 1,\dots p\}$ need not be uniquely defined but the canonical angles always are. Also, the vectors V's can be complemented with (q - p) orthonormal vectors $V_{p+1},\dots,V_q$ so that the U's and the V's are orthonormal bases for the subspaces L_x and L_y.

We also note that the orthogonal projectors

(14) $$P_{L_x} = U'U \text{ and } P_{L_y} = V'V ,$$

so that if $\hat{V}_i$ denotes the least-squares predictor of V_i in L_x

$$\begin{aligned}\hat{V}_i &= V_i P_{L_x} \\ &= V_i (U'U) \\ &= \lambda_i U_i\end{aligned}$$

The dimension of inclination between L_x and L_y is defined as

(15) $$\dim (L_x, L_y) = \text{rank } P_{L_x} P_{L_y}$$

and the coefficient of inclination between L_x and L_y is defined as

(16) $$c(L_x, L_y) = \text{trace } P_{L_x} P_{L_y}$$

(for an introduction to these concepts, we refer to [1].)

As an illustration of dim $(\cdot,\cdot)$ and $c(\cdot,\cdot)$, consider simple linear regression. In this case, L_x and L_y are the rays spanned by the vectors X and Y, respectively. Then,

$$\begin{aligned}c(L_x, L_y) &= \text{trace } P_{L_x} P_{L_y} \\ &= \text{trace } (XX^+)(YY^+) && \text{by (2)} \\ &= \text{trace } [X(X'X)^{-1}X'][Y(Y'Y)^{-1}Y'] && \text{by (2')} \\ &= \text{trace } \frac{XX'}{|X|^2} \quad \frac{YY'}{|Y|^2} \\ &= \frac{|X'Y|^2}{|X|^2 |Y^2|} \quad = \cos^2 \alpha\end{aligned}$$

where α is the angle of inclination between the vectors X and Y.

On the other hand,

$$\dim (L_x, L_y) = \begin{cases} 0 \text{ if and only if } X \perp Y \\ 1 \text{ otherwise}\end{cases}$$

denotes the number of angles of inclination between L_x and L_y.

The properties and geometric interpretations of c(·,·) and dim(·,·) carry over in canonical analysis. The former yields Hooper's trace correlation, as is shown in the next section, while the latter is the number of non-zero canonical correlations.

3. THE GEOMETRY OF LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATION

Consider the familiar structural system of stochastic equations

$$\Gamma Y - BX = E \tag{17}$$

where X is the $m \times n$ matrix of exogenous variables, Y is the $k \times n$ matrix of endogenous variables, B is the $k \times m$ matrix to be estimated, Γ is the $k \times k$ non-singular matrix to be estimated, and E is the $k \times n$ matrix of residuals.

Hooper [12, 13] has applied canonical analysis to the reduced form of (17)

$$Y = \Pi X + W\ . \tag{18}$$

Using ordinary least-squares, estimate Π by $\hat{\Pi} = YX'(XX')^{-1}$ and write

$$Q = (Y'Y)^{-1}\hat{W}\hat{W}' = I - (Y'Y)^{-1}\hat{\Pi}\, XX'\, \hat{\Pi}'\ ,$$

it follows that the eigenvalues of Q are equal to $1 - \lambda_i^2$, where λ_i is the i^{th} canonical correlation between the sets of variates X and Y. Taking Hotelling's "vector alienation" coefficient α and "vector correlation" coefficient β [14], Hooper shows that

$$\alpha = \frac{|\hat{W}\hat{W}'|}{|YY'|} = \left(\prod_{i=1}^{k} (1-\lambda_i^2)\right)^{\frac{1}{2}}$$

$$= |I - Q| = \left(\prod_{i=1}^{k} \lambda_i^2\right)^{\frac{1}{2}}$$

Hence, $\alpha = 1$ is obtained if all canonical correlations are zero (representing independence of the variates X and Y). And $\beta = 0$ if at least one canonical correlation is equal to 0. Noting that both α and β tend to zero when k is large. Hooper proposes taking the "trace correlation",

$$(19) \qquad \Upsilon^2 = \frac{1}{k}\,tr(I - Q) = \frac{1}{k}\sum_{i=1}^{k} \lambda_i^2$$

and concludes that "Υ^2 can be thought of as that part of the total variance of the jointly dependent variables that is accounted for by the systematic part of the reduced from and $1 - \Upsilon^2$ as the unexplained part..."

