The Annals of Statistics
1999, Vol. 27, No. 4, 1141–1154

ASYMPTOTIC DISTRIBUTION OF THE REDUCED RANK REGRESSION ESTIMATOR UNDER GENERAL CONDITIONS

BY T. W. ANDERSON

Stanford University

In the regression model $\mathbf{Y} = \eta + \mathbf{BX} + \mathbf{Z}$ with $\mathbf{Z}$ unobserved, $\mathscr{E}\mathbf{Z} = \mathbf{0}$ and $\mathscr{E}\mathbf{ZZ}' = \mathbf{\Sigma}_{ZZ}$, the least squares estimator of $\mathbf{B}$ is $\hat{\mathbf{B}} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}$. If the rank of $\mathbf{B}$ is known to be k less than the dimensions of $\mathbf{Y}$ and $\mathbf{X}$, the reduced rank regression estimator of $\mathbf{B}$, say $\hat{\mathbf{B}}_k$, depends on the first k canonical variates of $\mathbf{Y}$ and $\mathbf{X}$. The asymptotic distribution of $\hat{\mathbf{B}}_k$ is obtained and compared with the asymptotic distribution of $\hat{\mathbf{B}}$. The advantage of $\hat{\mathbf{B}}_k$ is characterized.

1. Introduction. Reduced rank regression, introduced by Anderson (1951a), has been applied in many disciplines, including econometrics, time series analysis and signal processing. See, for example, Johansen (1995) for use of reduced rank regression in estimation of cointegration in economic time series, Tsay and Tiao (1985), and Ahn and Reinsel (1988) for applications in stationary processes and Stoica and Viberg (1996) for utilization in signal processing. In general the estimated reduced rank regression is a better estimator in a regression model than the unrestricted estimator. This paper shows exactly in what sense the reduced rank estimator is better.

A general model for the dependence of a vector of p dependent variables $\mathbf{Y}_\alpha$ on a vector of q independent variables $\mathbf{X}_\alpha$ is

$$\mathbf{Y}_\alpha = \boldsymbol{\eta} + \mathbf{BX}_\alpha + \mathbf{Z}_\alpha, \tag{1.1}$$

where the unobservable disturbance or error $\mathbf{Z}_\alpha$ is distributed independently of $\mathbf{X}_\alpha$ with $\mathscr{E}\mathbf{Z}_\alpha = \mathbf{0}$ and $\mathscr{E}\mathbf{Z}_\alpha\mathbf{Z}_\alpha' = \mathbf{\Sigma}_{ZZ}$. If the rank of $\mathbf{B}$ is k, only $k \le \min(p, q)$ linear combinations of the components of $\mathbf{X}$ suffice to predict or "explain" $\mathbf{Y}$. These linear combinations are the canonical variates of $\mathbf{X}$ (defined below). The model where $k < \min(p, q)$ is called a *reduced rank regression*. The independent variables may be nonstochastic or stochastic.

On the basis of a sample $(\mathbf{y}_1, \mathbf{x}_1), \dots, (\mathbf{y}_N, \mathbf{x}_N)$ an estimator of $\mathbf{B}$ is desired. Anderson (1951a) found the maximum likelihood estimator of $\mathbf{B}$ of preassigned rank k when $\mathbf{x}_1, \dots, \mathbf{x}_N$ are considered nonstochastic and $\mathbf{z}_1, \dots, \mathbf{z}_N$ are independently distributed according to $N(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$. If $(\mathbf{Y}_\alpha', \mathbf{X}_\alpha')'$ have a joint normal distribution with mean vector $\mathscr{E}(\mathbf{Y}_\alpha', \mathbf{X}_\alpha')' = (\boldsymbol{\mu}_Y', \boldsymbol{\mu}_X')'$ and covariance

Received April 1998; revised June 1999.
AMS 1991 *subject classifications*. Primary 62H10, 62E20; secondary 62H12.
Key words and phrases. Canonical variates, reduced rank regression, maximum likelihood estimators.

matrix

$$\mathscr{E}\begin{bmatrix}\mathbf{Y}_\alpha - \boldsymbol{\mu}_Y \\ \mathbf{X}_\alpha - \boldsymbol{\mu}_X\end{bmatrix}(\mathbf{Y}'_\alpha - \boldsymbol{\mu}'_Y, \mathbf{X}'_\alpha - \boldsymbol{\mu}'_X) = \begin{bmatrix}\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{XX}\end{bmatrix} = \boldsymbol{\Sigma}, \tag{1.2}$$

then the density of $(\mathbf{Y}'_\alpha, \mathbf{X}'_\alpha)'$ is

$$\begin{aligned} & n\left[\begin{pmatrix}\mathbf{y} \\ \mathbf{x}\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{\mu}_Y \\ \boldsymbol{\mu}_X\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{XX}\end{pmatrix}\right] \\ & \qquad = n[\mathbf{y}|\boldsymbol{\mu}_Y + \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}_X), \boldsymbol{\Sigma}_{ZZ}]\, n(\mathbf{x}|\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_{XX}), \end{aligned} \tag{1.3}$$

where

$$\mathbf{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}, \tag{1.4}$$

$$\boldsymbol{\Sigma}_{ZZ} = \boldsymbol{\Sigma}_{YY} - \mathbf{B}\boldsymbol{\Sigma}_{XX}\mathbf{B}' = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}. \tag{1.5}$$

The maximum likelihood estimator of $\mathbf{B}$ of rank k in model (1.3), (1.4) and (1.5) is the same as the maximum likelihood estimator in model (1.1) with $\mathbf{x}_1, \dots, \mathbf{x}_N$ nonstochastic. We call this estimator the reduced rank regression estimator. This estimator can be defined in terms of the canonical variates. One form is $\hat{\mathbf{B}}_k = \mathbf{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}'_1$, where $\mathbf{S}_{YX}$ is the sample covariance between $\mathbf{Y}$ and $\mathbf{X}$ and $\hat{\boldsymbol{\Gamma}}_1$ consists of the coefficients of the first k canonical variates of $\mathbf{X}$; other forms are given in (2.13) below.

The major objective of this paper is to obtain the asymptotic distribution of $\hat{\mathbf{B}}_k$ for $\mathbf{x}_1, \dots, \mathbf{x}_N$ nonstochastic and for $\mathbf{x}_1, \dots, \mathbf{x}_N$ observations on a random vector with $\mathscr{E}\mathbf{X} = \boldsymbol{\mu}_X$ and $\mathscr{E}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \boldsymbol{\Sigma}_{XX}$ under the assumption that $\mathbf{X}_\alpha$ and $\mathbf{Z}_\alpha$ are independent. In fact, the conditions for the asymptotic normal distribution of the reduced rank regression estimator are the same as for the asymptotic distribution of the ordinary last squares estimator of $\mathbf{B}$. A second aim of this paper is to relate the asymptotic distribution of $\hat{\mathbf{B}}_k$ to the asymptotic distribution of $\hat{\mathbf{R}}$, $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Gamma}}$, the sample canonical correlations and coefficients of the canonical variables. In asymptotics $\hat{\mathbf{B}}_k$ is a simple function of the sample covariance matrix of $\mathbf{Y}$ and $\mathbf{X}$ expressed in terms of the canonical variables.

The asymptotic distribution of $\hat{\mathbf{B}}_k$ has been obtained by Ryan, Hubert, Carter, Sprague and Parrott (1992), Schmidli (1996), Stoica and Viberg (1996) and Reinsel and Velu (1998) by use of the expected Fisher information on the assumption that $\mathbf{Z}_\alpha$ is normally distributed. These studies are summarized and compared to the results of this paper in Section 5.

2. Canonical correlations and variables. To express and develop the results it is convenient to review the canonical correlations and variables. More details are given in Anderson [(1984), Chapter 12] and Anderson (1999), for example. The equations defining the canonical correlations and variates (in the population) are

$$\begin{pmatrix} -\rho\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & -\rho\boldsymbol{\Sigma}_{XX} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{pmatrix} = \mathbf{0}, \tag{2.1}$$

where ρ satisfies

$$\begin{vmatrix} -\rho\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & -\rho\boldsymbol{\Sigma}_{XX} \end{vmatrix} = 0, \tag{2.2}$$

and

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha} = 1, \qquad \boldsymbol{\gamma}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\gamma} = 1. \tag{2.3}$$

The number of positive canonical correlations is the rank of $\boldsymbol{\Sigma}_{YX}$, which is the rank of $\mathbf{B}$. The canonical correlations are ordered $\rho_1 \geq \cdots \geq \rho_p \geq -\rho_p \geq \cdots \geq -\rho_1$ with $q - p$ additional roots of 0 if $q > p$. We shall assume that the rank of $\boldsymbol{\Sigma}_{YX}$ is k and $\rho_1 > \cdots > \rho_k$; then the solution of (2.1) and (2.3) for such a value of ρ is unique except for multiplication by -1. To eliminate this indeterminacy we shall require that $\alpha_{ii} > 0$, $i = 1, \dots, k$. [Since the matrix $\mathbf{A} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p)$ is nonsingular, the components of $\mathbf{Y}$ can be numbered in such a way that the ith component of $\boldsymbol{\alpha}_i$ is nonzero.]

