CANONICAL CORRELATIONS

Imagine that there are T observations on two sets of variables which give rise to the matrices Y and X of order $T\times n_y$ and $T\times n_x$ respectively, where $n_x+n_y=n$ and $T>n$. It is assumed that the elements of these matrices are adjusted by subtracting the column means and that the matrices have full-column rank.

The object is to find a set of mutually uncorrelated linear combinations of the columns of Y,

$$v_1 = Y\alpha_1, v_2 = Y\alpha_2, \ldots, v_p = Y\alpha_p, \tag{1}$$

and a similar set of combinations of the columns of X,

$$w_1 = X\beta_1, w_2 = X\beta_2, \ldots, w_p = X\beta_p, \tag{2}$$

such that, within the pairs $(v_1, w_1), (v_2, w_2), \ldots, (v_p, w_p)$, the correlations of the vectors are maximised. When these conditions are satisfied, the condition that $v_i = Y\alpha_i$ and $w_j = X\beta_j$ are uncorrelated when $i \neq j$ arises as a by-product. The vectors v_j and w_j are described as the jth canonical variates.

Let us denote the empirical variance–covariance matrix of the variables by

$$\begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} = \frac{1}{T}\begin{bmatrix} Y'Y & Y'X \\ X'Y & X'X \end{bmatrix}. \tag{3}$$

Then the constraints are that

$$\begin{gathered} \alpha_i' S_{yy}\alpha_i = 1 \quad \text{for all} \quad i, \qquad \beta_j' S_{xx}\beta_j = 1 \quad \text{for all} \quad j \\ \text{and} \quad \alpha_i' S_{yx}\beta_j = 0, \quad \text{when} \quad i \neq j. \end{gathered} \tag{4}$$

The canonical variates are found in succession beginning with v_1 and w_1. Consider the correlation

$$r(\alpha, \beta) = \frac{\alpha' S_{yx}\beta}{(\alpha' S_{yy}\alpha)^{1/2}(\beta' S_{xx}\beta)^{1/2}}. \tag{5}$$

Subject to the constraints of (4), the maximum is sought by evaluating the Lagrangean function

$$L(\alpha, \beta) = (\alpha' S_{yx}\beta)^2 - \lambda(\alpha' S_{yy}\alpha - 1) - \mu(\beta' S_{xx}\beta - 1). \tag{6}$$

Differentiating the function with respect to α and β and setting the results to zero gives

$$\begin{aligned} -\lambda S_{yy}\alpha + (\alpha' S_{yx}\beta) S_{yx}\beta &= 0, \\ (\alpha' S_{yx}\beta) S_{xy}\alpha - \mu S_{xx}\beta &= 0. \end{aligned} \tag{7}$$

Premultiplying the first equation by α' and the second by β' shows that

$$\lambda = \mu = (\alpha' S_{yx}\beta)^2, \tag{8}$$

which is to say that, under the constraints, the value of Lagrangean multiplier λ is the maximum value of $r(\alpha, \beta)$ for all α and β.

In view of (8), the equation systems of (7) can be written as

$$\begin{bmatrix} -\lambda^{1/2} S_{yy} & S_{yx} \\ S_{xy} & -\lambda^{1/2} S_{xx} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{9}$$

The equations which give α and β separately are

$$\begin{aligned} [S_{yx} S_{xx}^{-1} S_{xy} - \lambda S_{yy}]\alpha &= 0, \\ [S_{xy} S_{yy}^{-1} S_{yx} - \lambda S_{xx}]\beta &= 0. \end{aligned} \tag{10}$$

For the equations of (9) to have a solution other than $\alpha = 0$ and $\beta = 0$, it is necessary that their matrix should be singular or, equivalently, that the determinant of the matrix should vanish:

$$\begin{vmatrix} -\lambda^{1/2} S_{yy} & S_{yx} \\ S_{xy} & -\lambda^{1/2} S_{xx} \end{vmatrix} = 0. \tag{11}$$

The determinant is a polynomial of degree n. In consequence of the symmetry of the matrix, its roots are real-valued, and they can be ordered in declining magnitude: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let α_1 and β_1 be the solutions corresponding to λ_1. Then $v_1 = Y\alpha_1$ and $w_1 = X\beta_1$ are the first canonical variates and $\sqrt{\lambda_1}$ is their correlation.