From our geometrical treatment of canonical correlation, it is easy to show that $\Upsilon^2 = \frac{1}{k}\, c(L_x, L_y)$ where L_x and L_y are the subspaces generated by the rows of X and Y respectively.

Proof: Let U and V be the matrices of canonical variates corresponding to X and Y respectively. From (14) we have:

$$P_{L_x} = U'U$$

$$P_{L_y} = V'V$$

$$\begin{aligned} c(L_x, L_y) = tr\,(P_{L_x} P_{L_y}) &= tr(U'U\,V'V) \\ &= tr(U'(\Lambda,0)\;V) \\ &= tr(VU'\;(\Lambda,0)) \\ &= tr((\Lambda,0)'(\Lambda,0)) \\ &= \sum_{i=1}^{k} \lambda_i^2 \end{aligned}$$

Choose, without loss of generality, the first equation of (17)

$$(18) \qquad \Gamma_{1*}Y_* - B_{1*}X_* = \epsilon_1$$

where Γ_{1*}, B_{1*} and ϵ_1 are the first rows of Γ, B, and ϵ, respectively, with elements suitably rearranged in the equation to correspond to the included m_* exogenous and k_* endogenous variables X_* and Y_*, respectively. Assume (18) is over-identified and that Γ_{1*} is normalized by having its first element set at -1.

To estimate Γ_{1*} and B_{1*} one could find linear combinations of the vectors $Y_1, Y_2, \ldots, Y_{k*}$ and $X_1, X_2, \ldots, X_{m*}$, respectively, such that $\Gamma_{1*}Y_*$ is the sum of a vector $B_{1*}X_*$ in the subspace $L_{x*} = R(X_*')$ and a vector ϵ_1 in $L^{\perp}_{x*}$. The ordinary least-squares criterion for finding Γ_{1*} and B_{1*} is to minimize $|\epsilon_1|^2$; equivalently, the vectors $\Gamma_{1*}Y_*$ and $B_{1*}X_*$ can be chosen so that they have minimum angle θ_1 between them. From §2.6, it is clear that the first pair of canonical vectors with correlation λ_1 has this minimal property, i.e.,

$$V_1 = \Gamma_{1*}Y_* \quad , \quad U_1 = B_{1*}X_*$$

and if $\hat{V}_1$ denotes the least-squares preditor of V_1 in L_{x*},

$$\hat{V}_1 = V_1 P_{L_{x*}} = V_1(U'U) \text{ by (14) } = \lambda_1 U_1 \qquad (\cos \theta_1 = \lambda_1)$$

Note, however, that this minimum angle estimation procedure does not take into account the fact that the included endogenous variables Y_* depend on the exogenous variables X_{**} present in the model but excluded from the first equation. The aim of LIML is to let the addition of excluded exogenous variables make a minimal improvement in estimating Y_* from X_*.

From this vantage point, a variety of estimation procedures is possible. When $\Gamma_{1*}Y_*$ is projected on L_{x*} it yields a residual vector R_{x*}, when it is projected on L_x, it yields a residual vector R_x. In the context of minimum variance ratio, LIML minimizes the ratio $|R_{x*}|^2 / |R_x|^2$. In the context of canonical correlation, consider the subspaces $L^{\perp}_{x*}$ with respect to L_{y*} and L_{x**} and calculate the maximum angle of inclination between them. This is because the residuals in equation (18) should be as little correlated as possible with X_{**}.

The LIML procedure amounts to

(19)
$$\min_{\Gamma_{1*}} \frac{\Gamma_{1*}'Y_*'Q_*Y_*\Gamma_{1*}}{\Gamma_{1*}'Y_*'Q\,Y_*\Gamma_{1*}}$$

where Y_* is $k_* \times n$, X_* is $m_* \times n$

$$Q_* = I - X_*(X_*'X_*)^{-1}X_*'$$

and $$Q = I - X\,(X\,'X)^{-1}X'$$

Noting that (19) expresses the ratio of residual variation using X_* to residual variation using X, one may propose an estimation procedure that minimizes the difference between the two residual variations, i.e.