From (2.1) we obtain $\boldsymbol{\gamma} = (1/\rho)\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\alpha}$, $\boldsymbol{\alpha} = (1/\rho)\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{YX}\boldsymbol{\gamma}$,

$$\rho^2\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\Sigma}_{XX}\mathbf{B}'\boldsymbol{\alpha}, \tag{2.4}$$

$$\rho^2\boldsymbol{\Sigma}_{XX}\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1}\boldsymbol{\Sigma}_{XY}\boldsymbol{\gamma}. \tag{2.5}$$

The solutions of (2.1) corresponding to $\rho_1, \dots, \rho_p$ can be assembled as $\mathbf{A} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p)$ and $(\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$. If $q > p$, there are $q - p$ additional solutions $(\boldsymbol{\gamma}_{p+1}, \dots, \boldsymbol{\gamma}_q)$ to (2.1) with $\rho = 0$. Let $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_q)$ and let $\mathbf{R} = \operatorname{diag}(\rho_1, \dots, \rho_p)$, $\overline{\mathbf{R}} = (\mathbf{R}, \mathbf{0})$. Then the solutions can be chosen to satisfy

$$\begin{aligned} \begin{pmatrix} \mathbf{A}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}' \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{YY} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma} \end{pmatrix} &= \begin{pmatrix} \mathbf{A}'\boldsymbol{\Sigma}_{YY}\mathbf{A} & \mathbf{A}'\boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{XY}\mathbf{A} & \boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\Gamma} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I} & \overline{\mathbf{R}} \\ \overline{\mathbf{R}}' & \mathbf{I} \end{pmatrix}. \end{aligned} \tag{2.6}$$

This is the covariance matrix of the canonical variates $\mathbf{U} = \mathbf{A}'\mathbf{Y}$ and $\mathbf{V} = \boldsymbol{\Gamma}'\mathbf{X}$.

The unbiased sample means and covariances are $\bar{\mathbf{y}}$, $\bar{\mathbf{x}}$ and

$$\begin{pmatrix} \mathbf{S}_{YY} & \mathbf{S}_{YX} \\ \mathbf{S}_{XY} & \mathbf{S}_{XX} \end{pmatrix} = \frac{1}{n} \sum_{\alpha=1}^{N} \begin{pmatrix} \mathbf{y}_\alpha - \bar{\mathbf{y}} \\ \mathbf{x}_\alpha - \bar{\mathbf{x}} \end{pmatrix} (\mathbf{y}'_\alpha - \bar{\mathbf{y}}', \mathbf{x}'_\alpha - \bar{\mathbf{x}}'), \tag{2.7}$$

where $n = N - 1$. The sample equations corresponding to (2.1) and (2.3) defining the population canonical correlations and variates are

$$\begin{pmatrix} -r\mathbf{S}_{YY} & \mathbf{S}_{YX} \\ \mathbf{S}_{XY} & -r\mathbf{S}_{XX} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{c} \end{pmatrix} = \mathbf{0}, \tag{2.8}$$

$$\mathbf{a}'\mathbf{S}_{YY}\mathbf{a} = 1, \qquad \mathbf{c}'\mathbf{S}_{XX}\mathbf{c} = 1. \tag{2.9}$$

The solutions with $a_{ii} > 0$, $i = 1, \dots, p$, and $r_1 > r_2 > \cdots > r_p > 0$ define the estimators $\hat{\mathbf{A}} = (\mathbf{a}_1, \dots, \mathbf{a}_p)$, $\hat{\boldsymbol{\Gamma}} = (\mathbf{c}_1, \dots, \mathbf{c}_q)$, $\hat{\mathbf{R}} = \operatorname{diag}(r_1, \dots, r_p)$. These are uniquely defined except that if $q > p$, $\mathbf{c}_{p+1}, \dots, \mathbf{c}_q$ satisfy $\mathbf{c}'\mathbf{S}_{XX}\mathbf{c}_j = 0$, $j = 1, \dots, p$, and some other $(q-p)(q-p-1)$ arbitrary conditions. From (2.8) and (2.9) we obtain $\mathbf{c} = (1/r)\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{a}$, $\mathbf{a} = (1/r)\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{c}$,

$$\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{a} = r^2\mathbf{S}_{YY}\mathbf{a}, \tag{2.10}$$

$$\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{c} = r^2\mathbf{S}_{XX}\mathbf{c}. \tag{2.11}$$

Let $\hat{\mathbf{A}}_1 = (\mathbf{a}_1, \dots, \mathbf{a}_k)$, $\hat{\boldsymbol{\Gamma}}_1 = (\mathbf{c}_1, \dots, \mathbf{c}_k)$, $\hat{\mathbf{R}}_1 = \operatorname{diag}(r_1, \dots, r_k)$ and $\hat{\boldsymbol{\Phi}}_1 = \hat{\mathbf{A}}_1(\mathbf{I}_k - \hat{\mathbf{R}}_1^2)^{-1/2} = \mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\mathbf{R}}_1^{-1}(\mathbf{I}_k - \hat{\mathbf{R}}_1^2)^{-1/2}$. The least squares estimator of $\mathbf{B}$ in model (1.1) is

$$\hat{\mathbf{B}} = \boldsymbol{\Sigma}_{YX}\mathbf{S}_{XX}^{-1}; \tag{2.12}$$

this is the unrestricted maximum likelihood estimator under normality of the $\mathbf{Z}_\alpha$'s. The maximum likelihood estimator of $\mathbf{B}$ of rank k, found by Anderson (1951a), is

$$\hat{\mathbf{B}}_k = \mathbf{S}_{\hat{Z}\hat{Z}}\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Phi}}_1'\hat{\mathbf{B}} = \mathbf{S}_{YY}\hat{\mathbf{A}}_1\hat{\mathbf{R}}_1\hat{\boldsymbol{\Gamma}}_1' = \mathbf{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}_1', \tag{2.13}$$

where $\mathbf{S}_{\hat{Z}\hat{Z}}$ is the sample covariance matrix of the residual $\mathbf{z} = \mathbf{y} - \hat{\mathbf{B}}\mathbf{x}$. A column of $\hat{\boldsymbol{\Phi}}_1$ satisfies

$$\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\hat{\boldsymbol{\phi}} = t\mathbf{S}_{\hat{Z}\hat{Z}}\hat{\boldsymbol{\phi}}, \qquad \hat{\boldsymbol{\phi}}'\mathbf{S}_{\hat{Z}\hat{Z}}\hat{\boldsymbol{\phi}} = 1, \tag{2.14}$$

$t = r^2/(1-r^2)$; $\hat{\boldsymbol{\phi}}$ differs from the corresponding $\mathbf{a}$ only with respect to the normalization.

Another motivation of the estimator $\hat{\mathbf{B}}_k$ is that $\hat{\mathbf{B}}_k$ is the matrix $\mathbf{B}_k$ of rank k that minimizes

$$\begin{aligned} &\sum_{\alpha=1}^{N} [\mathbf{Y}_\alpha - \bar{\mathbf{y}} - \mathbf{B}_k(\mathbf{X}_\alpha - \bar{\mathbf{x}})]'\mathbf{S}_{\hat{Z}\hat{Z}}^{-1}[\mathbf{Y}_\alpha - \bar{\mathbf{y}} - \mathbf{B}_k(\mathbf{X}_\alpha - \bar{\mathbf{x}})] \\ &\quad = \operatorname{tr}(\mathbf{S}_{YY} - \mathbf{B}_k\mathbf{S}_{XY} - \mathbf{S}_{YX}\mathbf{B}_k' + \mathbf{B}_k\mathbf{S}_{XX}\mathbf{B}_k')\mathbf{S}_{\hat{Z}\hat{Z}}^{-1}. \end{aligned} \tag{2.15}$$

See Izenman (1975), Brillinger (1975) and Reinsel and Velu (1998) for other forms of this criterion and further discussions. The procedure of maximizing the normal likelihood shows that $\hat{\mathbf{B}}_k$ minimizes

$$\begin{aligned} &\frac{\left|\sum_{\alpha=1}^{N}[\mathbf{Y}_\alpha - \bar{\mathbf{y}} - \mathbf{B}_k(\mathbf{X}_\alpha - \bar{\mathbf{x}})][\mathbf{Y}_\alpha - \bar{\mathbf{y}} - \mathbf{B}_k(\mathbf{X}_\alpha - \bar{\mathbf{x}})]'\right|}{|\mathbf{S}_{\hat{Z}\hat{Z}}|} \\ &\quad = \left|(\mathbf{S}_{YY} - \mathbf{B}_k\mathbf{S}_{XY} - \mathbf{S}_{YX}\mathbf{B}_k' + \mathbf{B}_k\mathbf{S}_{XX}\mathbf{B}_k')\mathbf{S}_{\hat{Z}\hat{Z}}^{-1}\right|. \end{aligned} \tag{2.16}$$

The $\mathbf{B}_k$ that minimizes the trace criterion is identical to the $\mathbf{B}_k$ that minimizes the generalized variance.