The second canonical variates are the linear combinations $v_2 = Y\alpha_2$ and $w_2 = X\beta_2$ which have maximum correlation subject to the condition that they are uncorrelated with the first canonical variates v_1 and w_1. The conditions, all told, are that

$$\begin{aligned} &\text{(i)} \quad r(\alpha, \alpha_1) = \alpha' S_{yy}\alpha_1 = 0, && \text{(iii)} \quad r(\beta, \beta_1) = \beta' S_{xx}\beta_1 = 0, \\ &\text{(ii)} \quad r(\alpha, \beta_1) = \alpha' S_{yx}\beta_1 && \text{(iv)} \quad r(\beta, \alpha_1) = \beta' S_{xy}\alpha_1 \\ &\qquad\qquad\qquad = \lambda_1 \alpha' S_{yy}\alpha_1 = 0, && \qquad\qquad\qquad = \lambda_1 \beta' S_{xx}\beta_1 = 0; \end{aligned} \tag{12}$$

but it is clear that the conditions (ii) and (iv) are redundant in the sense that they follow automatically from (i) and (iii). The correlation of the second canonical variates is maximised via the Lagrangean function

(13)

$$L(\alpha, \beta) = (\alpha' S_{yx}\beta)^2 - \lambda(\alpha' S_{yy}\alpha - 1) - \mu(\beta' S_{xx}\beta - 1) + \pi\alpha' S_{yy}\alpha_1 + \rho\beta' S_{xx}\beta_1,$$

wherein λ, μ, π and ρ are the Lagrangean multipliers. Differentiating with respect to α and β and setting the results to zero yields the following conditions:

$$(14)\qquad \begin{aligned} -\lambda S_{yy}\alpha + (\alpha' S_{yx}\beta)S_{yx}\beta + \pi S_{yy}\alpha_1 &= 0, \\ (\alpha' S_{yx}\beta)S_{xy}\alpha - \mu S_{xx}\beta + \rho S_{xx}\beta_1 &= 0. \end{aligned}$$

Premultiplying these equations by α_1' and β_1', respectively, and using the conditions of (12) gives

$$(15)\qquad \begin{aligned} \pi\alpha_1' S_{yy}\alpha_1 = \pi &= 0, \\ \rho\beta_1' S_{xx}\beta_1 = \rho &= 0. \end{aligned}$$

Under these conditions, the equations of (14) assume the form of (7); and this shows that the solutions of the equations, other than α_1 and β_1, automatically satisfy the conditions of (12). If α_2 and β_2 are the solutions corresponding to λ_2, then $v_2 = Y\alpha_2$ and $w_2 = X\beta_2$ are the second canonical variates and $\sqrt{\lambda_2}$ is their correlation.

Successive canonical variates can be found which correspond to the nonzero roots of (11) ordered by magnitude. It can be show that there can be no more than $p = \min(n_x, n_y)$ roots, which is the maximum number of independent linear combinations that can be formed from whichever of X and Y has the least number of columns.

The matter can be understood best in reference to the equations of (10). There $S_{yx}S_{xx}^{-1}S_{xy}$ and $S_{xy}S_{yy}^{-1}S_{yx}$ are, respectively, the empirical moment matrices of the projection of Y on the column space of X and of the projection of X on the column space of Y.

The common rank of these moment matrices is equal to the minimum p of the ranks of X and Y. If $n_x \neq n_y$, then one of the moment matrices is bound to be rank-deficient. According to (8), the nonzero eigenvalues of $S_{yx}S_{xx}^{-1}S_{xy}$ in respect of the metric defined by S_{yy} are the same as the nonzero eigenvectors of $S_{xy}S_{yy}^{-1}S_{yx}$ in respect of the metric defined by S_{xx}. Their number p is equal to the common rank of the moment matrices.