ASSESSING DIMENSIONALITY IN MULTIVARIATE REGRESSION

by

Alan Julian Izenman

Technical Report #342
February 1979

Department of Applied Statistics
University of Minnesota
St. Paul, Minnesota 55108

To appear in Handbook of Statistics, Vol. 1. (ed. P.R. Krishnaiah), North-Holland Publishing Company.

SUMMARY

The rank of the regression coefficient matrix in a multivariate linear regression model is referred to in this paper as the dimensionality of that multivariate regression. The problem of assessing the value of that rank when both sets of variates in the regression are jointly distributed is related to the classical multivariate problems of determining the number of principal components or pairs of canonical variates to use in a given practical situation. Different methods for assessing the rank of the regression coefficient matrix are described here, with emphasis on inference from certain graphical displays. The concept of a rank trace plot is introduced and illustrated using a number of real-data examples. Further information regarding the dimensionality of the regression is obtained through comparing gamma probability plots of the multivariate residual vectors.

KEY WORDS: Reduced-rank regression, principal component analysis, canonical variate analysis, residual analysis, rank trace, gamma probability plots.

1. INTRODUCTION

In this paper we describe a certain generalization of the multivariate linear regression model which also provides a unified approach to the classical multivariate techniques of principal component and canonical variate and correlation analysis.

This regression model can be described as follows. Let

$$\begin{bmatrix} \underset{\sim}{X} \\ \underset{\sim}{Y} \end{bmatrix} \tag{1.1}$$

be a collection of r + s variables partitioned into two disjoint subcollections, where $\underset{\sim}{X} = [X_1,\ldots,X_r]^\tau$ has r components, $\underset{\sim}{Y} = [Y_1,Y_2,\ldots,Y_s]^\tau$ has s components, and $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are jointly distributed with mean vector and covariance matrix given by

$$E\begin{bmatrix} \underset{\sim}{X} \\ \underset{\sim}{Y} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\mu}_X \\ \underset{\sim}{\mu}_Y \end{bmatrix} \tag{1.2}$$

$$E\begin{bmatrix} \underset{\sim}{X} - \underset{\sim}{\mu}_X \\ \underset{\sim}{Y} - \underset{\sim}{\mu}_Y \end{bmatrix}\begin{bmatrix} \underset{\sim}{X} - \underset{\sim}{\mu}_X \\ \underset{\sim}{Y} - \underset{\sim}{\mu}_Y \end{bmatrix}^\tau = \begin{bmatrix} \underset{\sim}{\Sigma}_{XX} & \underset{\sim}{\Sigma}_{XY} \\ \underset{\sim}{\Sigma}_{YX} & \underset{\sim}{\Sigma}_{YY} \end{bmatrix} = \underset{\sim}{\Sigma} \tag{1.3}$$

respectively, where $\underset{\sim}{\Sigma}_{XX}$ and $\underset{\sim}{\Sigma}_{YY}$ are both assumed nonsingular, $\underset{\sim}{A}^\tau$ denotes the transpose of the matrix $\underset{\sim}{A}$, and E is the expectation operator defined by the distribution associated with the variate (1.1). Assume further that the variates $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are linearly related, so that

$$\underset{\sim}{Y}^{s\times 1} = \underset{\sim}{\mu}^{s\times 1} + \underset{\sim}{C}^{s\times r}\,\underset{\sim}{X}^{r\times 1} + \underset{\sim}{\varepsilon}^{s\times 1}, \tag{1.4}$$

where $\underset{\sim}{\mu}$ and $\underset{\sim}{C}$ are unknown parameters and $\underset{\sim}{\varepsilon}$ is the corresponding error variate of the model, uncorrelated with $\underset{\sim}{X}$ and having mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Sigma}_{\varepsilon\varepsilon}$.

Descriptions of this model in the literature assume implicitly that the regression coefficient matrix $\underset{\sim}{C}$ has full rank, and then demonstrate that simultaneous (unrestricted) least-squares estimation applied to all s equations of (1.4) yields the same results as does equation-by-equation least-squares. As a result, nothing is gained by estimating the equations jointly.

A true multivariate feature enters the model when we know (or suspect) that the regression coefficient matrix $\underset{\sim}{C}$ may not have full-rank, that in fact,

$$\text{rank } (\underset{\sim}{C}) = t \leq \min(r,s) = s, \tag{1.5}$$

say, so that a number of linear restrictions on the set of regression coefficients of the model may be present. The value of t in (1.5), and hence the number and nature of those restrictions, may or may not be known prior to analysis.

Although $\underset{\sim}{X}$ is presumed in (1.5) to be the larger of the two sets of variates, this reflects purely a mathematical convenience, and similar expressions as appear in this paper can also be obtained for the case in which $r \leq s$.

The statistical literature contains some discussion of this type of multivariate regression model. Most of the papers on the subject assume that the value of t in (1.5) is either known à priori ([5], [6,p.335], [13], [17,p. 505], [18], [19], [21]) or that a suitable hypothesized value can be stated for the value of t ([1], [2, Section 14.2], [21]). With this

in mind, it was found convenient (in [13], [18], [19]) to distinguish the case $t = s$ from the case $1 \leq t < s$ by terming the former full-rank regression and the latter reduced-rank regression. Similarly, the regression coefficient matrix $\underset{\sim}{C}$ is called 'full-rank' or 'reduced-rank' as appropriate, and to show dependence on its rank, the matrix $\underset{\sim}{C}$ is also written $C^{(t)}$.

Several of the above-mentioned papers were specifically concerned with the relationships between multivariate regression analysis and the dimensionality-reduction techniques of principal component analysis ([16], [5, Ch.9], [13]) and canonical variate and correlation analysis ([5, Ch. 10], [6], [13], [18], [19], [21], [24]). Bartlett [3], however, seems to have been the first to observe the important connections between these various methodologies.

What is probably of greatest interest to the statistician is the case in which the rank of $\underset{\sim}{C}$ cannot be so specified beforehand and has instead to be determined from a given multivariate sample of n independent observations,

$$\begin{bmatrix} \underset{\sim}{X}_j \\ \underset{\sim}{Y}_j \end{bmatrix}, \quad j = 1,2,\ldots,n, \tag{1.6}$$

on the variate (1.1). Such data will introduce noise into the relationship between $\underset{\sim}{Y}$ and $\underset{\sim}{X}$, and hence will tend to obscure the actual structure of the matrix $\underset{\sim}{C}$, so that rank determination for any particular problem will be made more difficult. There is, therefore, a need to make a distinction here between the "true" or "mathematical" rank of $\underset{\sim}{C}$, which will 'always' be full (since it will be based on a sample estimate of $\underset{\sim}{C}$), and the "practical" or "statistical" rank of $\underset{\sim}{C}$ --- the one of real interest --- which will typically be unknown.

The problem is, therefore, a selection problem: from the set of integers from 1 through s, we are to choose the smallest integer such that the reduced-rank regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X}$ with that integer as rank will be close (in some sense) to the corresponding full-rank regression. The sense by which one multivariate regression can be 'close' to another multivariate regression forms the subject of this paper. Section 2 gives the main results concerning reduced-rank regression and its relationship to principal component and canonical variate analysis. Section 3 discusses the nature of the residuals from a specific (known rank) reduced-rank regression. Section 4 introduces the problem of assessing the rank of $\underset{\sim}{C}$ and Section 5 illustrates some of these concepts through a simple but interesting real-data example. Sections 6 and 7 then consider new types of graphical displays by which that dimensionality may be determined.

2. REDUCED-RANK REGRESSION: MAIN RESULTS

The general objective of reduced-rank regression, therefore, is to approximate the s-vector variate $\underset{\sim}{Y}$ by a set of s linear combinations, $\underset{\sim}{\mu} + \underset{\sim}{C}\,\underset{\sim}{X}$, of the r-vector variate $\underset{\sim}{X}$ in which the number of linearly independent combinations (i.e. dimensionality of the regression) may be less than s. In the event that there exist t (<s) such combinations, the (regression coefficient) matrix $\underset{\sim}{C}$ will have rank t so that there will exist two (non-unique) matrices, an (s×t)-matrix $\underset{\sim}{A}$ and a (t×r)-matrix $\underset{\sim}{B}$, such that $\underset{\sim}{C} = \underset{\sim}{A}\,\underset{\sim}{B}$, where $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are both of rank t. The problem, therefore, becomes one of finding an $\underset{\sim}{A}$ and a $\underset{\sim}{B}$ such that the variate $\underset{\sim}{\mu} + \underset{\sim}{A}\,\underset{\sim}{B}\,\underset{\sim}{X}$ is approximately equal to $\underset{\sim}{Y}$. We have the following general result.

THEOREM. Let (1.1) be an (r + s) - vector-valued variate having mean vector (1.2) and covariance matrix (1.3). Suppose that $\underset{\sim XX}{\Sigma}$ is non-singular and that $\underset{\sim}{\Gamma}$ is positive-definite symmetric. Then, the (s × 1) - vector $\underset{\sim}{\mu}$, an (s × t) - matrix $\underset{\sim}{A}$, and a (t × r) - matrix $\underset{\sim}{B}$, where $1 \le t \le s \le r$, that minimize

$$E\{(\underset{\sim}{Y} - \underset{\sim}{\mu} - \underset{\sim}{ABX})^{T}\underset{\sim}{\Gamma}(\underset{\sim}{Y} - \underset{\sim}{\mu} - \underset{\sim}{ABX})\} \tag{2.1}$$

are given by

$$\underset{\sim}{A} = \underset{\sim}{A}^{(t)} = \underset{\sim}{\Gamma}^{-\frac{1}{2}}[\underset{\sim 1}{V},\ldots,\underset{\sim t}{V}] \tag{2.2}$$

$$\underset{\sim}{B} = \underset{\sim}{B}^{(t)} = \begin{bmatrix} \underset{\sim 1}{V}^{T} \\ \vdots \\ \underset{\sim t}{V}^{T} \end{bmatrix} \underset{\sim}{\Gamma}^{\frac{1}{2}}\ \underset{\sim YX}{\Sigma}\ \underset{\sim XX}{\Sigma}^{-1} \tag{2.3}$$

$$\underset{\sim}{\mu} = \underset{\sim}{\mu}^{(t)} = \underset{\sim Y}{\mu} - \underset{\sim}{A}^{(t)}\underset{\sim}{B}^{(t)}\underset{\sim X}{\mu}, \tag{2.4}$$

where $\underset{\sim}{V}_j$ is the latent vector corresponding to the j^{th} largest latent root, λ_j, of the matrix

$$\underset{\sim}{\Gamma}^{\frac{1}{2}} \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1} \underset{\sim}{\Sigma}_{XY} \underset{\sim}{\Gamma}^{\frac{1}{2}} , \tag{2.5}$$

$j = 1,2,\ldots,s$. At the minimum, the criterion (2.1) has the value

$$W(t) = \operatorname{tr}\{\underset{\sim}{\Sigma}_{YY} \underset{\sim}{\Gamma}\} - \Sigma_{j=1}^{t} \lambda_j . \tag{2.6}$$

PROOF. A straight forward application of the Eckart-Young Theorem (see [10]). Some remarks regarding this theorem are necessary.