The distribution of $n\mathbf{S}_{YY}, n\mathbf{S}_{YX}, n\mathbf{S}_{XX}$ is the same as the distribution of $\sum_{\alpha=1}^{n}\mathbf{Y}_\alpha\mathbf{Y}_\alpha', \sum_{\alpha=1}^{n}\mathbf{Y}_\alpha\mathbf{X}_\alpha' \sum_{\alpha=1}^{n}\mathbf{X}_\alpha\mathbf{X}_\alpha'$ when $\boldsymbol{\mu}_Y = \mathbf{0}$, $\boldsymbol{\mu}_X = \mathbf{0}$ and $(\mathbf{Y}', \mathbf{X}')$ is normally distributed. In any case, the limiting distribution of $\sqrt{n}\,(\mathbf{S}_{YY} -$

$\boldsymbol{\Sigma}_{YY}), \sqrt{n}(\mathbf{S}_{YX} - \boldsymbol{\Sigma}_{YX}), \sqrt{n}(\mathbf{S}_{XX} - \boldsymbol{\Sigma}_{XX})$ does not depend on $(\boldsymbol{\mu}'_Y, \boldsymbol{\mu}'_X)'$. Hence we shall consider the model as $\mathbf{Y}_\alpha = \mathbf{B}\mathbf{X}_\alpha + \mathbf{Z}_\alpha$ with

$$\begin{bmatrix} \mathbf{S}_{YY} & \mathbf{S}_{YX} \\ \mathbf{S}_{XY} & \mathbf{S}_{XX} \end{bmatrix} = \frac{1}{n} \sum_{\alpha=1}^{n} \begin{bmatrix} \mathbf{Y}_\alpha \\ \mathbf{X}_\alpha \end{bmatrix} [\mathbf{Y}'_\alpha, \mathbf{X}'_\alpha] \tag{2.17}$$

and $\mathscr{E}\mathbf{Y}_\alpha = \mathbf{0}$, $\mathscr{E}\mathbf{X}_\alpha = \mathbf{0}$.

3. Asymptotic distribution of the reduced rank regression when the independent variables are stochastic. We want to find the asymptotic distribution of $\hat{\mathbf{B}}_k = \mathbf{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}'_1$. Note that the fact that the number of columns of $\hat{\boldsymbol{\Gamma}}_1$ is the rank of $\mathbf{B}$ implies that the rank of $\mathbf{B}$ is known to the statistician. The transformation to canonical variables $\mathbf{U}_\alpha = \mathbf{A}'\mathbf{Y}_\alpha$, $\mathbf{V}_\alpha = \boldsymbol{\Gamma}'\mathbf{X}_\alpha$ and $\mathbf{W}_\alpha = \mathbf{A}'\mathbf{Z}_\alpha$ transforms $\mathbf{Y}_\alpha = \mathbf{B}\mathbf{X}_\alpha + \mathbf{Z}_\alpha$ to

$$\mathbf{U}_\alpha = \boldsymbol{\Psi}\mathbf{V}_\alpha + \mathbf{W}_\alpha, \tag{3.1}$$

$\boldsymbol{\Sigma}_{UU} = \mathbf{A}'\boldsymbol{\Sigma}_{YY}\mathbf{A} = \mathbf{I}_p$, $\boldsymbol{\Sigma}_{VV} = \boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\Gamma} = \mathbf{I}_q$, $\boldsymbol{\Sigma}_{UV} = \mathbf{A}'\boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma} = \overline{\mathbf{R}}$, $\boldsymbol{\Sigma}_{VW} = \mathbf{0}$, $\boldsymbol{\Sigma}_{WW} = \mathbf{A}'\boldsymbol{\Sigma}_{ZZ}\mathbf{A} = \mathbf{I}_p - \mathbf{R}^2$ and $\boldsymbol{\Psi} = \mathbf{A}'\mathbf{B}(\boldsymbol{\Gamma}')^{-1} = \overline{\mathbf{R}}$. Also $\mathbf{S}_{UU} = \mathbf{A}'\mathbf{S}_{YY}\mathbf{A}$, $\mathbf{S}_{UV} = \mathbf{A}'\mathbf{S}_{YX}\boldsymbol{\Gamma}$, $\mathbf{S}_{VV} = \boldsymbol{\Gamma}\mathbf{S}_{XX}\boldsymbol{\Gamma}'$, $\hat{\boldsymbol{\Psi}} = \mathbf{S}_{UV}\mathbf{S}_{VV}^{-1} = \mathbf{A}'\hat{\mathbf{B}}(\boldsymbol{\Gamma}')^{-1}$ (the unrestricted estimator of $\boldsymbol{\Psi}$) and the restricted estimator

$$\hat{\boldsymbol{\Psi}}_k = \mathbf{S}_{UV}\mathbf{H}_1\mathbf{H}'_1, \tag{3.2}$$

where $\mathbf{H}_1 = \boldsymbol{\Gamma}_1^{-1}\hat{\boldsymbol{\Gamma}}_1$ satisfies

$$\mathbf{S}_{VU}\mathbf{S}_{UU}^{-1}\mathbf{S}_{UV}\mathbf{H}_1 = \mathbf{S}_{VV}\mathbf{H}_1\hat{\mathbf{R}}_1^2, \qquad \mathbf{H}'_1\mathbf{S}_{VV}\mathbf{H}_1 = \mathbf{I}_k. \tag{3.3}$$

The limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi})$ will be found (Theorem 1 below) and transformed back to the original coordinates (Corollary 1).

Define $\mathbf{S}^*_{UU} = \sqrt{n}(\mathbf{S}_{UU} - \mathbf{I}_p)$, $\mathbf{S}^*_{UV} = \sqrt{n}(\mathbf{S}_{UV} - \overline{\mathbf{R}})$, $\mathbf{S}^*_{VV} = \sqrt{n}(\mathbf{S}_{VV} - \mathbf{I}_q)$, $\mathbf{H}^*_1 = \sqrt{n}(\mathbf{H}_1 - \mathbf{I}_{(k)})$ and $\overline{\mathbf{R}}^* = [\sqrt{n}(\hat{\mathbf{R}} - \mathbf{R}), \mathbf{0}]$, where $\mathbf{I}_{(k)} = (\mathbf{I}_k, \mathbf{0})'$ is $q \times k$. Then substitution of these quantities into (3.3) yields

$$\begin{aligned} &\overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{I}_{(k)} + \frac{1}{\sqrt{n}}\Big[\mathbf{S}^*_{VU}\overline{\mathbf{R}}\mathbf{I}_{(k)} + \overline{\mathbf{R}}'\mathbf{S}^*_{UV}\mathbf{I}_{(k)} - \overline{\mathbf{R}}'\mathbf{S}^*_{UU}\overline{\mathbf{R}}\mathbf{I}_{(k)} + \overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{H}^*_1\Big] \\ &\quad = \mathbf{I}_{(k)}\mathbf{R}_1^2 + \frac{1}{\sqrt{n}}\Big[\mathbf{S}^*_{VV}\mathbf{I}_{(k)}\mathbf{R}_1^2 + 2\mathbf{I}_{(k)}\mathbf{R}_1\mathbf{R}^*_1 + \mathbf{H}^*_1\mathbf{R}_1^2\Big] + o_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{3.4}$$

or equivalently,

$$\begin{aligned} &\mathbf{S}^*_{VU}\overline{\mathbf{R}}\mathbf{I}_{(k)} + \overline{\mathbf{R}}'\mathbf{S}^*_{UV}\mathbf{I}_{(k)} - \overline{\mathbf{R}}'\mathbf{S}^*_{UU}\overline{\mathbf{R}}\mathbf{I}_{(k)} - \mathbf{S}^*_{VV}\mathbf{I}_{(k)}\mathbf{R}_1^2 \\ &\quad = 2\mathbf{I}_{(k)}\mathbf{R}_1\mathbf{R}^*_1 + \mathbf{H}^*_1\mathbf{R}_1^2 - \overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{H}^*_1 + o_p(1). \end{aligned} \tag{3.5}$$

