AD-A096 958 STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
GENERALIZED CORRELATIONS IN THE SINGULAR CASE.(U)
NOV 80 A S GUPTA N00014-75-C-0442
UNCLASSIFIED TR-46 NL

1 OF 1
AD A
096958

END
DATE
FILMED
4-81
DTIC

LEVEL II 12

GENERALIZED CORRELATIONS IN THE SINGULAR CASE

BY

ASHIS SEN GUPTA

TECHNICAL REPORT NO. 46
NOV[redacted]80

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

GENERALIZED CORRELATIONS IN THE SINGULAR CASE

by

ASHIS SEN GUPTA
Stanford University

TECHNICAL REPORT NO. 46
NOVEMBER 1980

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Also issued as Technical Report No. 161 under National Science
Foundation Grant MCS78-07736, Dept. of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

GENERALIZED CORRELATIONS IN THE SINGULAR CASE

Ashis Sen Gupta

1. Introduction. When the covariance matrix is singular, the usual expressions for the multiple, partial, canonical [see, e.g., Rao (1973)] and some generalized canonical correlations [for a review, see Sen Gupta (1979)] need to be revised. Tucker, et al. (1972), Khatri (1976) and Rao (1979), (1980) have provided formulae for some of these correlation coefficients in the general case by using g-inverses. A review of their results on multiple, partial and canonical correlations is given first. Next, it is shown that there exists a general representation which covers several generalized canonical correlations and as special cases the multiple, partial and canonical correlations, too. Then a general theory is formulated which deals with the singular case for the representation. Previous results on multiple, partial and canonical correlations follow as special cases of this theory. Further, appropriate formulae are also provided through this formulation for various generalized canonical correlations in the singular case. Finally, the numbers of various critical generalized correlations are derived for the general case.

2. Multiple, partial and canonical correlations in the singular case. Let $R = (R_1 : \ldots : R_p)$ be the correlation matrix of p variables. Further, let $R^- = (r^{ij}) = (T_1 : \ldots : T_p)$ be any g-inverse of R. Define, $RR^- = Q = (Q_1 : \ldots : Q_p)$. Let

I_p have the unit vector e_i, as its i-th column, i = 1,...,p.

Result 1. The squared multiple correlation of X_1 on $X_2,\ldots,X_p$ is

$$R^2_{1\cdot(2\ldots p)} = 1 \text{ if } Q_1 \neq e_1$$
$$= 1 - (r^{11})^{-1} \text{ if } Q_1 = e_1 \; .$$

Result 2. The partial correlation between X_1 and X_2 eliminating $X_3, X_4, \ldots, X_p$ is

$$r_{12\cdot(34\ldots p)} = 0 \text{ if } Q_1 \neq e_1 \text{ and } Q_2 = e_2 \text{ or if } Q_1 = e_1 \text{ and } Q_2 \neq e_2$$
$$= 1 \text{ if } Q_1 \neq e_1 \text{ and } Q_2 \neq e_2$$
$$= -r^{12}/(r^{11}r^{22})^{1/2} \text{ if } Q_1 = e_1 \text{ and } Q_2 = e_2 \; .$$

Let X_1 and X_2 be two sets of variables with the joint dispersion matrix Σ, partitioned accordingly.

Result 3. The squared canonical correlations are the non-zero roots of the determinantal equation $|\Sigma_{11}^{-}\Sigma_{12}\Sigma_{22}^{-}\Sigma_{21} - \rho^2 I| = 0$ where Σ_{11}^{-} and Σ_{22}^{-} are any g-inverses of Σ_{22} and Σ_{11} respectively.

For proofs and further discussions on the results see Rao [(1979), (1980).]