1. The matrix $\underset{\sim}{B}^{(t)}$ in (2.3) can be re-expressed as

$$\underset{\sim}{B} = \underset{\sim}{B}^{(t)} = \begin{bmatrix} \underset{\sim}{U}_1^{\tau} \\ \vdots \\ \underset{\sim}{U}_t^{\tau} \end{bmatrix} \underset{\sim}{\Sigma}_{XX}^{-\frac{1}{2}} \tag{2.7}$$

where $\underset{\sim}{U}_j$ is the latent vector corresponding to the j^{th} largest latent root, λ_j, of the matrix

$$\underset{\sim}{\Sigma}_{XX}^{-\frac{1}{2}} \underset{\sim}{\Sigma}_{XY} \underset{\sim}{\Gamma} \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-\frac{1}{2}} , \tag{2.8}$$

$j = 1,2,\ldots,r$. Since it has been assumed here that $s \leq r$, then $\lambda_j = 0$ for $j = s + 1,\ldots,r$.

2. The regression coefficient matrix $\underset{\sim}{C}$ in (1.4) with rank t is, therefore, given by

$$\underset{\sim}{C} = \underset{\sim}{C}^{(t)} = \underset{\sim}{\Gamma}^{-\frac{1}{2}} (\Sigma_{j=1}^{t} \underset{\sim}{V}_j \underset{\sim}{V}_j^{\tau}) \underset{\sim}{\Gamma}^{\frac{1}{2}} \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1}, \tag{2.9}$$

which, if $t = s$, reduces to the full-rank regression coefficient matrix, to be denoted henceforth by

$$\Theta = \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1} . \tag{2.10}$$

3. A principal components analysis of the r-vector variate $\underset{\sim}{X}$ corresponds to setting $\underset{\sim}{Y} = \underset{\sim}{X}$, $s = r$, and $\underset{\sim}{\Gamma} = \underset{\sim}{I}_r$ in the above theorem. This gives the following versions of (2.2) - (2.6):

$$\underset{\sim}{A}^{(t)} = [\underset{\sim}{V}_1, \ldots, \underset{\sim}{V}_t] = \underset{\sim}{B}^{(t)\tau}, \tag{2.11}$$

$$\underset{\sim}{C}^{(t)} = \Sigma_{j=1}^{t} \underset{\sim}{V}_j \underset{\sim}{V}_j^{\tau}, \qquad \underset{\sim}{\Theta} = \underset{\sim}{I}_r, \tag{2.12}$$

where $\underset{\sim}{V}_j$ is the latent vector corresponding to the j^{th} largest latent root, λ_j, of $\underset{\sim}{\Sigma}_{XX}$. The minimum value of the criterion (2.1) in this case is given by $\Sigma_{j=t+1}^{r} \lambda_j$, the sum of the residual $r - t$ latent roots of $\underset{\sim}{\Sigma}_{XX}$. The first t principal components of $\underset{\sim}{X}$ are given by the elements of the vector $\underset{\sim}{\xi}^{(t)} = \underset{\sim}{B}^{(t)}\underset{\sim}{X}$, (i.e., $\xi_j = \underset{\sim}{V}_j^{\tau} \underset{\sim}{X}$, $j = 1,2,\ldots,t$), where $\text{var}\{\xi_j\} = \lambda_j$ and $\text{cov}\{\xi_j,\xi_k\} = 0$ for $j \neq k$.

4. A canonical variate and correlation analysis of the two sets of variates, $\underset{\sim}{X}$ and $\underset{\sim}{Y}$, corresponds to setting $\underset{\sim}{\Gamma} = \underset{\sim}{\Sigma}_{YY}^{-1}$ in the above theorem, so that the minimum of (2.1) over choice of $\underset{\sim}{A}$, $\underset{\sim}{B}$ and $\underset{\sim}{\mu}$ is invariant under simultaneous linear transformations of the variates $\underset{\sim}{X}$ and $\underset{\sim}{Y}$. The reduced-rank regression coefficient matrix (2.9), therefore, becomes

$$\underset{\sim}{C} = \underset{\sim}{C}^{(t)} = \underset{\sim}{\Sigma}_{YY}^{\frac{1}{2}} (\Sigma_{j=1}^{t} \underset{\sim}{V}_j \underset{\sim}{V}_j^{\tau}) \underset{\sim}{\Sigma}_{YY}^{-\frac{1}{2}} \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1}, \tag{2.13}$$

where $\underset{\sim}{V}_j$ is the latent vector corresponding to the j^{th} largest latent root, λ_j, of the matrix

$$\underset{\sim}{R} = \underset{\sim}{\Sigma}_{YY}^{-\frac{1}{2}} \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1} \underset{\sim}{\Sigma}_{XY} \underset{\sim}{\Sigma}_{YY}^{-\frac{1}{2}} \tag{2.14}$$

The matrix, $\underset{\sim}{R}$, is a multivariate generalization of the simple squared correlation coefficient between two variables (r=s=1), and also of the squared multiple correlation coefficient between a single variable and a number of other variables (s=1, and any r). Set

$$\underset{\sim}{A}^{(t)-} = \begin{bmatrix} \underset{\sim}{V}_1^T \\ \vdots \\ \underset{\sim}{V}_t^T \end{bmatrix} \underset{\sim}{\Sigma}_{YY}^{-\frac{1}{2}} . \tag{2.15}$$

The matrix $\underset{\sim}{A}^{(t)-}$ is for $1 \leq t < s$, a reflexive generalized-inverse of $\underset{\sim}{A}^{(t)}$, and for t=s, $\underset{\sim}{A}^{(s)-}$ is the unique inverse $\{\underset{\sim}{A}^{(s)}\}^{-1}$ (see [20]). Note the symmetric relationship between $\underset{\sim}{B}^{(t)}$ (as given in (2.7)) and $\underset{\sim}{A}^{(t)-}$ (in (2.15)). The transformed variates

$$\underset{\sim}{\xi}^{(t)} = \underset{\sim}{B}^{(t)} \underset{\sim}{X} \quad \text{and} \quad \underset{\sim}{\omega}^{(t)} = \underset{\sim}{A}^{(t)-} \underset{\sim}{Y} . \tag{2.16}$$

have correlation matrix

$$\text{corr}\{\underset{\sim}{\xi}^{(t)}, \underset{\sim}{\omega}^{(t)}\} = \begin{bmatrix} \underset{\sim}{I}_t & \vdots & \underset{\sim}{P}_t \\ \cdots & \vdots & \cdots \\ \underset{\sim}{P}_t & \vdots & \underset{\sim}{I}_t \end{bmatrix}$$

where $\underset{\sim}{P}_t = \text{diag}\{\rho_1, \rho_2, \ldots, \rho_t\}$, $\rho_j = \lambda_j^{\frac{1}{2}}$, $j = 1,2,\ldots,t$, and the j^{th} components of both $\underset{\sim}{\xi}^{(t)}$ and $\underset{\sim}{\omega}^{(t)}$, namely

$$\xi_j = \underset{\sim}{U}_j^T \underset{\sim}{\Sigma}_{XX}^{-\frac{1}{2}} \underset{\sim}{X} \quad \text{and} \quad \omega_j = \underset{\sim}{V}_j^T \underset{\sim}{\Sigma}_{YY}^{-\frac{1}{2}} \underset{\sim}{Y},$$

respectively, are the j^{th} pair of canonical variates, and ρ_j, the correlation between them is the j^{th} canonical correlation coefficient ($j=1,2,\ldots,t$).

3. RESIDUALS FROM A REDUCED-RANK REGRESSION

Estimation of the vector and matrix quantities in Section 2 above is carried out using the sample of values (1.6). First, (1.2) is estimated by $\hat{\underset{\sim}{\mu}}_X = n^{-1}\Sigma_{j=1}^n \underset{\sim}{X}_j$ and $\hat{\underset{\sim}{\mu}}_Y = n^{-1}\Sigma_{j=1}^n \underset{\sim}{Y}_j$, and (1.3) by

$$(n-1)^{-1}\Sigma_{j=1}^n \begin{bmatrix} \underset{\sim}{X}_j - \hat{\underset{\sim}{\mu}}_X \\ \underset{\sim}{Y}_j - \hat{\underset{\sim}{\mu}}_Y \end{bmatrix} \begin{bmatrix} \underset{\sim}{X}_j - \hat{\underset{\sim}{\mu}}_X \\ \underset{\sim}{Y}_j - \hat{\underset{\sim}{\mu}}_Y \end{bmatrix}^\tau = \begin{bmatrix} \hat{\underset{\sim}{\Sigma}}_{XX} & \hat{\underset{\sim}{\Sigma}}_{XY} \\ \hat{\underset{\sim}{\Sigma}}_{YX} & \hat{\underset{\sim}{\Sigma}}_{YY} \end{bmatrix} = \hat{\underset{\sim}{\Sigma}} . \qquad (3.1)$$

All estimates of unknowns are then based on the appropriate elements of (3.1), and denoted by placing a circumflex above the quantity to be estimated. In this way, we denote $\hat{\underset{\sim}{\mu}}^{(t)}$ to be an estimate of (2.4) and $\hat{\underset{\sim}{C}}^{(t)}$ to be an estimate of (2.9).