In terms of partitions into submatrices of k and $q - k$ rows, (3.5) is

$$\begin{aligned} &\begin{bmatrix} \mathbf{S}^{*11}_{VU}\mathbf{R}_1 + \mathbf{R}_1\mathbf{S}^{*11}_{UV} - \mathbf{R}_1\mathbf{S}^{*11}_{UU}\mathbf{R}_1 - \mathbf{S}^{*11}_{VV}\mathbf{R}_1^2 \\ \mathbf{S}^{*21}_{VU}\mathbf{R}_1 - \mathbf{S}^{*21}_{VV}\mathbf{R}_1^2 \end{bmatrix} \\ &\quad = \begin{bmatrix} 2\mathbf{R}_1\mathbf{R}^*_1 + \mathbf{H}^*_{11}\mathbf{R}_1^2 - \mathbf{R}_1^2\mathbf{H}^*_{11} \\ \mathbf{H}^*_{21}\mathbf{R}_1^2 \end{bmatrix} + o_p(1), \end{aligned} \tag{3.6}$$

where $\mathbf{H}_1^* = (\mathbf{H}_{11}^{*\prime}, \mathbf{H}_{21}^{*\prime})'$. From (3.6) we obtain

$$\mathbf{H}_{21}^*\mathbf{R}_1 + o_p(1) = \mathbf{S}_{VU}^{*21} - \mathbf{S}_{VV}^{*21}\mathbf{R}_1 = \mathbf{S}_{VW}^{*21} = \left(\mathbf{S}_{WV}^{*12}\right)'. \tag{3.7}$$

From the second part of (3.3) we obtain

$$\mathbf{I}_{(k)}'\mathbf{I}_{(k)} + \frac{1}{\sqrt{n}}\left[\mathbf{H}_1^{*\prime}\mathbf{I}_{(k)} + \mathbf{I}_{(k)}'\mathbf{H}_1^* + \mathbf{I}_{(k)}'\mathbf{S}_{VV}^*\mathbf{I}_{(k)}\right] = \mathbf{I}_k + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{3.8}$$

from which we obtain

$$\mathbf{H}_{11}^{*\prime} + \mathbf{H}_{11}^* = -\mathbf{S}_{VV}^{*11} + o_p(1). \tag{3.9}$$

From (3.2) and the definition $\boldsymbol{\Psi}_k^* = \sqrt{n}\,(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi})$, we obtain

$$\begin{aligned}
\boldsymbol{\Psi}_k^* &= \mathbf{S}_{UV}^*\mathbf{I}_{(k)}\mathbf{I}_{(k)}' + \overline{\mathbf{R}}\mathbf{H}_1^*\mathbf{I}_{(k)}' + \overline{\mathbf{R}}\mathbf{I}_{(k)}\mathbf{H}_1^{*\prime} + o_p(1) \\
&= \begin{bmatrix} \mathbf{R}_1(\mathbf{H}_{11}^* + \mathbf{H}_{11}^{*\prime}) + \mathbf{S}_{UV}^{*11} & \mathbf{R}_1\mathbf{H}_{21}^{*\prime} \\ \mathbf{S}_{UV}^{*21} & \mathbf{0} \end{bmatrix} + o_p(1) \\
&= \begin{bmatrix} \mathbf{S}_{WV}^{*11} & \mathbf{S}_{WV}^{*12} \\ \mathbf{S}_{WV}^{*21} & \mathbf{0} \end{bmatrix} + o_p(1),
\end{aligned} \tag{3.10}$$

where the partitioning is into k and $p-k$ rows and k and $q-k$ columns. The last equality follows from (3.7), (3.9) and $\mathbf{S}_{UV}^* - \overline{\mathbf{R}}\mathbf{S}_{VV}^* = \mathbf{S}_{WV}^*$.

The maximum likelihood estimator of $\mathbf{B}$ unrestricted with respect to rank is $\hat{\mathbf{B}} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}$, the unrestricted estimator in terms of canonical variables is $\hat{\boldsymbol{\Psi}} = \mathbf{S}_{UV}\mathbf{S}_{VV}^{-1}$ and $\hat{\boldsymbol{\Psi}}^* = \sqrt{n}\,(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \mathbf{S}_{WV}^*$. The effect of the rank restriction is to replace the lower right-hand corner of $\mathbf{S}_{WV}^*$ by $\mathbf{0}$.

To characterize the asymptotic distribution of $\hat{\boldsymbol{\Psi}}_k$ and $\hat{\mathbf{B}}_k$, we use the notation $\operatorname{vec}\mathbf{A} = \operatorname{vec}(\mathbf{a}_1, \dots, \mathbf{a}_m) = (\mathbf{a}_1', \dots, \mathbf{a}_m')'$ and $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$ and the property $\operatorname{vec}\mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A})\operatorname{vec}\mathbf{B}$, which implies $\operatorname{vec}\mathbf{xy}' = \operatorname{vec}\mathbf{x}\mathbf{1}\mathbf{y}' = (\mathbf{y} \otimes \mathbf{x})\operatorname{vec}\mathbf{1} = \mathbf{y} \otimes \mathbf{x}$. Then

$$\begin{aligned}
\operatorname{vec}\hat{\boldsymbol{\Psi}}_k^* &= \operatorname{vec}\frac{1}{\sqrt{n}}\left[\sum_{\alpha=1}^n \mathbf{W}_\alpha\mathbf{V}_\alpha^{(1)\prime}, \sum_{\alpha=1}^n \begin{pmatrix} \mathbf{W}_\alpha^{(1)} \\ \mathbf{0} \end{pmatrix}\mathbf{V}_\alpha^{(2)\prime}\right] \\
&= \frac{1}{\sqrt{n}}\sum_{\alpha=1}^n \begin{bmatrix} \mathbf{V}_\alpha^{(1)} \otimes \mathbf{W}_\alpha \\ \mathbf{V}_\alpha^{(2)} \otimes \begin{pmatrix} \mathbf{W}_\alpha^{(1)} \\ \mathbf{0} \end{pmatrix} \end{bmatrix}.
\end{aligned} \tag{3.11}$$

Since $\mathbf{V}_\alpha$ and $\mathbf{W}_\alpha$ are assumed to be independent, we obtain

$$\begin{aligned}
&\mathscr{E}\operatorname{vec}\hat{\boldsymbol{\Psi}}_k^*\left(\operatorname{vec}\hat{\boldsymbol{\Psi}}_k^*\right)' \\
&\quad\to \begin{bmatrix} \mathbf{I}_k \otimes \left(\mathbf{I}_p - \mathbf{R}^2\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{q-k} \otimes \begin{bmatrix} \mathbf{I}_k - \mathbf{R}_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{bmatrix} \\
&\quad= \operatorname{diag}\left(\mathbf{I}_p - \mathbf{R}^2, \dots, \mathbf{I}_p - \mathbf{R}^2, \mathbf{I}_k - \mathbf{R}_1^2, \mathbf{0}, \dots, \mathbf{I}_k - \mathbf{R}_1^2, \mathbf{0}\right),
\end{aligned} \tag{3.12}$$

where there are k blocks of $\mathbf{I}_p - \mathbf{R}^2$ and $q-k$ blocks of $\operatorname{diag}(\mathbf{I}_k - \mathbf{R}_1^2, \mathbf{0})$. The rank of (3.12) is $kp + (q-k)k = k(p+q-k)$.

THEOREM 1. *Let* $(\mathbf{u}'_\alpha, \mathbf{v}'_\alpha)$, $\alpha = 1, \dots, n$, *be observations on the random vector* $(\mathbf{U}', \mathbf{V}')$ *with mean* $\mathbf{0}$ *and covariance matrix* (2.6). *Let* $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{UV}\boldsymbol{\Sigma}_{VV}^{-1}$, $\mathbf{S}_{UU} = n^{-1}\sum_{\alpha=1}^n \mathbf{u}_\alpha \mathbf{u}'_\alpha$, $\mathbf{S}_{UV} = n^{-1}\sum_{\alpha=1}^n \mathbf{u}_\alpha \mathbf{v}'_\alpha$, $\mathbf{S}_{VV} = n^{-1}\sum_{\alpha=1}^n \mathbf{v}_\alpha \mathbf{v}'_\alpha$. *distribution of* $\operatorname{vec}\sqrt{n}(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi})$, *where* $\hat{\boldsymbol{\Psi}}_k$ *is defined by* (3.2), *is normal with mean* $\mathbf{0}$ *and covariance matrix* (3.12).