3. Generalized canonical correlations in the singular case. Canonical correlations have been generalized in various ways. Formulae in the general case will be provided here for those obtained by extending the concepts of tests of independence for two sets of

variates--giving rise to partial, part and bipartial canonical correlations [see Timm (1975) pp. 352-353] and g_1 - , g_2 - bipartial canonical correlations [see Lee (1978)] and some association measures [see McKeon (1965), pp. 16-19]. Various other authors [Horst (1961); Edgerton and Kolbe (1936); Wilks (1938); Lord (1958)] arrived at the same solution as that of McKeon for the particular case of a single variable per set. Appropriate formula will also be provided for the new generalized canonical correlation arising out of the concept of minimum generalized variance [proposed by Anderson (1958) Problem 5, pp. 305-306 and derived by the author {see SenGupta (1979)} under constraint of equi-correlation structure of the generalized canonical variables].

Let $X = (X_1, \ldots, X_k)$, $X_i : p_i \times 1$, $p_1 + \ldots + p_k = p$, $\mathrm{Disp}(X) = {}_k\Sigma$, $\mathrm{Cov}\,(X_i, X_j) = \Sigma_{ij}$ and non-zero ρs be the generalized canonical correlations. Starting with the defining equations it can be easily seen that for all the above cases, the generalized canonical correlations are obtained from the eigen values of ${}_k\Sigma^*$ in the metric of ${}_k\Sigma^*_d$, i.e. from the solutions of

$$(3.1) \qquad |{}_k\Sigma^* - \rho^* \, {}_k\Sigma^*_d| = 0$$

where $\rho^* = 1 + (k - 1)\rho$ and ${}_k\Sigma^*_d$ is a diagonal super matrix with elements Σ^*_{ii} $i = 1, \ldots, k$. In the notation of Lee, ${}_k\Sigma^*$, with $k = 2$, is the covariance matrix of the residual vectors $(\tilde{e}_{1.34}, \tilde{e}_{2.35})$ and $(e_{1.34}, e_{2.35})$ for the g_1- and

g_2-bipartial canonical correlations, respectively. In the notation of Timm, ${}_k\Sigma^* = \Sigma_{\cdot 3}$, $\Sigma_{1(2\cdot 3)}$ and $\Sigma_{(1\cdot 4)(2\cdot 3)}$ with k = 2 for partial, part and bipartial canonical correlations respectively. Also for McKeon's and the new generalized canonical correlations, ${}_k\Sigma^* = {}_k\Sigma$.

Theorem. The generalized canonical correlations, for the methods quoted above, are given by $\rho = (\rho^* - 1)/(k - 1)$ where ρ^*'s are the non-zero roots of $|{}_k\Sigma^* \, {}_k\Sigma_d^{*-} - \rho^* I| = 0$, ${}_k\Sigma_d^{*-}$ being any g-inverse of ${}_k\Sigma_d^*$.

Proof. First note the representation (3.1). Consider next the following Lemmas.

Lemma 1. Let A be a hermitian matrix of order n and rank s, and B be non-negative definite matrix of order n and rank r such that $S(A) \subset S(B)$ [where S(M) represents the vector space spanned by the column vectors of M]. Then

(i) There exists a matrix L of order $n \times r$ such that $L'AL = \Lambda$, $L'BL = I_r$, where Λ is a diagonal matrix with s non-zero elements, some of which may be repeated and I_r is the identity matrix of order r.

(ii) The non-zero elements of Λ are the same as the roots of the determinantal equation, $|AB^- - \lambda I| = 0$ with repetitions allowed, for any g-inverse B^- of B.

Proof of Lemma 1. See Lemma 3, Rao (1979).

Lemma 2. $S({}_k\Sigma^*) \subset S({}_k\Sigma_d^*)$.