$$\min_Z \; Z'Q_*Z - Z'QZ \qquad , Z = \Gamma_{1*}'Y_*$$

$$= \min_Z \; Z'Z - Z'P_*Z - Z'Z + Z'PZ \qquad , P = I - Q$$
$$P_* = I - Q_*$$

$$= \max_Z \; Z'P_*Z - Z'PZ$$

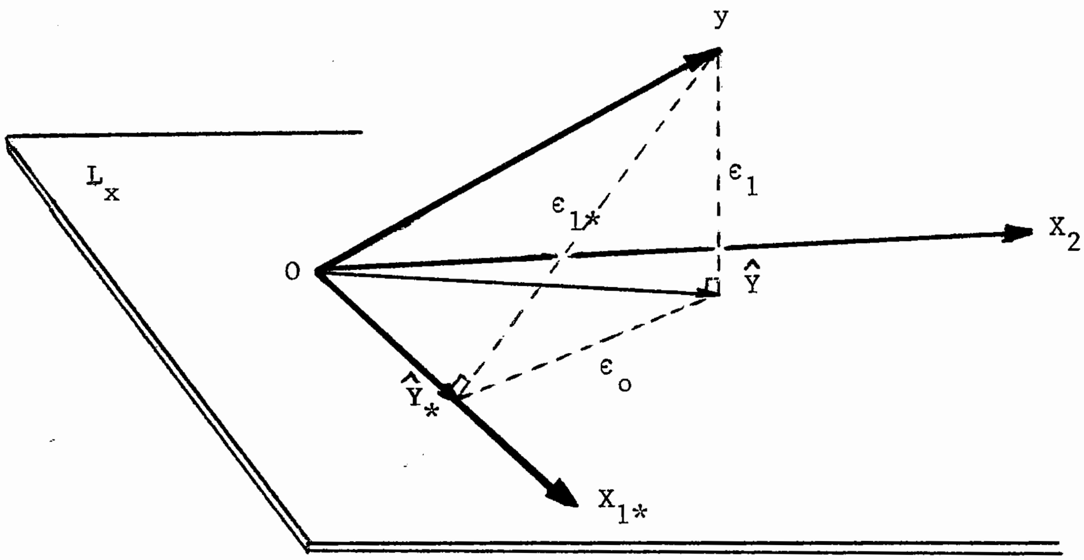
Since the second term of the expression represents variation of endogenous variables explained by the full set of exogenous variables, one constrains the optimization problem with $|Z'PZ| = c$, a constant.

$$\max_Z \frac{Z'P_*Z}{Z'PZ} \leq 1$$

This can be solved via the eigenvalue problem

$$P_*Z = \mu PZ$$

A simple graphical interpretation of these estimation procedures is presented in the next page for the case of $Y = \{y\}$, $X = \{X_1, X_2\}$, $X_* = \{X_1\}$.



LIML $\quad \min \dfrac{y'P^{\perp}_{x*}y}{y'P^{\perp}_{x}\,y} = \min \dfrac{\varepsilon_{1*}}{\varepsilon_1}$, "minimum variance", "minimum angle between residuals"

HANNAN $\quad \min \dfrac{y'(P_x - P_{x*})y}{y'P^{\perp}_{x*}\,y} = \min \dfrac{\varepsilon_o}{\varepsilon_{1*}}$, "maximum angle between ε_{1*} and X_{**}"

$\max \dfrac{y'P_{x*}y}{y'P_x\,y} = \max \dfrac{\hat{Y}_*}{\hat{Y}}$, "minimum difference between ε_{1*} and ε_1"

(length of all denominator vectors = constant)

For a comparison of LIML with other estimators and an extension of canonical correlation to multiple equation systems, see the excellent papers of Chow [3] and Hannan [11], respectively. A more recent minimum-distance interpretation of LIML can also be found in [8].

4. COMPUTATIONS IN LIML ESTIMATION

Only recently has an effort been made to graft numerical algorithms of linear algebra to statistical estimation procedures. Statisticians are just beginning to recognize in numerical analysis the algorithms they need for obtaining accurate solutions to their problems. For example, the least squares/linear regression model has been thoroughly investigated by several authors.