For the least squares estimator $\hat{\boldsymbol{\Psi}}$ we have

$$\sqrt{n}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \hat{\boldsymbol{\Psi}}^* = \begin{bmatrix} \mathbf{S}_{WV}^{*11} & \mathbf{S}_{WV}^{*12} \\ \mathbf{S}_{WV}^{*21} & \mathbf{S}_{WV}^{*22} \end{bmatrix} + o_p(1). \tag{3.13}$$

Since the four submatrices on the right-hand side of (3.13) are uncorrelated $[\mathbf{S}_{WV}^{*gh} = (1/\sqrt{N})\sum_{\alpha=1}^n \mathbf{W}_\alpha^{(g)}\mathbf{V}_\alpha^{(h)\prime}]$, they are independent in the limiting normal distribution. Then $\hat{\boldsymbol{\Psi}}_k^*$ and

$$\hat{\boldsymbol{\Psi}}^* - \hat{\boldsymbol{\Psi}}_k^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{WV}^{*22} \end{bmatrix} \tag{3.14}$$

are asymptotically independent with

$$\begin{aligned} &\mathscr{E} \operatorname{vec}\left(\hat{\boldsymbol{\Psi}}^* - \hat{\boldsymbol{\Psi}}_k^*\right)\left[\operatorname{vec}\left(\hat{\boldsymbol{\Psi}}^* - \hat{\boldsymbol{\Psi}}_k\right)\right]' \\ &\quad \to \mathscr{E}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{V}^{(2)} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{W}^{(2)} \end{pmatrix}\right]\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{V}^{(2)} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{W}^{(2)} \end{pmatrix}\right]' \\ &\quad = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \mathbf{I}_{q-k} \otimes & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-k} \end{bmatrix} \end{bmatrix}. \end{aligned} \tag{3.15}$$

In the original coordinate system we obtain

$$\begin{aligned} \operatorname{vec}\left(\hat{\mathbf{B}}_k - \mathbf{B}\right) &= \operatorname{vec}\left[(\mathbf{A}')^{-1}\left(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi}\right)\boldsymbol{\Gamma}'\right] \\ &= \left[\boldsymbol{\Gamma} \otimes (\mathbf{A}')^{-1}\right]\operatorname{vec}\left(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi}\right) \\ &= \left[(\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) \otimes \boldsymbol{\Sigma}_{ZZ}(\mathbf{A}_1, \mathbf{A}_2)\left(\mathbf{I}_p - \mathbf{R}^2\right)^{-1}\right]\operatorname{vec}\left(\hat{\boldsymbol{\Psi}}_k - \boldsymbol{\Psi}\right). \end{aligned} \tag{3.16}$$

From (3.12) and (3.16) we obtain

$$\begin{aligned} &\mathscr{E} \operatorname{vec} n\left(\hat{\mathbf{B}}_k - \mathbf{B}\right)\left[\operatorname{vec}\left(\hat{\mathbf{B}}_k - \mathbf{B}\right)\right]' \\ &\quad \to \left[(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}'_1) \otimes \boldsymbol{\Sigma}_{ZZ}\mathbf{A}\left(\mathbf{I}_p - \mathbf{R}^2\right)^{-1}\right] \\ &\qquad + \left[\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}'_2 \otimes \boldsymbol{\Sigma}_{ZZ}\mathbf{A}_1\left(\mathbf{I}_k - \mathbf{R}_1^2\right)^{-1}\mathbf{A}'_1\boldsymbol{\Sigma}_{ZZ}\right] \\ &\quad = [\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}'_1 \otimes \boldsymbol{\Sigma}_{ZZ}] + \left[\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}'_2 \otimes \boldsymbol{\Sigma}_{ZZ}\mathbf{A}_1\left(\mathbf{I}_k - \mathbf{R}_1^2\right)^{-1}\mathbf{A}'_1\boldsymbol{\Sigma}_{ZZ}\right] \\ &\quad = \boldsymbol{\Sigma}_{XX}^{-1} \otimes \boldsymbol{\Sigma}_{ZZ} - (\boldsymbol{\Gamma}_2\boldsymbol{\Gamma}'_2 \otimes \boldsymbol{\Sigma}_{ZZ}\mathbf{A}_2\mathbf{A}'_2\boldsymbol{\Sigma}_{ZZ}). \end{aligned} \tag{3.17}$$

If we define $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma}_1 = \boldsymbol{\Sigma}_{ZZ}\mathbf{A}_1\mathbf{R}_1(\mathbf{I} - \mathbf{R}_1^2)^{-1}$ and $\boldsymbol{\Pi} = \boldsymbol{\Gamma}_1$, then $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$. We have

$$\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}_{ZZ} - \boldsymbol{\Sigma}_{ZZ}\mathbf{A}_2\mathbf{A}_2'\boldsymbol{\Sigma}_{ZZ}, \tag{3.18}$$

$$\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}' = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1' = \boldsymbol{\Sigma}_{XX}^{-1} - \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2'. \tag{3.19}$$

Thus (3.17) can be written

$$\begin{aligned}\mathscr{E}\operatorname{vec}\hat{\mathbf{B}}_k^*\left(\operatorname{vec}\hat{\mathbf{B}}_k^*\right)' \to \boldsymbol{\Sigma}_{XX}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ} - \left[\boldsymbol{\Sigma}_{XX}^{-1} - \boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}'\right]&\\ \otimes\left[\boldsymbol{\Sigma}_{ZZ} - \boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\right]&.\end{aligned} \tag{3.20}$$

COROLLARY 1. *Let* $(\mathbf{y}_\alpha', \mathbf{x}_\alpha')'$, $\alpha = 1, \dots, n$, *be observations on the random vector* $(\mathbf{Y}', \mathbf{X}')'$ *with mean* $\mathbf{0}$ *and covariance matrix* (1.2). *Let* $\mathbf{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}$, $\mathbf{S}_{YY} = n^{-1}\sum_{\alpha=1}^n \mathbf{y}_\alpha\mathbf{y}_\alpha'$, $\mathbf{S}_{XY} = n^{-1}\sum_{\alpha=1}^n \mathbf{x}_\alpha\mathbf{y}_\alpha'$, $\mathbf{S}_{XX} = n^{-1}\sum_{\alpha=1}^n \mathbf{x}_\alpha\mathbf{x}_\alpha'$. *Let the columns of* $\hat{\boldsymbol{\Gamma}}_1$ *satisfy* (2.9), (2.11) *and* $\hat{\gamma}_{ii} > 0$. *Suppose that* $\mathbf{Y} - \mathbf{B}\mathbf{X} = \mathbf{Z}$ *is independent of* $\mathbf{X}$. *Then the limiting distribution of* $\operatorname{vec}\hat{\mathbf{B}}_k^* = \sqrt{n}\,\operatorname{vec}(\hat{\mathbf{B}}_k - \mathbf{B})$, *with* $\hat{\mathbf{B}}_k = \mathbf{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}_1'$, *is normal with mean* $\mathbf{0}$ *and covariance matrix* (3.17) *or* (3.20).

Note that $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}' = \boldsymbol{\Lambda}\mathbf{M}'(\boldsymbol{\Pi}\mathbf{M}^{-1})'$ for arbitrary nonsingular $\mathbf{M}$; however, (3.18) and (3.19) are invariant with respect to the transformation $\boldsymbol{\Lambda} \to \boldsymbol{\Lambda}\mathbf{M}$ and $\boldsymbol{\Pi} \to \boldsymbol{\Pi}\mathbf{M}^{-1}$. Thus (3.20) holds for any factorization $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$.

It is interesting that the limiting distribution of $\hat{\mathbf{B}}_k^*$ only depends on $\sqrt{n}\,\mathbf{S}_{ZX}\mathbf{S}_{XX}^{-1} = (\mathbf{A}')^{-1}\mathbf{S}_{WV}^*\mathbf{S}_{VV}^{-1}\boldsymbol{\Gamma}'$ and hence holds under the same conditions as the asymptotic normality of the least squares estimator $\hat{\mathbf{B}}$. However, $\hat{\mathbf{B}}_k$ is a function of $\hat{\boldsymbol{\Gamma}}_1$, and the asymptotic distribution of $\hat{\boldsymbol{\Gamma}}_1$ requires some properties of the normal distribution of $\mathbf{Z}$.

From (3.6) we have

$$\begin{gathered} h_{ij}^*\left(\rho_j^2 - \rho_i^2\right) = s_{ij}^{*VU}\rho_j + \rho_i s_{ij}^{*UV} - \rho_i s_{ij}^{*UU}\rho_j - s_{ij}^{*VV}\rho_j^2 + o_p(1),\\ i \neq j,\ i = 1, \dots, q,\ j = 1, \dots, k. \end{gathered} \tag{3.21}$$

The second-order moments of h_{ij}^*, $i \neq j$, depend on the second-order moments of terms like v_iu_j, u_iu_j and v_iv_j; under normality $\mathscr{E}(v_iu_j)^2 = \mathscr{E}v_i^2u_j^2 = \mathscr{E}v_i^2\mathscr{E}u_j^2 = 1$, for example. On the other hand, the limiting distribution of $S_{WV}^* = (1/\sqrt{n})\sum_{\alpha=1}^n \mathbf{W}_\alpha\mathbf{V}_\alpha'$ depends only on $\mathbf{W}_\alpha$ and $\mathbf{V}_\alpha$ ($\mathbf{Z}_\alpha$ and $\mathbf{X}_\alpha$) being independent. The covariances of h_{ij}^* are valid for $(\mathbf{Y}, \mathbf{X})$ normally distributed and $\rho_1 > \cdots > \rho_k$; they depend on the second-order moments of the sample covariance matrices (hence, on the fourth-order moments of the observed variables). However, the limiting distribution of $\operatorname{vec}\hat{\mathbf{B}}^* = \sqrt{N}(\hat{\mathbf{B}} - \mathbf{B})$ is $N(\mathbf{0}, \boldsymbol{\Sigma}_{XX}^{-1}\otimes\boldsymbol{\Sigma}_{ZZ})$, irrespective of whether $\mathbf{Y}$ and $\mathbf{X}$ are normal.