Proof of Lemma 2. Note that $S(\Sigma_{ij}^*) \subset S(\Sigma_{ii}^*)$ for all the ${}_k\Sigma^*$ considered above. This follows immediately from the result [see Proposition 3.31, pp. 3.15-3.16 of Eaton

that, if $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \geq 0$ then $N(A_{22}) \subset N(A_{12})$ and

and $N(A_{11}) \subset N(A_{21})$ where $N(M)$ is the null space of M.
Then, there exist matrices B_{ij} such that $B_{ij}\Sigma^*_{jj} = \Sigma^*_{ij}$, $i,j = 1,\ldots,k$.
Hence, there exist matrices C_i such that $(C_1 \ldots C_k)_k\Sigma^*_d = C\,{}_k\Sigma^*_d = {}_k\Sigma^*$
which proves Lemma 2.
Coupling Lemma 2 with Lemma 1 proves the Theorem.
Note: For $k = 2$, if $p_1 = 1$, $p_2 > 1$ and if $p_1 > 1$, $p_2 > 1$ then we have the cases of multiple and canonical correlations respectively. Further, with $k = 2$, consideration of residual variables leads to partial correlation. Thus the above Theorem unifies the Results 1 through 3 and also considers simple (and not squared) multiple, partial and canonical correlations.

4. Numbers of critical generalized correlations.
Let A and B be two hermitian matrices and B be non-negative definite. If λ is a constant and v a vector such that $Av = \lambda Bv$, $Bv \neq 0$, then λ is called a proper eigen value and v a proper eigen vector of A with respect to B. In the context of Lemma 1, the elements of Λ are called the proper eigen values and the corresponding columns of L, the proper eigen vectors of A with respect to B. For the generalized correlations, we consider from (3.1) only the proper eigen

values of ${}_k\Sigma^*$ with respect to ${}_k\Sigma^*_d$. Also note that for $k \geq 2$, 1 and $-1/(k - 1)$ are the maximum and minimum possible values, respectively, for the generalized correlations. Let ${}_k\Sigma^*_{od}$ be the super off-diagonal matrix such that ${}_k\Sigma^* = {}_k\Sigma^*_d + {}_k\Sigma^*_{od}$. Also let R(M) denote the rank of the matrix M.

<u>Lemma 3</u>. The numbers of zero, unit and $-1/(k - 1)$-valued generalized correlations are given by $r - R({}_k\Sigma^*_{od})$, $r - R[{}_k\Sigma^*_{od} - (k - 1)\,{}_k\Sigma^*_d]$ and $r - R({}_k\Sigma^*)$ respectively, where $r = R({}_k\Sigma^*_d)$.

<u>Proof</u>: The proof follows by rewriting (3.1) as $|{}_k\Sigma^*_1 - \lambda\,{}_k\Sigma^*_d| = 0$ where $({}_k\Sigma^*_1, \lambda) = [{}_k\Sigma^*_{od}, (k - 1)\rho], [{}_k\Sigma^*_{od} - (k - 1)\,{}_k\Sigma^*_d, (k - 1)(\rho - 1)]$ and $[{}_k\Sigma^*, (k - 1)\rho + 1]$ for the zero, unit and $-1/(k - 1)$-valued generalized correlations respectively and noting the one-one relationship between λ and ρ.

<u>Acknowledgements</u>. I am grateful to Professor C. R. Rao and Professor T. W. Anderson for their kind remarks on the subject.

REFERENCES

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

[2] Eaton, L. M. (1972). Multivariate Statistical Analysis. Institute of Mathematical Statistics, University of Copenhagen.

[3] Edgerton, H. and Kolbe, E. L. (1936). The method of minimal variation for the combination of criteria. Psychometrika I, 183-187.

[4] Horst, P. (1961). Generalized canonical correlations. J. Clin. Psychol., Mon. Suppl. No. 14.

[5] Khatri, C. G. (1976). A note on multiple and canonical correlation for a singular covariance matrix. Psychometrika 41, 465-470.

[6] Lee, S. (1978). Generalizations of the partial, part and bipartial canonical correlation analysis. Psychometrika 43, 427-431.

[7] Lord, F. M. (1958) Some relations between Guttman's principal components of scale analysis and other psychometric theory. Psychometrika 23, 291-295.