The collection of n residual s-vectors from a rank t reduced-rank regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X}$ is given by the matrix

$$\hat{\underset{\sim}{\varepsilon}}^{(t)} = [\underset{\sim}{\varepsilon}_1^{(t)},\ldots,\underset{\sim}{\varepsilon}_n^{(t)}], \qquad (3.2)$$

where

$$\hat{\underset{\sim}{\varepsilon}}_j^{(t)} = \underset{\sim}{Y}_j - \hat{\underset{\sim}{\mu}}^{(t)} - \hat{\underset{\sim}{C}}^{(t)}\underset{\sim}{X}_j = (\underset{\sim}{Y}_j - \hat{\underset{\sim}{\mu}}_Y) - \hat{\underset{\sim}{C}}^{(t)}(\underset{\sim}{X}_j - \hat{\underset{\sim}{\mu}}_X),$$

$$j=1,2,\ldots,n. \qquad (3.3)$$

The columns of the matrix $\hat{\underset{\sim}{\varepsilon}}^{(t)}$ in (3.2) are each asymptotically (large n) s-variate normally distributed with mean zero and covariance matrix

$$\underset{\sim}{\Sigma}_{YY} - \underset{\sim}{\Gamma}^{-\frac{1}{2}}(\Sigma_{j=1}^{t} \lambda_j \underset{\sim}{V}_j \underset{\sim}{V}_j^{\tau})\underset{\sim}{\Gamma}^{-\frac{1}{2}}, \tag{3.4}$$

where λ_j and V_j are the j^{th} latent root and vector respectively of (2.5), if indeed the variate (1.1) is (r + s)-variate normally distributed. Furthermore, the columns of (3.2) are asymptotically pairwise uncorrelated. (These results can be obtained using perturbation expansions as in [13].) For the full-rank case, the corresponding set of residual vectors are each asymptotically jointly normal with mean zero and covariance matrix

$$\underset{\sim}{\Sigma}_{YY} - \underset{\sim}{\Sigma}_{YX} \underset{\sim}{\Sigma}_{XX}^{-1} \underset{\sim}{\Sigma}_{XY}.$$

In view of these remarks, we estimate the residual covariance matrix $\underset{\sim}{\Sigma}_{\varepsilon\varepsilon}^{(t)}$, by

$$\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}^{(t)} = (n-1-r)^{-1} \Sigma_{j=1}^{n} \hat{\underset{\sim}{\varepsilon}}_j^{(t)} \hat{\underset{\sim}{\varepsilon}}_j^{(t)\tau}, \tag{3.5}$$

and we write, for the full-rank case only (where t=s), $\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}^{(s)} = \hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}$.

4. THE CASE OF UNKNOWN RANK

Discussion so far has centered around the case in which the regression coefficient matrix $\underset{\sim}{C}$ has a specific rank, t say. The remainder of this paper is concerned with the case in which the value of t is unknown à priori and has to be assessed from the sample data (1.6).

It was pointed out in Section 2 of this paper that a reduced-rank regression of rank t corresponds to a choice of either the first t principal components of $\underset{\sim}{X}$ or of the first t pairs of canonical variates of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$. Recall that W(t) denotes the minimum value of (2.1) for a fixed value of t. The reduction in W(t) obtained by increasing the rank from $t=t_0$ to $t=t_1$ $(t_0<t_1)$ is, therefore, given by

$$W(t_0) - W(t_1) = \Sigma_{j=t_0+1}^{t_1} \lambda_j \ , \tag{4.1}$$

where λ_j is the j^{th} largest latent root of (2.5). That is, one method for assessing the rank of $\underset{\sim}{C}$ can be based either on the sequence of ordered latent roots, $\{\hat{\lambda}_j, j=1,2,\ldots,s\}$, in which $\hat{\lambda}_j$ is compared with suitable reference values for each j, or on the sum of the $(s-t_0)$ residual latent roots (see, e.g., Kshirsagar [14, sections 8.7 and 11.7]).

An obvious disadvantage of relying solely on such formal testing procedures is that any routine application of them might fail to take into account the possible need for a preliminary screening of the multidimensional data set in question. Robustness of sample estimates of the latent roots and hence of the various tests when outliers or distributional peculiarities are present in the data is a major statistical problem. In the context of this paper, disregard for such details could lead to incorrect inferences regarding the dimensionality of the regression.

5. A SIMPLE EXAMPLE

The data for this example was taken from Rao [15, p. 245], where it is attributed to Frets [11]. The same data appears in Anderson [2, p.58], Dixon [7, p. 212], and various other places for illustrating certain statistical procedures. While the original investigation consisted of about 3600 measurements on 360 families, this particular subset consists of two measurements, head-length and head-breadth, on each of the first and second sons in a sample of 25 families. Thus, X_1 (X_2) is the head-length (head-breadth) of the first son, and Y_1 (Y_2) is the head-length (head-breadth) of the second son.

Estimates of the mean vector and covariance matrix of these four variables can be found on p. 303 of Anderson [2] and will not be repeated here. Two regressions were made on the data, a reduced-rank regression (t=1) and a full-rank regression (t=2), using the canonical variates set up. The results are as follows:

$$\hat{\underset{\sim}{C}}^{(1)} = \begin{bmatrix} 0.43 & 0.54 \\ 0.29 & 0.36 \end{bmatrix}, \quad \hat{\underset{\sim}{C}}^{(2)} = \begin{bmatrix} 0.45 & 0.51 \\ 0.27 & 0.38 \end{bmatrix},$$

$$\hat{\underset{\sim}{\mu}}^{(1)} = \begin{bmatrix} 23.39 \\ 41.40 \end{bmatrix}, \quad \hat{\underset{\sim}{\mu}}^{(2)} = \begin{bmatrix} 23.74 \\ 37.17 \end{bmatrix}.$$

An initial inspection of these results shows that the matrices $\hat{\underset{\sim}{C}}^{(1)}$ and $\hat{\underset{\sim}{C}}^{(2)}$ are not very different from each other.

Further detailed study of this data indicated certain unexplainable peculiarities. The 25 observations were checked against the complete collection in Frets [11]. It was found that <u>six</u> of these 25 observations were different from those in the original source data. The incorrect

observation numbers are 11, 12, 13, 14, 15, and 25. From a close examination of the original source data tables it appears that the values of the first five of these errors were taken from the wrong columns of the original data; the sixth appears to be an independent error. Both sets of data are given here in Table 5.1.

The 'corrected' sample was analyzed in a similar manner as was the previous set; the results are summarized as follows:

$$\hat{\underset{\sim}{\mu}}_X = \begin{bmatrix} 187.40 \\ 151.12 \end{bmatrix}, \quad \hat{\underset{\sim}{\mu}}_Y = \begin{bmatrix} 183.32 \\ 149.36 \end{bmatrix}$$

$$\hat{\underset{\sim}{\Sigma}} = \left[\begin{array}{cc:cc} 102.83 & 59.62 & 70.32 & 52.68 \\ (0.82) & 51.86 & 44.25 & 40.21 \\ \hdashline (0.70) & (0.62) & 97.98 & 51.71 \\ (0.76) & (0.82) & (0.77) & 46.24 \end{array}\right]$$

(values in parentheses are correlations between appropriate variables)

$$\hat{\underset{\sim}{C}}^{(1)} = \begin{bmatrix} 0.25 & 0.62 \\ 0.21 & 0.53 \end{bmatrix}, \quad \hat{\underset{\sim}{C}}^{(2)} = \begin{bmatrix} 0.57 & 0.20 \\ 0.19 & 0.56 \end{bmatrix}$$

$$\hat{\underset{\sim}{\mu}}^{(1)} = \begin{bmatrix} 42.18 \\ 29.99 \end{bmatrix}, \quad \hat{\underset{\sim}{\mu}}^{(2)} = \begin{bmatrix} 46.61 \\ 29.62 \end{bmatrix}.$$

This time the estimates of $\underset{\sim}{C}$ for t = 1 and t = 2 look very different from each other.

6. THE RANK TRACE

We now propose a more elaborate method for assessing the dimensionality of a multivariate regression. It is described in terms of the following steps:

1. Carry out a sequence of reduced-rank regressions for specific values of t.
2. From each of the regressions of step (1), compute $\hat{\underset{\sim}{C}}(t)$ and $\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}(t)$.
3. Make a scatterplot of the s+1 points

$$(\Delta\hat{\underset{\sim}{C}}(t),\ \Delta\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}(t)),\ t=0,1,2,\ldots,s, \tag{6.1}$$

where

$$\Delta\hat{\underset{\sim}{C}}(t) = \frac{\|\hat{\underset{\sim}{C}}(t) - \hat{\underset{\sim}{\Theta}}\|}{\|\hat{\underset{\sim}{\Theta}}\|} \tag{6.2}$$

$$\Delta\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}(t) = \frac{\|\hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}(t) - \hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}\|}{\|\hat{\underset{\sim}{\Sigma}}_{YY} - \hat{\underset{\sim}{\Sigma}}_{\varepsilon\varepsilon}\|}, \tag{6.3}$$

and join up successive points in the plot. This is the <u>rank trace</u> for the regression of $\underset{\sim}{Y}$ on $\underset{\sim}{X}$.

4. Assess the rank of C as the <u>smallest</u> rank for which <u>both</u> (6.2) and (6.3) are approximately zero.

In small problems (where the value of s is at most 10), all values of t should be examined. For larger problems (s > 10), the costs of computation become critical and it is, therefore, recommended to be more selective in

choices of t; one possible way is to carry out regressions for a few small values of t, say $t=1,2,\ldots,\hat{t}_0+1$, where $\hat{t}_0$ might be an initial estimate of t (perhaps based on the sequence of sample latent roots) plus the usual full-rank regression model (in which t=s) for purposes of comparison.

For the examples in this paper, the classical Euclidean norm

$$\| \underset{\sim}{A} \| = (\mathrm{tr}\underset{\sim\sim}{A}\underset{\sim\sim}{A}^{T})^{\frac{1}{2}} = (\Sigma_i \Sigma_j a_{ij}^2)^{\frac{1}{2}}$$

is used in computing (6.2) and (6.3). For the case when t=0, define $\hat{\underset{\sim}{C}}^{(0)}$ to be the null matrix with all entries equal to zero, and $\hat{\underset{\sim\varepsilon\varepsilon}{\Sigma}}^{(0)}$ to be $\hat{\underset{\sim YY}{\Sigma}}$.