Note that $\mathbf{H}_{21}^*$ is $\mathbf{S}_{VW}^{*21}\mathbf{R}_1^{-1} + o_p(1)$ and asymptotically does not depend on $\mathbf{S}_{WW}^*$; although $\mathbf{H}_{11}^*$ does depend on $\mathbf{S}_{WW}^*$, it enters $\hat{\boldsymbol{\Psi}}_k^*$ only through $\mathbf{H}_{11}^* + \mathbf{H}_{11}^{*\prime} = -\mathbf{S}_{VV}^{*11} + o_p(1)$. Thus the limiting distribution of $\hat{\boldsymbol{\Psi}}_k^*$ and hence of $\hat{\mathbf{B}}_k^*$ does not depend on $\mathbf{S}_{WW}^*$. In particular, the distributions of Theorem 1 and Corollary 1 are valid for $\mathbf{S}_{WW}^* = 0$, which is equivalent to replacing $\mathbf{S}_{WW}$ or $\mathbf{S}_{ZZ}$ by $\boldsymbol{\Sigma}_{WW}$ or $\boldsymbol{\Sigma}_{ZZ}$, respectively.

4. Asymptotic distribution when the independent variables are nonstochastic. Now suppose that $\mathbf{X}_\alpha = \mathbf{x}_\alpha$, $\alpha = 1, \dots, n$, is nonstochastic. We assume that

$$\mathbf{S}_{XX} = \frac{1}{n} \sum_{\alpha=1}^{n} \mathbf{x}_\alpha \mathbf{x}'_\alpha \to \boldsymbol{\Sigma}_{XX}. \tag{4.1}$$

The model is

$$\mathbf{Y}_\alpha = \mathbf{B}\mathbf{x}_\alpha + \mathbf{Z}_\alpha, \tag{4.2}$$

where $\mathscr{E}\mathbf{Z}_\alpha = \mathbf{0}$ and $\mathscr{E}\mathbf{Z}_\alpha \mathbf{Z}'_\alpha = \boldsymbol{\Sigma}_{ZZ}$.

We shall find a suitable canonical form by replacing (2.1) and (2.3) by

$$\begin{bmatrix} -\rho(\boldsymbol{\Sigma}_{ZZ} + \mathbf{B}\mathbf{S}_{XX}\mathbf{B}') & \mathbf{B}\mathbf{S}_{XX} \\ \mathbf{S}_{XX}\mathbf{B}' & -\rho\mathbf{S}_{XX} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} = \mathbf{0}, \tag{4.3}$$

$$\boldsymbol{\alpha}'(\boldsymbol{\Sigma}_{ZZ} + \mathbf{B}\mathbf{S}_{XX}\mathbf{B}')\boldsymbol{\alpha} = 1, \qquad \boldsymbol{\gamma}'\mathbf{S}_{XX}\boldsymbol{\gamma} = 1. \tag{4.4}$$

Solving the second vector equation in (4.3) for $\rho\boldsymbol{\gamma} = \mathbf{B}'\boldsymbol{\alpha}$ and substituting in the first gives

$$\mathbf{B}\mathbf{S}_{XX}\mathbf{B}'\boldsymbol{\alpha} = \rho^2(\boldsymbol{\Sigma}_{ZZ} + \mathbf{B}\mathbf{S}_{XX}\mathbf{B}')\boldsymbol{\alpha}. \tag{4.5}$$

This equation and (4.4) imply $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\alpha} = 1 - \rho^2$, $\theta = \rho^2/(1-\rho^2)$ and $\boldsymbol{\phi} = \boldsymbol{\alpha}(1-\rho^2)^{-1/2}$. The solutions to (4.3) and (4.4) and $\alpha_{ii} > 0$, $\rho_1 > \cdots > \rho_k$, define $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_k$. The other columns of $\mathbf{A}_n = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p)$ and $\boldsymbol{\Gamma}_n = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_q)$ can be defined so

$$\boldsymbol{\Sigma}_{WW} = \mathbf{A}'_n \boldsymbol{\Sigma}_{ZZ} \mathbf{A}_n = \mathbf{I}_p - \mathbf{R}_n^2, \tag{4.6}$$

$$\mathbf{S}_{VV} = \boldsymbol{\Gamma}'_n \mathbf{S}_{XX} \boldsymbol{\Gamma}_n = \mathbf{I}_q, \tag{4.7}$$

$$\mathbf{A}'_n \mathbf{B}\mathbf{S}_{XX}\boldsymbol{\Gamma}_n = \overline{\mathbf{R}}_n = \mathbf{A}'_n \mathbf{B}(\boldsymbol{\Gamma}'_n)^{-1}, \tag{4.8}$$

where $\mathbf{U} = \mathbf{A}'_n\mathbf{Y}$, $\mathbf{v}_\alpha = \boldsymbol{\Gamma}'_n\mathbf{x}_\alpha$, $\mathbf{W} = \mathbf{A}'_n\mathbf{Z}$. Where convenient, the subscript n is used to emphasize that the matrices of transformed parameters depend on n through $\mathbf{S}_{XX}$. We write the model for $\mathbf{U}$ in terms of $\mathbf{v}$ and $\mathbf{W}$ as

$$\mathbf{U} = \boldsymbol{\Psi}\mathbf{v} + \mathbf{W}, \tag{4.9}$$

where $\overline{\mathbf{R}}_n = \mathbf{A}'_n\mathbf{B}(\boldsymbol{\Gamma}'_n)^{-1}$ has been replaced by $\boldsymbol{\Psi}$.

The unrestricted maximum likelihood estimators of $\mathbf{B}$ and $\boldsymbol{\Sigma}_{ZZ}$ are given by (2.12) and

$$\mathbf{S}_{\hat{Z}\hat{Z}} = \frac{1}{n} \sum_{\alpha=1}^{n} \left(\mathbf{y}_\alpha - \hat{\mathbf{B}}\mathbf{x}_\alpha\right)\left(\mathbf{y}_\alpha - \hat{\mathbf{B}}\mathbf{x}_\alpha\right)' = \mathbf{S}_{YY} - \hat{\mathbf{B}}\mathbf{S}_{XX}\hat{\mathbf{B}}'. \tag{4.10}$$

The estimators of $\mathbf{A}_n$, $\boldsymbol{\Gamma}_n$ and $\mathbf{R}_n^2$ are formed from the solution of (2.9), (2.10) and (2.11) as in Section 3.

When we transform from $\mathbf{Y}$, $\mathbf{x}$ and $\mathbf{Z}$ to $\mathbf{U}$, $\mathbf{v}$ and $\mathbf{W}$, the estimators of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}_{WW}$ are

$$\hat{\boldsymbol{\Psi}} = \mathbf{S}_{UV}\mathbf{S}_{VV}^{-1} = \mathbf{S}_{UV}, \tag{4.11}$$

$$\begin{aligned} \mathbf{S}_{\hat{W}\hat{W}} &= \frac{1}{n}\sum_{\alpha=1}^{n}\left(\mathbf{u}_\alpha - \hat{\boldsymbol{\Psi}}\mathbf{v}_\alpha\right)\left(\mathbf{u}_\alpha - \hat{\boldsymbol{\Psi}}\mathbf{v}_\alpha\right)' \\ &= \mathbf{S}_{UU} - \hat{\boldsymbol{\Psi}}\mathbf{S}_{VV}\hat{\boldsymbol{\Psi}}' = \mathbf{S}_{UU} - \hat{\boldsymbol{\Psi}}\hat{\boldsymbol{\Psi}}' \\ &= \mathbf{S}_{UU} - \mathbf{S}_{UV}\mathbf{S}_{VV}^{-1}\mathbf{S}_{VU} = \mathbf{S}_{UU} - \mathbf{S}_{UV}\mathbf{S}_{VU}. \end{aligned} \tag{4.12}$$

Now $\mathbf{H}_1 = \boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\Gamma}}_1$ satisfies

$$\mathbf{S}_{VU}\mathbf{S}_{UU}^{-1}\mathbf{S}_{UV}\mathbf{H}_1 = \mathbf{S}_{VV}\mathbf{H}_1\hat{\mathbf{R}}_1^2 = \mathbf{H}_1\hat{\mathbf{R}}_1^2, \tag{4.13}$$

$$\mathbf{I}_k = \mathbf{H}_1'\mathbf{S}_{VV}\mathbf{H}_1 = \mathbf{H}_1'\mathbf{H}_1. \tag{4.14}$$