[8] McKeon, J. J. (1965). Canonical Analysis: some relations between canonical correlations, factor analysis, discriminant function analysis and scaling theory. Psychometric Monograph No. 13.

[9] Rao, C. R. (1973). Linear Statistical Inference. Wiley, New York.

[10] Rao, C. R. (1979). Multiple, partial and canonical correlations in the singular case. Tech. Report 174, Dept. of Statistics, The Ohio State University.

[11] Rao, C. R. (1980). A lemma on g-inverse of a matrix and computation of correlation coefficients in the singular case. Tech. Report 80-5, Institute for Statistics and Applications, University of Pittsburgh.

[12] SenGupta, A. (1979). On the problems of construction and statistical inference associated with a generalization of canonical variables. Ph.D. Dissertation. Dept. of Statistics, The Ohio State University.

[13] Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology. Brooks Cole Publishing Co., Monterey, Calif.

[14] Tucker, L. R., Cooper, L. G. and Meredith, W. Obtaining squared multiple correlations from a correlation matrix which may be singular. Psychometrika 37, 143-148.

[15] Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika 3, 23-40.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

33. "Canonical Correlations with Respect to a Complex Structure," *Steen A. Andersson*, July 1978.

34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and I. Olkin, July 1978.

35. " Maximum likelihood Estimation for Vector Autoregressive Moving Average Models," T. W. Anderson, July 1978.

36. "Maximum likelihood Estimation of the Covariances of the Vector Moving Average Models in the Time and Frequency Domains," F. Ahrabi, August 1978.

37. "Efficient Estimation of a Model with an Autoregressive Signal with White Noise," Y. Hosoya, March 1979.

38. "Maximum Likelihood Estimation of the Parameters of a Multivariate Normal Distribution,"T.W. Anderson and I. Olkin, July 1979.

39. "Maximum Likelihood Estimation of the Autoregressive Coefficients and Moving Average Covariances of Vector Autoregressive Moving Average Models," Fereydoon Ahrabi, August 1979.

40. "Smoothness Priors and the Distributed Lag Estimator," Hirotugu Akaike, August, 1979.

41. "Approximating Conditional Moments of the Multivariate Normal Distribution," Joseph G. Deken, December 1979.

42. "Methods and Applications of Time Series Analysis - Part I: Regression, Trends, Smoothing, and Differencing," T.W. Anderson and N.D. Singpurwalla, July 1980.

43. "Cochran's Theorem, Rank Additivity, and Tripotent Matrices." T.W. Anderson and George P.H. Styan, August, 1980.

44. "On Generalizations of Cochran's Theorem and Projection Matrices," Akimichi Takemura, August, 1980.

45. "Existence of Maximum Likelihood Estimators in Autoregressive and Moving Average Models," T.W. Anderson and Raúl P. Mentz, Oct. 1980.

46. "Generalized Correlations in the Singular Case," Ashis Sen Gupta, November 1980.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 46	2. GOVT ACCESSION NO. AD-A096 958	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) GENERALIZED CORRELATIONS IN THE SINGULAR CASE		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) ASHIS SEN GUPTA		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, Stanford University, Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Statistics and Probability Program Code 436, Arlington, Virginia 22217		12. REPORT DATE November 1980
		13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This report was also issued as Technical Report No. 161 under National Science Foundation Grant MCS78-07736, Department of Statistics, Stanford University, Calif.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Generalized canonical correlations, g-inverse, hermitian matrix, singular covariance matrix.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general result is given which provides appropriate formulae for various generalizations of canonical correlations in the singular case. This covers as special cases the results for multiple correlation due to Tucker, Cooper and Meredith (1972) and Khatri (1976) and for partial and canonical correlations due to Rao (1979), (1980). The numbers of zero, unit and other critical generalized correlations are also given for the general case.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
FILMED
-8