Thus, the first point (corresponding to t=0) is always plotted at (1,1) and the last point (corresponding to t=s) is always plotted at (0,0). The horizontal coordinate (6.2) gives a quantitative representation of the difference between a reduced-rank regression coefficient matrix and the full-rank regression coefficient matrix, while the vertical coordinate (6.3) shows the proportionate reduction in the residual variance matrix in using a simple full-rank model rather than the computationally more elaborate reduced-rank model. The reason for including a special point for t=0 is that without such a point, it would be impossible to deduce in many applications that the statistical rank of $\underset{\sim}{C}$ should be t=1. In this formulation, t=0 corresponds to the completely random model, $\underset{\sim}{Y} = \underset{\sim}{\mu} + \underset{\sim}{\varepsilon}$.

Assessing the dimensionality of the regression by using rule (4) above involves a certain amount of subjective judgment, but from experience with many of these types of plots, the choice should not be too difficult. Due to the nature of $\hat{\underset{\sim}{C}}^{(t)}$, the sequence of values for the horizontal coordinate (6.2) is not guaranteed to decrease monotonically from 1 to 0. It does

appear, however, that in many of the applications of this method (in particular, for the canonical variates case), the plotted points appear within the unit square, but below the (1,1) - (0,0) diagonal-line, indicating that the residual variance matrices typically stabilize faster than do the regression coefficient matrices.

For the principal components case, the expressions (6.2) and (6.3) reduce to the following simple forms:

$$\Delta\hat{C}^{(t)}_{\sim} = (1 - \frac{t}{r})^{\frac{1}{2}} \tag{6.4}$$

$$\Delta\hat{\Sigma}^{(t)}_{\sim\varepsilon\varepsilon} = \{(\Sigma_{j=t+1}^{r}\hat{\lambda}_j^2)/(\Sigma_{j=1}^{r}\hat{\lambda}_j^2)\}^{\frac{1}{2}}. \tag{6.5}$$

It is clear from (6.4) and (6.5) that: (a) we are really looking at the residual latent roots again (although this time they are each squared); (b) all the information regarding dimensionality of the regression is contained in the residual covariance matrices and not in the regression coefficients; and (c) the r + 1 plotted points do indeed decrease monotonically from (1,1) to (0,0) in this special case. (Unfortunately, no similar reduction of (6.2) and (6.3) can be obtained for the canonical variates case.)

In view of (6.4), a different criterion of assessing dimensionality from the rank trace plot in the principal components case needs to be applied. A natural rule (which has also been proposed for obtaining multidimensional scaling solutions; see, e.g., Gnanadesikan [12, p. 46]) is that of assessing the rank of $\underset{\sim}{C}$ by the smallest integer value between 1 and r at which an "elbow" can be detected in the PC rank trace plot.

Example 1 (continued). The CV rank trace of the data from Rao [15] is plotted in Figure 6.1(a). From the plot it appears that the rank of

$\underset{\sim}{C}$ is best estimated by $\hat{t}=1$, which also seems reasonable on the basis of the canonical correlations, 0.7885 and 0.0537. The 'corrected' data, however, yield a very different result with the CV rank trace plotted in Figure 6.1 (b). The plots suggest that the estimated rank of $\underset{\sim}{C}$ should be $\hat{t}=2$ (the canonical correlations are 0.8386 and 0.3256). A third analysis (not shown here) was made on the complete data in Frets [11] on the same four variables. (The set of extensive data tables was screened very carefully since cross-referencing of observations there often proved inconsistent; this meant that the data were boiled down to 247 points.) This larger set (which also contained the 'corrected' 25 values) gave a similar plot of the CV rank trace to the 'corrected' data, again suggesting that the rank estimate should be $\hat{t}=2$. The sample canonical correlations, 0.6588 and 0.5077, appear to reflect the same information.

Example 2. U.S. and European Temperature Records. These data, which were made available by J. M. Craddock (Meterological Office, Bracknell, Herts., England), the World Weather Records, Smithsonian Miscellaneous Collections, and the U.S. Weather Bureau, consist of 516 mean monthly temperatures (1918-1960) for five U.S. cities (New Haven, Cape Hatteras, Cincinnati, Nashville, and St. Louis) and for five European cities (Copenhagen, DeBilt, Paris, Odessa, and Valentia). Before analysis, the series for each city was seasonally adjusted by subtracting out the mean for each of the 12 months of the year. The European cities were treated as the $\underset{\sim}{X}$ variables and the U.S. cities as the $\underset{\sim}{Y}$ variables, so that r=s=5 and n=516. As in the previous example, all points of the CV rank trace plot (see Figure 6.2) are interior to the unit square, and the rank of $\underset{\sim}{C}$ is assessed at $\hat{t}=3$. It is worth noting that a formal χ^2 test for the significance of

the residual canonical correlations gives only the first two canonical correlations as being non zero. The five correlations are 0.3836, 0.2746, 0.1340, 0.0507, and 0.0090, and the CV rank trace plot has, therefore, yielded additional information for this example.

Example 3. L.A. Heart Study Data. These data were taken from Dixon and Massey [9, pp. 14-17], and consist of measurements on 200 men who were survivors of a group having had an initial examination in 1952 and who were re-examined in 1962. For this example, the variables in [9] were divided into two groups: a set of r=6 (1952) $\underset{\sim}{X}$ variables (i.e., A, C, D, G, I, and J) and a set of s=4 (1962) $\underset{\sim}{Y}$ variables (i.e., E, F, H, and K). Only those cases with L=0 were used here, where L is coded 1 (or 0) if a coronary incident occurred (or, did not occur) between 1952 and 1962; this reduced the size of the sample to n=174. The plot of the CV rank trace (see Figure 6.3) yields an exterior value for the reduced-rank regression of rank one; all other points are interior to the unit square. The rank of $\underset{\sim}{C}$ is assessed at $\hat{t}=3$, which agrees with the appropriate χ^2 test for significance of the canonical correlations (namely, 0.6704, 0.5443, 0.4790, and 0.0932).

Example 4. Fisher's Iris Data. This is a classical data set of n=50 measurements on the r=4 variables, sepal length, sepal width, petal length, and petal width, of the species Iris Versicolor. See, e.g., Anderson [2, Section 11.5]. The PC rank trace plot is given in Figure 6.4, and the rank is assessed as $\hat{t}=1$. The corresponding latent roots are 0.4879, 0.0724, 0.0548, and 0.0098.

Example 5. Jarvik Smoking Questionnaire Data. These data were taken from Dixon and Brown [8, p. 624]. They refer to r=12 answers to a smoking questionnaire administered to n=110 subjects. Each question was coded 1 to 5 such that a high score represents a desire to smoke. The PC rank trace plot (see Figure 6.5) shows an "elbow" at $\hat{t}$=3. The latent roots for this example (which were calculated from the (12 × 12)-correlation matrix of the data, are given by 5.426, 2.997, 1.361, 0.560, 0.363, 0.302, 0.241, 0.200, 0.158, 0.146, 0.137, and 0.110.

7. COMPARING GAMMA PLOTS OF MULTIVARIATE RESIDUALS

The methodology proposed in Sections 4 and 6 for assessing the rank of the regression coefficient matrix used various summary measures resulting from each regression, namely the set of latent roots, the sequence of regression coefficient matrices themselves, and their corresponding residual variance matrices. The purpose of this section is to describe an additional method using the set of multivariate residuals from a series of reduced-rank regressions.

The residuals are the n s-dimensional vectors $\hat{\underset{\sim}{\varepsilon}}_j^{(t)}$, j=1,2,...,n, obtained from a reduced-rank regression of rank t. See Section 3 above. One method of comparing these vectors simultaneously is to construct a quadratic form for each vector in which the choice of 'compounding' matrix is positive-definite, but otherwise arbitrary. The n derived quadratic forms (for a given compounding matrix) may then be compared with each other (for example, through a linear ordering of their values).

If $\underset{\sim}{M}$ is some positive-definite matrix, the quadratic form

$$f^{(n)}(\underset{\sim}{\varepsilon}_j^{(t)}) = \hat{\underset{\sim}{\varepsilon}}_j^{(t)\tau} \underset{\sim}{M} \hat{\underset{\sim}{\varepsilon}}_j^{(t)}, \tag{7.1}$$

converges in distribution to the random variable

$$f(\underset{\sim}{\varepsilon}_j^{(t)}) = \underset{\sim}{\varepsilon}_j^{(t)\tau} \underset{\sim}{M} \underset{\sim}{\varepsilon}_j^{(t)}, \tag{7.2}$$

with distribution

$$\Sigma_{k=1}^{s} \mu_k^{(t)} \chi_{1,k}^2, \tag{7.3}$$

where $\{\mu_k^{(t)}\}$ are the latent roots of the matrix $\underset{\sim}{\Sigma}_{\varepsilon\varepsilon}^{(t)}\underset{\sim}{M}$ and χ_1^2 denotes an independent chi-squared variate having one degree of freedom (Box [4, Theorem 2.1]. In the special case when $\underset{\sim}{M} = \{\underset{\sim}{\Sigma}_{\varepsilon\varepsilon}^{(t)}\}^{-1}$, (7.1) converges in distribution to a central chi-squared variate with s degrees of freedom.

The distribution (7.3) is approximated here by a gamma distribution with density

$$g(x;\ \lambda,\eta) = \lambda^{\eta}x^{\eta-1}e^{-\lambda x}/\Gamma(\eta),\ (x>0,\ \lambda>0,\ \eta>0), \tag{7.4}$$

where $\lambda = \lambda^{(t)}$ is the unknown scale parameter and $\eta = \eta^{(t)}$ is the unknown shape parameter, both depending on the value of t. Estimation of $\lambda^{(t)}$ and $\eta^{(t)}$ is carried out by the method of maximum likelihood using the first K order-statistics of the n values,

$$f^{(n)}(\underset{\sim}{\varepsilon}_1^{(t)}),\ f^{(n)}(\underset{\sim}{\varepsilon}_2^{(t)}),\ldots,\ f^{(n)}(\underset{\sim}{\varepsilon}_n^{(t)}). \tag{7.5}$$

The details may be found in Wilk et al [23].