Substitution for $\mathbf{S}_{VU}, \mathbf{S}_{UU}, \mathbf{H}_1, \hat{\mathbf{R}}_1^2$ in (4.13) yields

$$\begin{aligned} &\mathbf{S}_{UV}^*\overline{\mathbf{R}}\mathbf{I}_{(k)} + \overline{\mathbf{R}}'\mathbf{S}_{UV}^*\mathbf{I}_{(k)} - \overline{\mathbf{R}}'\mathbf{S}_{UU}^*\overline{\mathbf{R}}\mathbf{I}_{(k)} \\ &\qquad = 2\mathbf{I}_{(k)}\mathbf{R}_1\mathbf{R}_1^* + \mathbf{H}_1^*\mathbf{R}_1^2 - \overline{\mathbf{R}}'\overline{\mathbf{R}}\mathbf{H}_1^* + o_p(1), \end{aligned} \tag{4.15}$$

which is (3.5) with $\mathbf{S}_{VV}\mathbf{I}_{(k)}\mathbf{R}_1^2$ omitted. As in Section 3, (4.15) implies

$$\mathbf{H}_{21}^*\mathbf{R}_1 + o_p(1) = \left(\mathbf{S}_{WV}^{*12}\right)'. \tag{4.16}$$

Substitution in (4.14) yields

$$\mathbf{H}_{11}^{*\prime} + \mathbf{H}_{11}^* = \mathbf{0} + o_p(1). \tag{4.17}$$

Then again

$$\hat{\boldsymbol{\Psi}}_k^* = \begin{bmatrix} \mathbf{S}_{WV}^{*11} & \mathbf{S}_{WV}^{*12} \\ \mathbf{S}_{WV}^{*21} & \mathbf{0} \end{bmatrix} + o_p(1). \tag{4.18}$$

In (3.11) $\mathbf{v}_\alpha$ is nonstochastic and (3.12) holds.

The conclusions of Theorem 1 and Corollary 1 hold for $\hat{\boldsymbol{\Psi}}_k$ and $\hat{\mathbf{B}}_k$, respectively, when the independent variables are nonstochastic.

The distribution of the roots and vectors of (2.14) have been given by Anderson (1951b) for arbitrary multiplicities of the population roots when $\mathbf{Z}_\alpha$ is normally distributed and the independent variables are nonstochastic.

As in Section 3, the limiting distribution of $\mathbf{H}_{21}^*$ does not depend on $\mathbf{S}_{WW}^*$ and $\mathbf{H}_{11}^*$ does not enter (asymptotically) $\hat{\boldsymbol{\Psi}}_k^*$. Hence, the limiting distribution of $\hat{\mathbf{B}}_k^*$ is the same for $\boldsymbol{\Sigma}_{WW}$ known as for $\boldsymbol{\Sigma}_{WW}$ estimated.

5. Discussion.

5.1. *Asymptotic distribution of the reduced rank regression estimator under normality.* Ryan, Hubert, Carter, Sprague and Parrott (1992) let $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$ and $\boldsymbol{\eta} = \mathbf{0}$ and assume $\mathbf{X}_\alpha$ in (1.1) is nonstochastic and $\mathbf{Z}_a$ is normally distributed. They differentiate the log-likelihood function with respect to the elements of $\boldsymbol{\Lambda}$, $\boldsymbol{\Pi}$ and $\boldsymbol{\Sigma}_{ZZ}$ to obtain the expected Fisher information matrix. From this they derive the asymptotic covariance matrix of the maximum likelihood estimators of $\boldsymbol{\Lambda}$, $\boldsymbol{\Pi}$, and $\boldsymbol{\Sigma}_{ZZ}$, presumably as a generalized inverse. Then the asymptotic covariance matrix of the estimator of $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$ is the asymptotic covariance matrix of $\boldsymbol{\Lambda}\hat{\boldsymbol{\Pi}}' + \hat{\boldsymbol{\Lambda}}\boldsymbol{\Pi}'$. This agrees with (3.20). The authors do not comment on the indeterminacy in $\boldsymbol{\Lambda}\boldsymbol{\Pi}'$.

Schmidli (1996) goes through the same steps, but with more care, and reaches the same asymptotic covariance matrix of the estimator of $\mathbf{B}$.

Stoica and Viberg (1996) have obtained an expression for the covariance of the limiting distribution of $\sqrt{n}\,(\hat{\mathbf{B}}_k - \mathbf{B})$ where the $\mathbf{x}$'s are nonstochastic satisfying (4.1) by assuming that the $\mathbf{Y}$'s (or $\mathbf{Z}$'s) are normally distributed and calculating the Fisher information matrix. The expression is

$$
\begin{aligned}
\mathscr{E} n \operatorname{vec}\bigl(\hat{\mathbf{B}}_k - \mathbf{B}\bigr)\bigl[\operatorname{vec}\bigl(\hat{\mathbf{B}}_k - \mathbf{B}\bigr)\bigr]' \to{} & \bigl[(\boldsymbol{\Gamma}_1 \otimes \mathbf{I}_p), (\mathbf{I}_q \otimes \boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma}_1)\bigr] \\
& \times \left\{ \begin{bmatrix} \boldsymbol{\Gamma}_1' \otimes \mathbf{I}_p \\ \mathbf{I}_q \otimes \boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{XY} \end{bmatrix} \bigl[\boldsymbol{\Sigma}_{XX} \otimes \boldsymbol{\Sigma}_{ZZ}^{-1}\bigr] \right. \\
& \qquad \left. \times \bigl[(\boldsymbol{\Gamma}_1 \otimes \mathbf{I}_p), (\mathbf{I}_q \otimes \boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma}_1)\bigr] \right\}^{+} \\
& \times \begin{bmatrix} \boldsymbol{\Gamma}_1' \otimes \mathbf{I}_p \\ \mathbf{I}_q \otimes \boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{XY} \end{bmatrix},
\end{aligned}
\tag{5.1}
$$

where $[\ ]^+$ denotes the Moore–Penrose inverse. That (5.1) is equivalent to (3.20) can be shown conveniently by transforming (5.1) to the canonical variable framework as

$$
\begin{aligned}
& \bigl[(\mathbf{I}_{(k)} \otimes \mathbf{I}_p), (\mathbf{I}_q \otimes \overline{\mathbf{R}}\mathbf{I}_{(k)})\bigr] \\
& \times \begin{Bmatrix}
\mathbf{I}_k \otimes \begin{bmatrix} (\mathbf{I}_k - \mathbf{R}_1^2)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-k} \end{bmatrix} & \mathbf{I}_k \otimes \begin{bmatrix} \mathbf{R}_1(\mathbf{I}_k - \mathbf{R}_1^2)^{-1} \\ \mathbf{0} \end{bmatrix} & \mathbf{0} \\
\mathbf{I}_k \otimes \bigl[\mathbf{R}_1(\mathbf{I}_k - \mathbf{R}_1^2)^{-1}, \mathbf{0}\bigr] & \mathbf{I}_k \otimes \mathbf{R}_1^2(\mathbf{I}_k - \mathbf{R}_1^2)^{-1} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{q-k} \otimes \mathbf{R}_1^2(\mathbf{I}_k - \mathbf{R}_1^2)^{-1}
\end{Bmatrix}^{+} \\
& \times \begin{bmatrix} \mathbf{I}_{(k)}' \otimes \mathbf{I}_p \\ \mathbf{I}_q \otimes \mathbf{I}_{(k)}'\overline{\mathbf{R}} \end{bmatrix};
\end{aligned}
\tag{5.2}
$$

the Moore–Penrose inverse can be calculated explicitly. The treatment of Stoica and Viberg does not show that these covariances hold when the $\mathbf{Z}$'s are not normal or when the $\mathbf{X}$'s are stochastic.

Reinsel and Velu (1998) also parameterize $\mathbf{B}$ as $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$ (my notation), where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$ are $p \times k$ and normalized so that $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{ZZ}\boldsymbol{\Phi}_1$, and $\boldsymbol{\Pi} = \mathbf{B}'\boldsymbol{\Phi}_1$. The estimator of $\boldsymbol{\Lambda}$, say $\mathbf{L}$, that Reinsel and Velu defines (page 41) satisfies

$$
\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\boldsymbol{\Sigma}_{ZZ}^{-1}\mathbf{L} = \mathbf{L}\mathbf{T}_1, \qquad \mathbf{L}'\boldsymbol{\Sigma}_{ZZ}^{-1}\mathbf{L} = \mathbf{I}_k,
\tag{5.3}
$$

but the estimator $\hat{\mathbf{\Lambda}} = \mathbf{S}_{ZZ}\hat{\mathbf{\Phi}}_1$ defined in this paper satisfies (5.3) with $\mathbf{\Sigma}_{ZZ}$ replaced by $\mathbf{S}_{\hat{Z}\hat{Z}}$. Similarly, their estimator of $\mathbf{\Pi}$, say $\mathbf{P}$, satisfies

$$\hat{\mathbf{B}}'\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{S}_{YX}\mathbf{P} = \mathbf{P}\mathbf{T}_1, \qquad \mathbf{P}'\left(\hat{\mathbf{B}}'\mathbf{\Sigma}_{ZZ}\hat{\mathbf{B}}\right)^{-1}\mathbf{P} = \mathbf{I}_k, \tag{5.4}$$

while $\hat{\mathbf{\Pi}} = \hat{\mathbf{B}}'\hat{\mathbf{\Phi}}$, satisfies (5.4) with $\mathbf{\Sigma}_{ZZ}$ replaced by $\mathbf{S}_{\hat{Z}\hat{Z}}$.