The LIML procedure (19) reduces in turn to the eigenproblem

(20) $$(Y_*'Q_*Y_* - \gamma Y'QY)\,\Gamma_{1*} = 0$$

for which we choose the smallest eigenvalue γ_{min}, then

(21) $$F_{1*} = -(X_*'X_*)^{-1}X_*'Y_*\,\Gamma_{1*}\,.$$

Noting that (20) is a generalized symmetric eigenproblem, it can be solved by the method proposed in section 2.5 ($Y_*'Q_*Y$ takes the place of the matrix A, and $Y'QY$ takes the place of the matrix B). The smallest eigenvalue can then be extracted by standard methods.

In a recent report [6], Dent and Golub have proposed computing LIML estimators by Householder decomposition of the data matrices X and Y and application of the singular value decomposition. Their approach also exhibits the equivalence of 2SLS (two-stage least-squares) to LIML in the case where (18) is just identified.

It is not easy to state which computational method is preferable. In a previous report [5], the author has discussed some of the practical considerations for choosing between Cholesky decomposition/SGE and a Householder decomposition/SVD.

For practical implementation in a statistical programming system*, Cholesky/SGE has storage advantages: a compact storage scheme can be used to handle the symmetric sum-of-squares and cross-product matrices. Provided that these matrices are computed accurately, the method is numerically stable and cheapest in computations. On the other hand, Householder/SVD avoids possible ill-conditioning and provides a good global definition of the rank of the data matrices involved. But all singular values have to be computed.

---

*Both methods are being included in SPSS (Statistical Package for the Social Sciences) for comparison on large problems using Northwestern University's CDC 6400 computer.

References

1. Afriat, S. N., "Orthogonal and oblique projectors and the characteristics of vector spaces", Proc. Camb. Philos. Soc., 53 (1957), pp. 800-816.

2. Businger, P. and G. Golub, "Linear least squares solutions by Householders' transformations", Num. Math., 7, (1965), pp. 269-276.

3. Chow, C. G., "A comparaison of alternative estimators for simultaneous equations", Econometrica, 32, (1964), pp. 532-553.

4. Cohen, C. and A. Ben-Israel, "On the computation of canonical correlations", Cahier du Centre d'Etudes de Recherche Operationnelle, 11, (1969), pp. 121-132.

5. Cohen, C., "The generalized symmetric eigenproblem in multivariate statistical models", Discussion Paper no. 96, June 1974, Center for Mathematical Studies in Economics and Management Science, Northwestern University, Evanston, Illinois.

6. Dent, W. and G. Golub, "Computation of the limited information maximum likelihood estimator", STAN-CS-73-339, (1973), Computer Science Department, Stanford University, Stanford, California.

7. Eckart, C. and G. Young, "A principal axis transformation for non-hermitian matrices", Bull. Amer. Math. Soc., 45, (1939), pp. 118-121.

8. Goldberger, A. and I. Olkin, "A minimum distance interpretation of limited information estimation", Econometrica, 39, (1971), pp. 635-639.

9. Golub, G. "Matrix decompositions and statistical computations" in Statistical Computation, (Milton and Nelder, eds.), Academic Press, 1969.

10. Golub, G. and C. Reinsch, "Singular value decomposition and least-squares solutions", Num. Math., 14, (1970), pp. 403-420.

11. Hannan, E.J. "Canonical correlation and multiple equation systems in economics", Econometrica, 35, (1967), pp. 123-138.

12. Hooper, J. W., "Simultaneous equations and canonical correlation theory", Econometrica, 27, (1959), pp. 245-256.

13. Hooper, J. W., "Partial trace correlations", Econometrica, 30, (1962), pp. 324-331.

14. Hotelling, H., "Relations between two sets of variates", Biometrika, 28, (1936), pp. 321-377.

15. Martin, R.S. and J. H. Wilkinson, "Reduction of the symmetric eigenproblem $Ax = \lambda Bx$ and related problems to standard form", Num. Math., 11, (1968), pp. 99-110.

16. Penrose, E., "A generalized inverse for matrices", Proc. Camb. Phil. Soc., 51, (1955), pp. 406-413.

17. Wilkinson, J. H. and C. Reinsch, Linear Algebra, Handbook for Automatic Computation, Volume II, Springer-Verlag (1971).