Following the estimation of $\lambda^{(t)}$ and $\eta^{(t)}$ in (7.4), gamma probability plots are prepared in the manner of Wilk et al. [22] using gamma quantiles computed from the estimates $\hat{\lambda}^{(t)}$ and $\hat{\eta}^{(t)}$. Such a plot should resemble a straight-line configuration whenever the values (7.5) are from the estimated gamma distribution. If several of the largest values of (7.5) appear too large, or if a certain degree of 'curvature' is visible in the plot, then the assumption that all the values in (7.5) are gamma distributed may be invalid.

For purposes of comparing several reduced-rank regressions (each having a different rank), it is important that the same number K, of smallest order-statistics be nominated for each value of t. Revised gamma plots omitting any 'overly-large' values might be made to check better agreement of the model to the remaining data.

As long as the statistical rank of the regression coefficient matrix is larger than those values of t being considered, the corresponding gamma plots should differ markedly for different values of t. When t reaches this rank, the plots should cease to change significantly and should settle down. The characteristics of these gamma plots that yield information regarding degree of stability are:

(1) the sequence of estimates of (λ,η), namely, $(\hat{\lambda}^{(1)}, \hat{\eta}^{(1)}),\ldots, (\hat{\lambda}^{(s)}, \hat{\eta}^{(s)})$; and

(2) the general 'shape' of the plots.

For a given value of t, the gamma plot indicates the presence of outliers and any distributional peculiarities that may exist in the data. On the other hand, a comparative analysis of gamma plots for different values of t will help to assess the value of t. For the latter type of analysis, the shape of each plot is, therefore, important only in so far as it allows two or more plots to be compared with each other. Choices of $\underset{\sim}{M}$ include the $(s \times s)$-identity matrix $\underset{\sim}{I}_s$, and $\{\hat{\underset{\sim}{\Sigma}}^{(t)}_{\varepsilon\varepsilon}\}^{-1}$, for t=1,2,...,s. Results so far indicate that the estimates $(\hat{\lambda}^{(t)}, \hat{\eta}^{(t)})$ are much smaller when $\underset{\sim}{I}_s$ is used as a compounding matrix than when $\{\hat{\underset{\sim}{\Sigma}}^{(t)}_{\varepsilon\varepsilon}\}^{-1}$ is used.

Example 3 (continued). The gamma plots for the multivariate residuals are shown in Figures 7.1 ($\underset{\sim}{M} = \underset{\sim}{I}_4$) and 7.2 ($\underset{\sim}{M} = \{\hat{\underset{\sim}{\Sigma}}^{(t)}_{\varepsilon\varepsilon}\}^{-1}$, t=1,2,3,4); for these plots, the smallest K=150 order-statistics of (7.5) were used to estimate the gamma quantiles. Internal features of these plots show a lack of near-zero values and a possible outlier, as well as evidence of non-normality in the residuals. However, comparisons of the plots over the four values of t reveal that the configurations of the quadratic forms in the residuals change much less markedly following t=3 than for any previous value of t, again suggesting that $\hat{t}$=3.

TABLE 5.1

CORRECTED DATA FOR EXAMPLE 1; r=s=2, n=25
(*ASTERISKS DENOTE INCORRECT OBSERVATIONS USED BY RAO, ANDERSON AND DIXON.)

	Head Length, First Son X_1	Head Breadth, First Son X_2	Head Length, Second Son Y_1	Head Breadth, Second Son Y_2
1	191	155	179	145
2	195	149	201	152
3	181	148	185	149
4	183	153	188	149
5	176	144	171	142
6	208	157	192	152
7	189	150	190	149
8	197	159	189	152
9	188	152	197	159
10	192	150	187	151
11	186 (179*)	161 (158*)	179 (186*)	158 (148*)
12	179 (183*)	147 (147*)	183 (174*)	147 (147*)
13	195 (174*)	153 (150*)	174 (185*)	150 (152*)
14	202 (190*)	160 (159*)	190 (195*)	159 (157*)
15	194 (188*)	154 (151*)	188 (187*)	151 (158*)
16	163	137	161	130
17	195	155	183	158
18	186	153	173	148
19	181	145	182	146
20	175	140	165	137
21	192	154	185	152
22	174	143	178	147
23	176	139	176	143
24	197	167	200	158
25	190	153 (163*)	187	150

LEGENDS FOR FIGURES

FIGURE 6.1

Plot of CV rank trace for example 1, (a) Rao's data, and (b) corrected data, on heredity of headform in man ($r = s = 2$, $n = 25$).

FIGURE 6.2

Plot of CV rank trace for example 2 on U.S. and European temperature records ($r = s = 5$, $n = 516$).

FIGURE 6.3

Plot of CV rank trace for example 3 on L.A. heart study data ($r = 6$, $s = 4$, $n = 174$).

FIGURE 6.4

Plot of PC rank trace for example 4 on Fisher's iris versicolor data ($r = 4$, $n = 50$).

FIGURE 6.5

Plot of PC rank trace for example 5 on Jarvik's smoking questionnaire data ($r = 12$, $n = 110$).

FIGURE 7.1

Gamma probability plots of observed residuals for example 3.

FIGURE 7.2

Gamma probability plots of observed residuals for example 3.

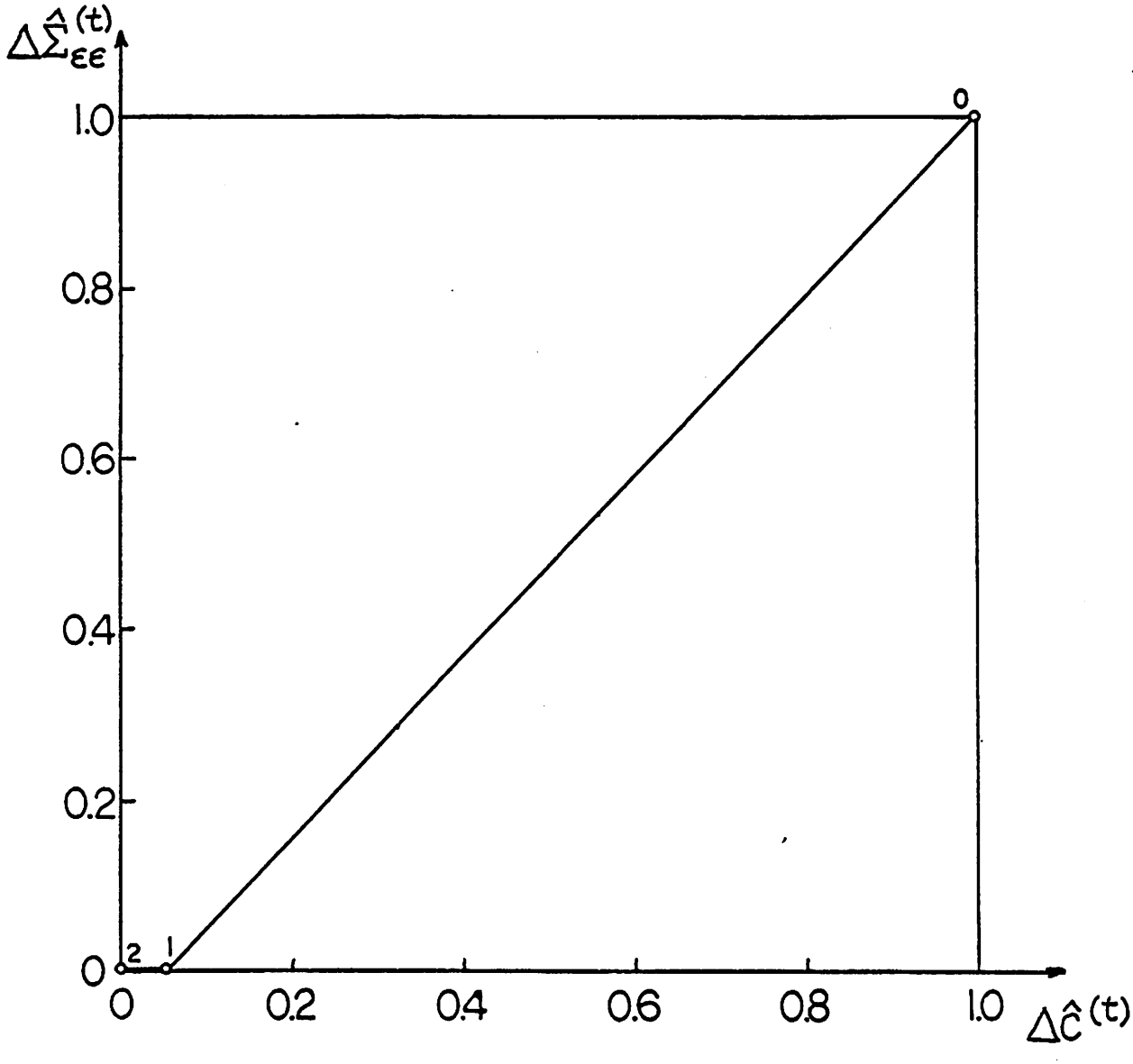
$\Delta\hat{\Sigma}_{\epsilon\epsilon}(t)$
1.0
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{C}(t)$
0
1
2

FIG. 6.1 (a)