The difference between the definitions of the pair $\mathbf{L}$ and $\mathbf{P}$ and the pair $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Pi}}$ is that $\mathbf{L}$ and $\mathbf{P}$ are defined in terms of $\mathbf{\Sigma}_{ZZ}$ known, while $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Pi}}$ are defined in terms of $\mathbf{S}_{\hat{Z}\hat{Z}}$, the estimator of $\mathbf{\Sigma}_{ZZ}$. However, as shown in Sections 3 and 4, the asymptotic distribution of the estimator $\mathbf{LP}'$ for $\mathbf{L}$ and $\mathbf{P}$ defined by (5.3) and (5.4) is the same as the asymptotic distribution of $\hat{\mathbf{B}}_k = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Pi}}'$. Reinsel and Velu [(1998), page 84] also obtain the expression (5.1) as the covariance matrix of the limiting distribution of their estimator. [They have provided the author with a direct verification that (5.1) is equivalent to (3.20).]

Reinsel and Velu [(1998), page 45] also approach the asymptotic distribution of $\mathbf{LP}'$ in an alternative way by finding the limiting distribution of $\mathbf{L}^* = \sqrt{n}(\mathbf{L} - \mathbf{\Lambda})$ and $\mathbf{P}^* = \sqrt{n}(\mathbf{P} - \mathbf{\Pi})$. Although their limiting distribution of $\mathbf{L}^*$ and $\mathbf{P}^*$ disagrees with Anderson (1999) because the variability in $\mathbf{S}_{XX}$ was neglected [e.g., in the first asymptotic covariance in Theorem 2.4 the coefficient $\lambda_j^2 + \lambda_i^2$ (in their notation) should be $\lambda_j^2 + \lambda_i^2 + \lambda_i^2\lambda_j^2$], this approach leads to the correct limiting distribution of $\hat{\mathbf{B}}_k$ by the results of Section 4 ($\mathbf{X}$ nonstochastic). However, they only give an explicit expression for the case of $k = 1$ [which agrees with (3.20)].

Velu, Reinsel and Wichern (1986) give the same asymptotic covariances for $\mathbf{L}$ and $\mathbf{P}$.

Lütkepohl (1993) repeats the incorrect asymptotic distribution of $\mathbf{L}$ and $\mathbf{P}$ in Velu, Reinsel and Wichern and asserts that the asymptotic distribution of $\mathbf{LP}'$ is that of $\mathbf{L\Pi}' + \mathbf{\Lambda P}'$, but does not explicitly calculate the covariance matrix.

5.2. *Further comments.* As a measure of the accuracy of the estimator $\tilde{\mathbf{B}}$, we might consider $\lim_{n\to\infty} n \operatorname{tr} \mathscr{E}(\tilde{\mathbf{B}} - \mathbf{B})\mathbf{\Sigma}_{XX}(\tilde{\mathbf{B}} - \mathbf{B})'\mathbf{\Sigma}_{YY}^{-1}$. For the estimators $\hat{\mathbf{B}}_k$ and $\hat{\mathbf{B}}$, the measure is

$$\begin{aligned}\lim_{n\to\infty} n \operatorname{tr} \mathscr{E}\left(\hat{\mathbf{B}}_k - \mathbf{B}\right)\mathbf{\Sigma}_{XX}\left(\hat{\mathbf{B}}_k - \mathbf{B}\right)'\mathbf{\Sigma}_{YY}^{-1} &= \lim_{n\to\infty} \mathscr{E} \operatorname{tr} \hat{\mathbf{\Psi}}_k^* \hat{\mathbf{\Psi}}_k^{*\prime} \\ &= \mathscr{E}\left(\operatorname{vec} \hat{\mathbf{\Psi}}_k^*\right)' \operatorname{vec} \hat{\mathbf{\Psi}}_k^* \\ &= q\sum_{i=1}^k \left(1 - \rho_i^2\right) + k(p-k),\end{aligned} \tag{5.5}$$

$$\lim_{n\to\infty} n \operatorname{tr} \mathscr{E}(\hat{\mathbf{B}} - \mathbf{B})\mathbf{\Sigma}_{XX}(\hat{\mathbf{B}} - \mathbf{B})'\mathbf{\Sigma}_{YY}^{-1} = q\sum_{i=1}^k \left(1 - \rho_i^2\right) + q(p-k). \tag{5.6}$$

If it were known that the rank of $\mathbf{B}$ was k and that $\mathbf{U}_\alpha^{(1)} = (U_{1\alpha}, \dots, U_{k\alpha})'$ and $\mathbf{V}_\alpha^{(1)} = (V_{l\alpha}, \dots, V_{k\alpha})'$ were the canonical variables with positive canonical correlations, the estimator would be

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \mathbf{S}_{UV}^{11}\left(\mathbf{S}_{VV}^{11}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.7}$$

The error in the upper left-hand corner of $\tilde{\boldsymbol{\Psi}}$ is

$$\begin{aligned} \mathbf{S}_{UV}^{11}\left(\mathbf{S}_{VV}^{11}\right)^{-1} - \overline{\mathbf{R}}_1 &= \left(\mathbf{S}_{UV}^{11} - \overline{\mathbf{R}}_1 \mathbf{S}_{VV}^{11}\right)\left(\mathbf{S}_{VV}^{11}\right)^{-1} \\ &= \mathbf{S}_{WV}^{11}\left(\mathbf{S}_{VV}^{11}\right)^{-1} \\ &= \frac{1}{\sqrt{n}} \mathbf{S}_{WV}^{*11}\left(\mathbf{I}_k + \frac{1}{\sqrt{n}} \mathbf{S}_{VV}^{*11}\right)^{-1} \\ &= \frac{1}{\sqrt{n}} \mathbf{S}_{WV}^{*11} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{5.8}$$

The error in the other three submatrices in (5.8) would be zero. Of course, this estimator is not feasible, but it shows what use could be made of prior information.

Note that the number of elements in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Pi}$ is $(p+q)k$, but k^2 restrictions can be imposed to eliminate the indeterminacy implied by $\boldsymbol{\Lambda}\boldsymbol{\Pi}' = \boldsymbol{\Lambda}\mathbf{M}'(\boldsymbol{\Pi}\mathbf{M}^{-1})'$, resulting in $(p+q-k)k$ coordinates in $\mathbf{B} = \boldsymbol{\Lambda}\boldsymbol{\Pi}'$. This number, which is the number of elements in $\boldsymbol{\Psi}_k^*$ not $o_p(1)$, can be much smaller than pq, the number of elements in $\mathbf{B}$.

Since the limiting distribution of $\hat{\mathbf{B}}_k^*$ under general conditions is the same as for normally distributed errors, confidence regions and test procedures for $\mathbf{B}$ based on normal theory can also be used for nonnormal errors.

Acknowledgment. The author thanks Yasuo Amemiya and Tore Schweder for helpful suggestions.

REFERENCES

Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced-rank autoregressive models for multiple time series. *J. Amer. Statist. Assoc.* **83** 849–856.

Anderson, T. W. (1951a). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22** 327–351. [Correction: (1980) *Ann. Statist.* **8** 1400.]

Anderson, T. W. (1951b). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 103–130 Univ. California Press, Berkeley.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis. *J. Multivariate Anal.* **70** 1–29.

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

IZENMAN, A. J. (1975). Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal.* **5** 248–264.

JOHANSEN, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford Univ. Press.

LÜTKEPOHL, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer, New York.

REINSEL, G. C. and VELU, R. P. (1998). *Multivariate Reduced-rank Regression*. Springer, New York.

RYAN, D. A. J., HUBERT, J. J., CARTER, E. M., SPRAGUE, J. B. and PARROTT, J. (1992). A reduced-rank multivariate regression approach to joint toxicity experiments. *Biometrics* **48** 155–162.

SCHMIDLI, H. (1996). *Reduced-rank Regression*. Physica, Berlin.

STOICA, P. and VIBERG, M. (1996). Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Trans. Signal Processing* **44** 3069–3078.

TSAY, R. S. and TIAO, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika* **72** 299–315.

VELU, R. P., REINSEL, G. C. and WICHERN, D. W. (1986). Reduced rank models for multiple times series. *Biometrika* **73** 105–118.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065
E-MAIL: twa@stat.stanford.edu