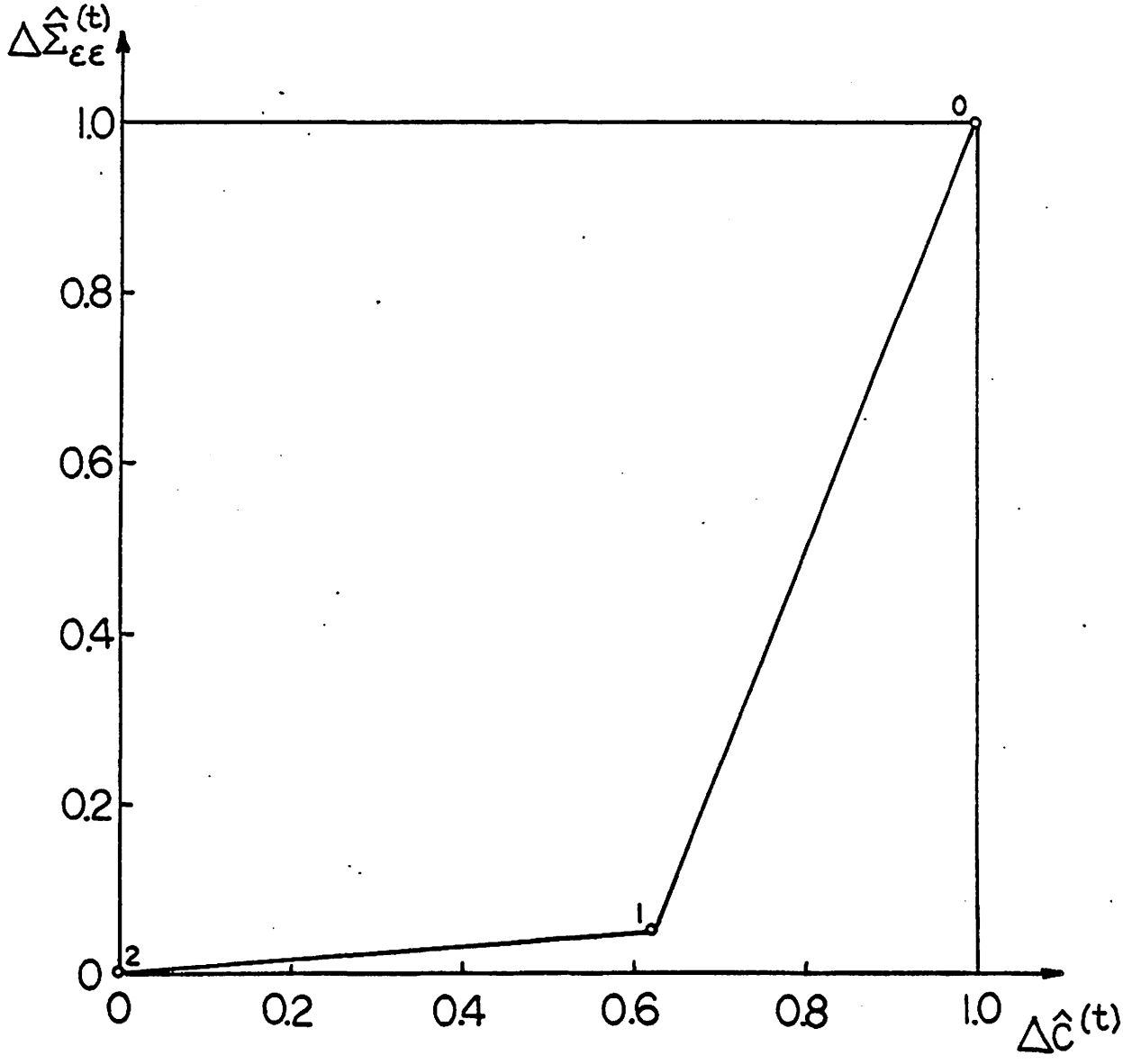
$\Delta\hat{\Sigma}_{\varepsilon\varepsilon}^{(t)}$
1.0
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{C}^{(t)}$
0
1
2

FIG. 6.1 (b)

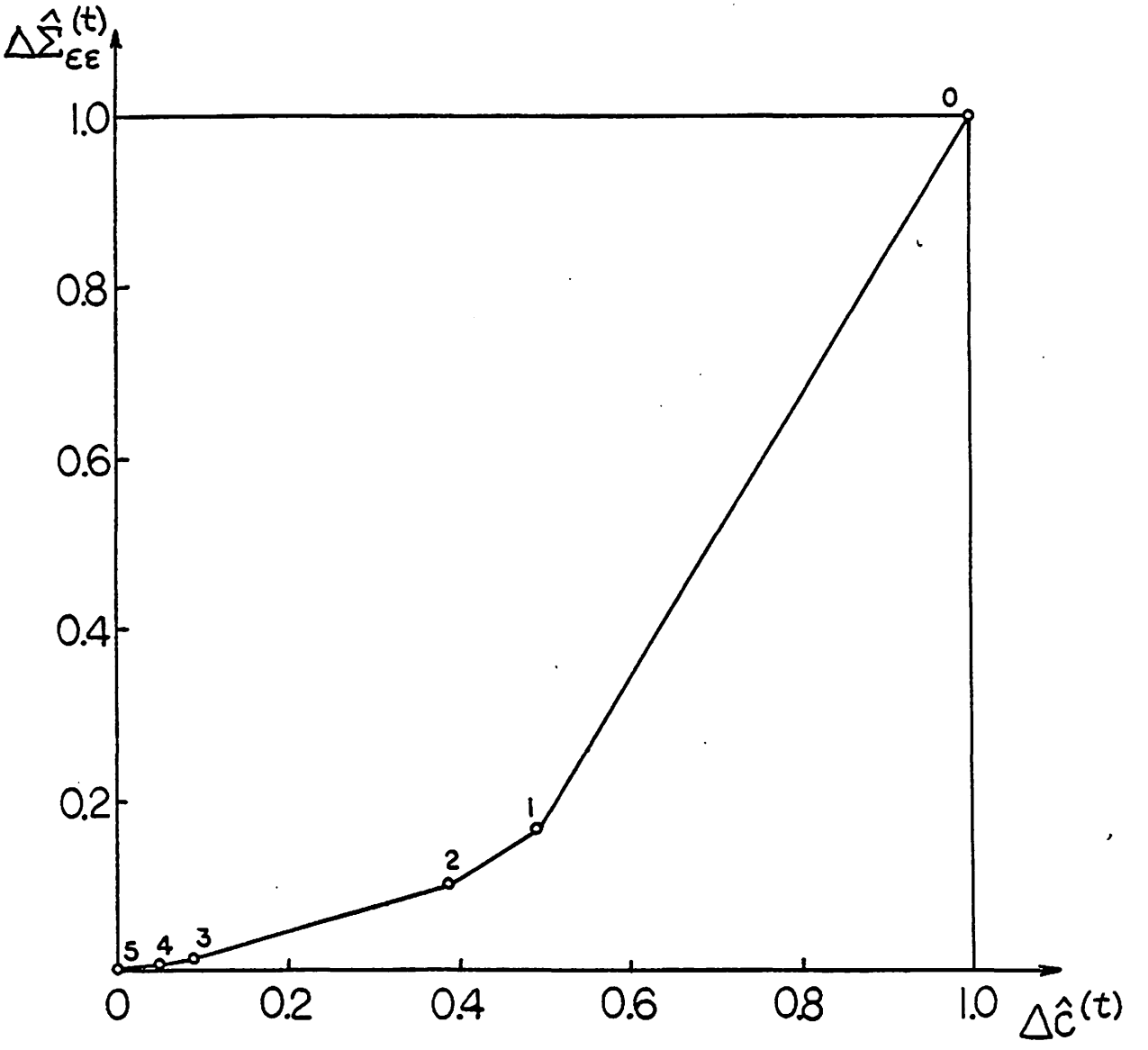
$\Delta\hat{\Sigma}_{\epsilon\epsilon}(t)$
1.0
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{c}(t)$
0
1
2
3
4
5

FIG. 6.2

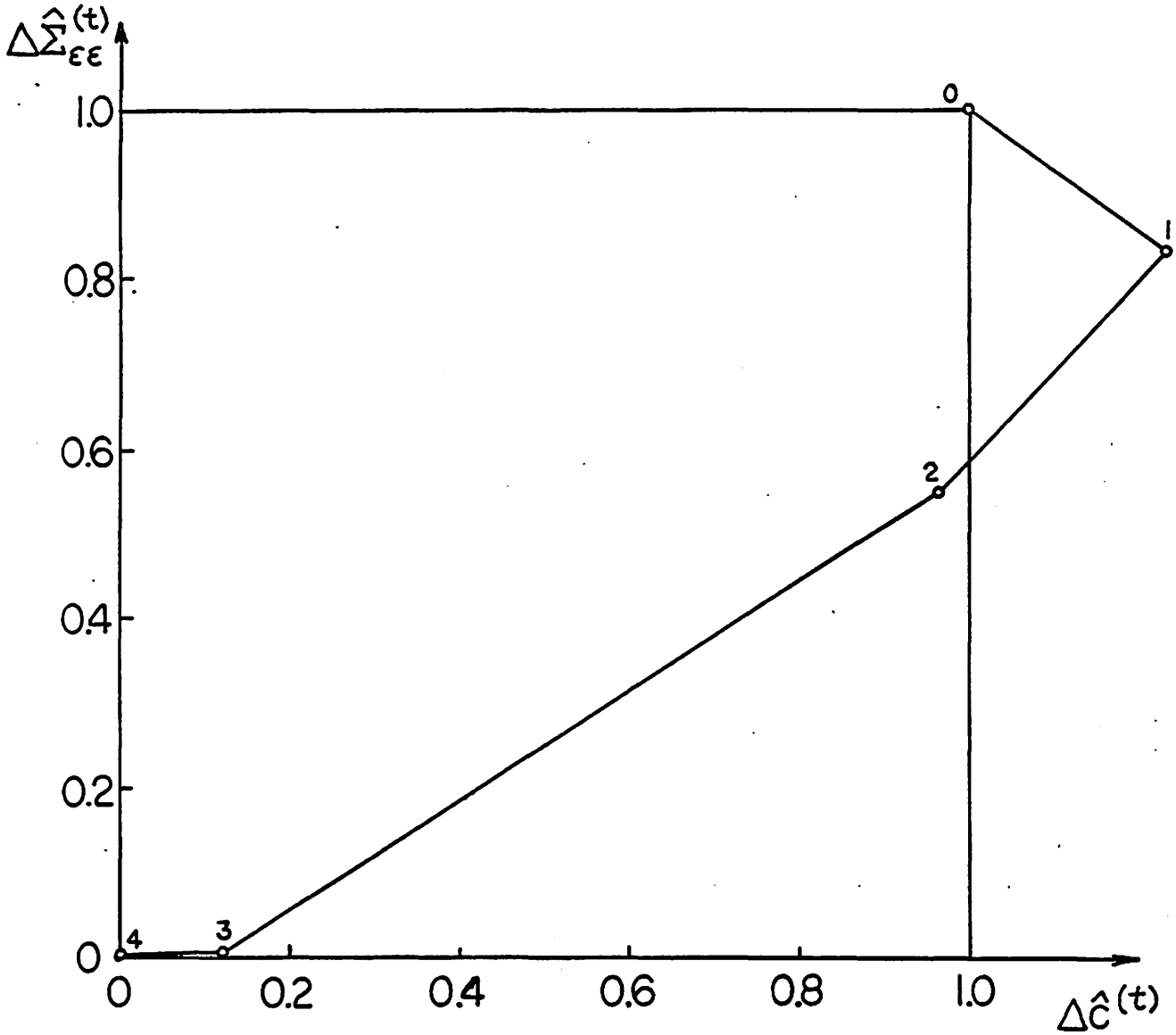
$\Delta\hat{\Sigma}^{(t)}_{\varepsilon\varepsilon}$
1.0
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{C}^{(t)}$
0
1
2
3
4

FIG. 6.3

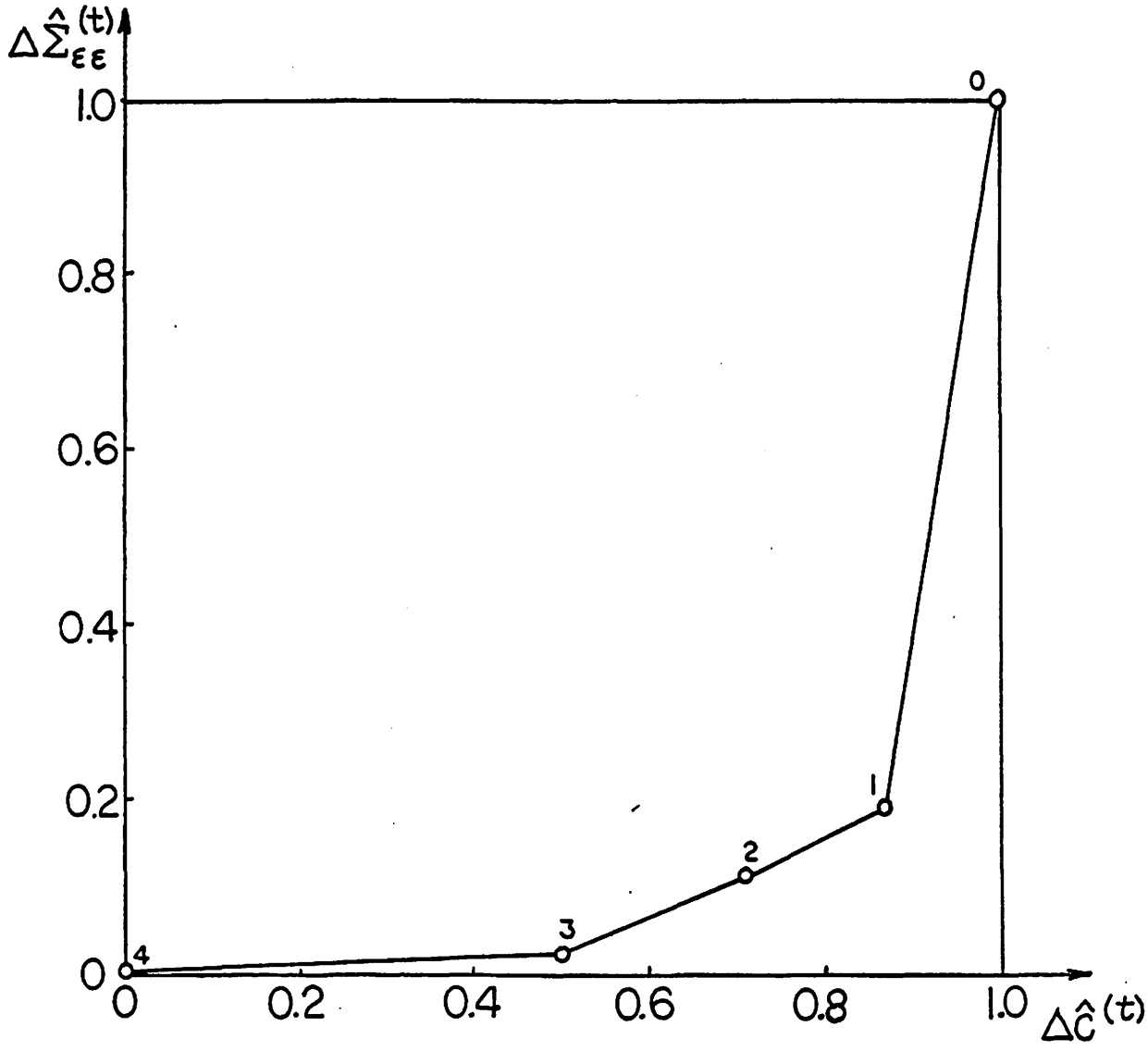
$\Delta\hat{\Sigma}_{\varepsilon\varepsilon}(t)$
1.0
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{C}(t)$
0
1
2
3
4

FIG. 6.4

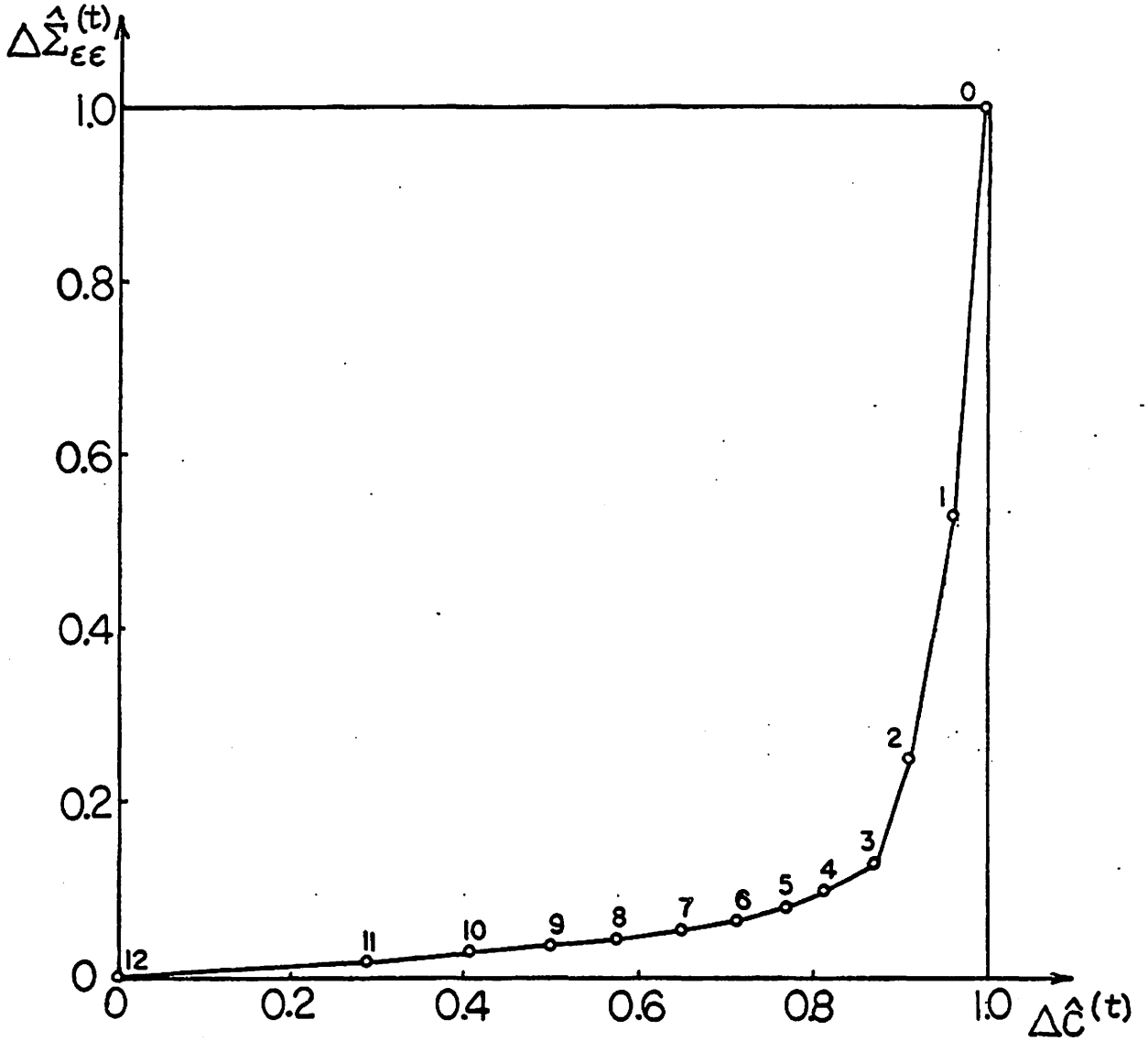
$\Delta\hat{\Sigma}_{\epsilon\epsilon}(t)$
1.0
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1.0
$\Delta\hat{c}(t)$
0
1
2
3
4
5
6
7
8
9
10
11
12

FIG. 6.5

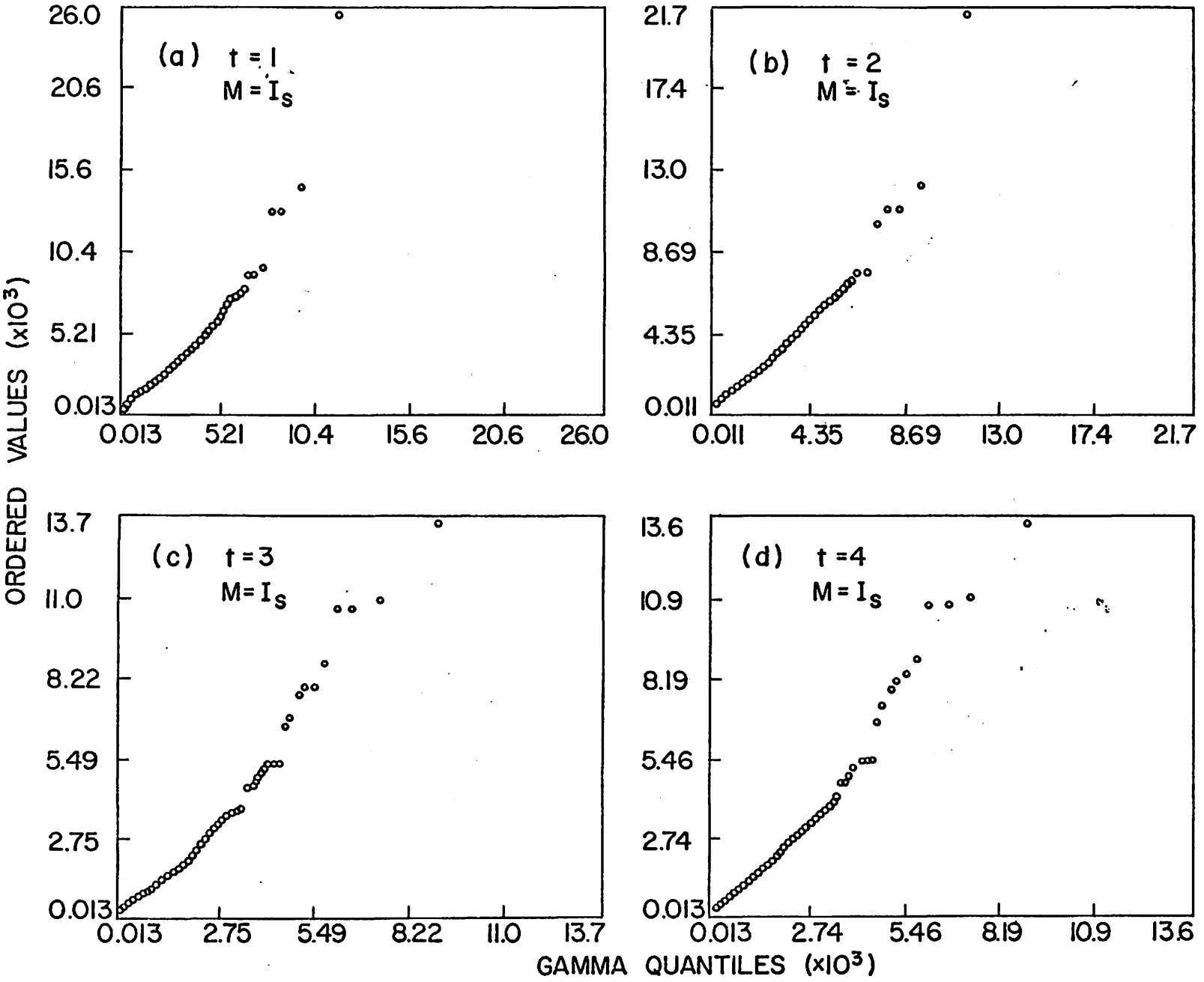
(a) t = 1
M = I_S
(b) t = 2
M = I_S
(c) t = 3
M = I_S
(d) t = 4
M = I_S
ORDERED VALUES (x10^3)
GAMMA QUANTILES (x10^3)
26.0
20.6
15.6
10.4
5.21
0.013
21.7
17.4
13.0
8.69
4.35
0.011
13.7
11.0
8.22
5.49
2.75
0.013
13.6
10.9
8.19
5.46
2.74
0.013

FIG. 7.1

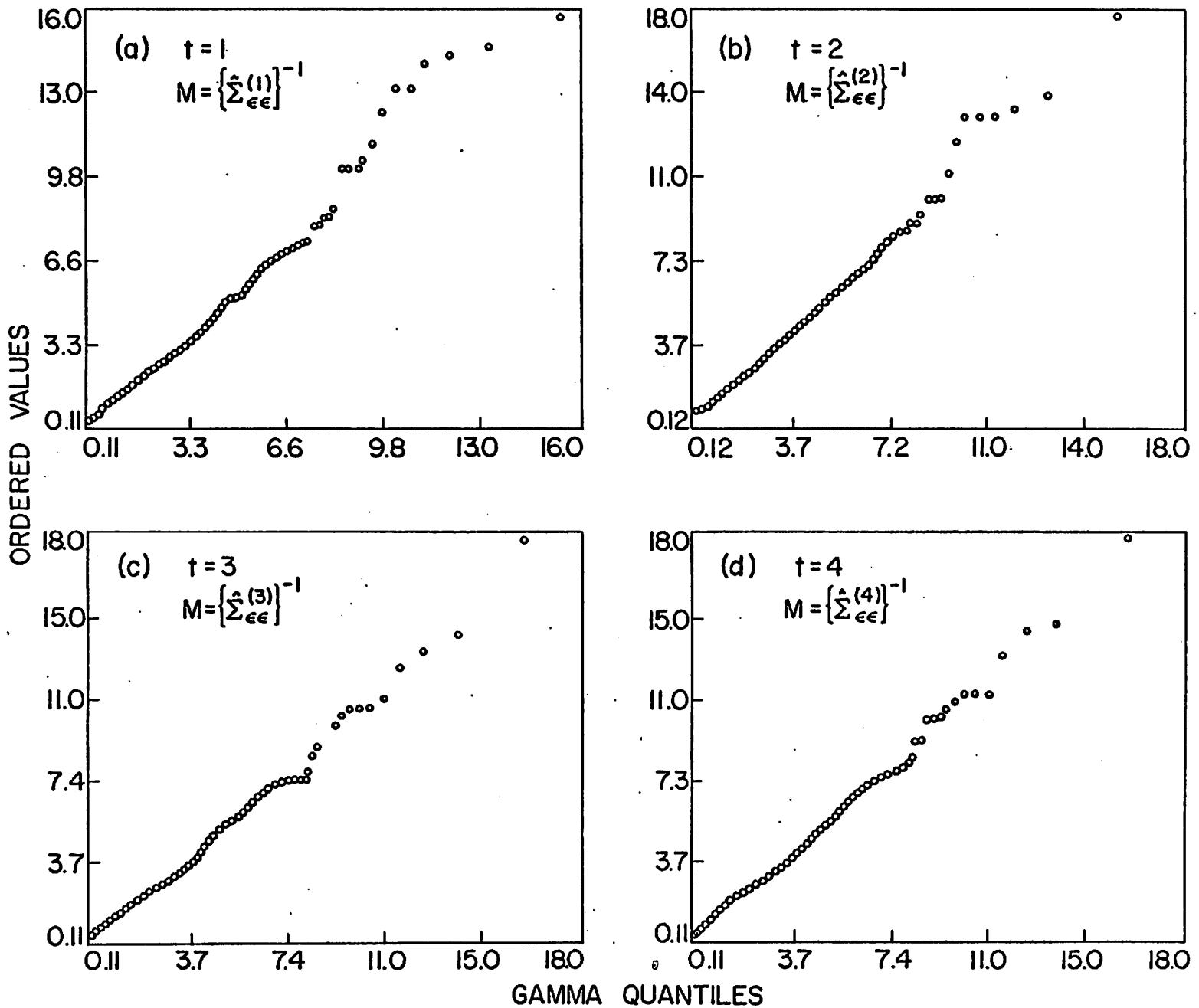
(a) t = 1
$M = \{\hat{\Sigma}^{(1)}_{\epsilon\epsilon}\}^{-1}$
16.0
13.0
9.8
6.6
3.3
0.11
0.11 3.3 6.6 9.8 13.0 16.0
(b) t = 2
$M = \{\hat{\Sigma}^{(2)}_{\epsilon\epsilon}\}^{-1}$
18.0
14.0
11.0
7.3
3.7
0.12
0.12 3.7 7.2 11.0 14.0 18.0
(c) t = 3
$M = \{\hat{\Sigma}^{(3)}_{\epsilon\epsilon}\}^{-1}$
18.0
15.0
11.0
7.4
3.7
0.11
0.11 3.7 7.4 11.0 15.0 18.0
(d) t = 4
$M = \{\hat{\Sigma}^{(4)}_{\epsilon\epsilon}\}^{-1}$
18.0
15.0
11.0
7.3
3.7
0.11
0.11 3.7 7.4 11.0 15.0 18.0
ORDERED VALUES
GAMMA QUANTILES

FIG. 7.2

REFERENCES

[1] Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist. 22, 327-351.

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

[3] Bartlett, M.S. (1947). Multivariate analysis. Journal of the Royal Statistical Society, (Supplement), 9, 176-197.

[4] Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Ann. Math. Statist. 25, 290-302.

[5] Brillinger, D.R. (1975). Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.

[6] Dempster, A.P. (1971). An overview of multivariate data analysis. J. Multivariate Anal. 1, 316-346.

[7] Dixon, W.J. (1968) (ed.). BMD Biomedical Computer Programs. University of California Press.

[8] Dixon, W.J. and Brown, M.B. (1977) (eds.). BMDP Biomedical Computer Programs, P-Series. University of California Press.

[9] Dixon, W.J. and Massey, F.J. (1969). Introduction to Statistical Analysis, Third Edition. McGraw-Hill, New York.

[10] Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika 1, 211-218.

[11] Frets, G.P. (1921). Heredity of headform in man. Genetica 3, 193-400.

[12] Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. Wiley, New York.

[13] Izenman, A.J. (1975). Reduced-rank regression for the multivariate linear model. J. Multivariate Anal. 5, 248-264.

[14] Kshirsagar, A. M. (1972). Multivariate Analysis. Marcel Dekker, New York.

[15] Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.

[16] Rao, C.R. (1965a). The use and interpretation of principal component analysis in applied research. Sankhyā A, 26, 329-358.

[17] Rao, C.R. (1965b). Linear Statistical Inference and its Applications. Wiley, New York.

[18] Rao, C.R. (1978a). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. To appear in Multivariate Analysis V (P.R. Krishnaiah, ed.). North-Holland Publishing Company.

[19] Rao, C.R. (1978b). Separation theorems for singular values of matrices and their applications in multivariate analysis. Technical Report No. 78-01, Department of Mathematics and Statistics, University of Pittsburgh.

[20] Rao, C.R. and Mitra, S.K. (1971). Generalized Inverse of Matrices and its Applications. Wiley, New York.

[21] Robinson, P.M. (1973). Generalized canonical analysis for time series. J. Multivariate Anal. 3, 140-160.

[22] Wilk, M.B., Gnanadesikan, R. and Huyett, M.J. (1962a). Probability plots for the gamma distribution. Technometrics 4, 1-20.

[23] Wilk, M.B., Gnanadesikan, R. and Huyett, M.J. (1962b). Estimation of parameters of the gamma distribution using order statistics. Biometrika 49, 525-545.

[24] Williams, E.J. (1967). The analysis of association among many variates. Journal of the Royal Statistical Society. Series B. 29, 199-242.