Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 193-205

SPECIFICATION AND MISSPECIFICATION IN REDUCED RANK REGRESSION

By T.W. ANDERSON
Stanford University, USA

SUMMARY. In the classical linear regression model with p dependent variables constituting the vector $\mathbf{Y}$ and q independent variables constituting the vector $\mathbf{X}$ the rank k of the regression matrix $\mathbf{B}$ of $\mathbf{Y}$ on $\mathbf{X}$ may be less than p and q. In that case the estimator of $\mathbf{B}$, called the "reduced rank regression estimator," is composed of the k canonical variables corresponding to the k largest canonical correlations (Anderson, 1951). The asymptotic distribution of this estimator has been found when the rank is correctly specified and $\mathbf{X}$ and the residual $\mathbf{Y} - \mathbf{BX}$ are independent with finite variances (Anderson, 1999b). The reduced rank regression estimator is more efficient than the least squares estimator, markedly so if k is small. This paper considers the properties of the estimator when the rank is not correctly specified. When the specified rank of $\mathbf{B}$ is less than the true rank, biases occur; when the specified rank is greater, variances of estimators and predictors are unnecessarily large. These results are related to tests concerning the rank.

1. **Introduction**

A frequently used model of statistical analysis of two sets of variables is the linear regression model

$$\mathbf{Y} = \mathbf{BX} + \mathbf{Z}. \tag{1.1}$$

Here $\mathbf{Y}$ and $\mathbf{X}$ are observable vector variables of p and q components, respectively, and $\mathbf{Z}$ is an unobservable error vector of p components, uncorrelated with the components of $\mathbf{X}$. The usual estimator of the matrix $\mathbf{B}$ from a set

Paper received October 2000; revised January 2002.
AMS (2000) subject classification. Primary 62H12; secondary 62J05.
Keywords and phrases. Canonical correlations and variates, reduced rank regression, maximum likelihood estimators, test of rank.

of observations on $\mathbf{Y}$ and $\mathbf{X}$ is the least squares estimator. However, when some further knowledge about the matrix is available, it may be useful to consider modifications of the least squares estimator that take into account the additional information.

There are situations in which it is known or surmised that the rank of $\mathbf{B}$ is less than the orders p and q. In macroeconomic models of simultaneous equations, for example, the rank of the "reduced form" (the matrix $\mathbf{B}$ in (1.1)) or some submatrix is implied by economic theory to be of lower rank. The reduced rank regression estimator was introduced by Anderson (1951) to estimate the reduced form of assigned rank by maximum likelihood under the assumption that $\mathbf{Z}$ is normally distributed. (The limited information maximum likelihood estimator of a single simultaneous equation [Anderson and Rubin (1949)] is based on a special case.) The reduced rank regression estimator was used in cointegrated models [Johansen (1995)], where the rank is based on the dimension of the nonstationary part of the process.

The rank may alternatively be determined from the data. Tsay and Tsiao (1985) and Ahn and Reinsel (1988) have applied the estimator in stationary time series analysis, and Stoica and Viberg (1996) in signal processing. See Reinsel and Velu (1998) for a comprehensive treatment.

More generally, the independent variables may be divided into two sets, say $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$ and $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$. The model (1.1) is written as

$$\mathbf{Y} = \mathbf{B}_1\mathbf{X}_1 + \mathbf{B}_2\mathbf{X}_2 + \mathbf{Z} \ . \tag{1.2}$$

When $\mathbf{B}_1$ is assumed of lower rank, it is estimated by the reduced rank regression of $\mathbf{Y}$ on $\mathbf{X}_1$ conditional on $\mathbf{X}_2$. In this paper the model (1.1) will be treated first; later modification of the procedures to the more general model (1.2) are made.

The focus of the study is on the effects of misspecification of the model, specifically, the effects of under- and over-estimating the rank of $\mathbf{B}$. The rank of $\mathbf{B}$ is the number of positive canonical correlations between $\mathbf{Y}$ and $\mathbf{X}$. In the first part of the study $\mathbf{X}$ is a random vector of q components with mean $\mathcal{E}\mathbf{X} = \mathbf{0}$ and covariance matrix $\mathcal{E}\mathbf{X}\mathbf{X}' = \mathbf{\Sigma}_{XX}$. The unobserved error vector $\mathbf{Z}$ is a random vector of p components with $\mathcal{E}\mathbf{Z} = \mathbf{0}$, $\mathcal{E}\mathbf{Z}\mathbf{Z}' = \mathbf{\Sigma}_{ZZ}$, and $\mathcal{E}\mathbf{Z}\mathbf{X}' = \mathbf{0}$. Then $\mathcal{E}\mathbf{Y}\mathbf{X}' = \mathbf{\Sigma}_{YX} = \mathbf{B}\mathbf{\Sigma}_{XX}$ and $\mathcal{E}\mathbf{Y}\mathbf{Y}' = \mathbf{\Sigma}_{YY} = \mathbf{B}\mathbf{\Sigma}_{XX}\mathbf{B}' + \mathbf{\Sigma}_{ZZ}$. Suppose $p \leq q$.

Observations $(\mathbf{y}_1, \mathbf{x}_1), \ldots, (\mathbf{y}_N, \mathbf{x}_N)$ are made on $\mathbf{Y}, \mathbf{X}$. We define the sample matrices

$$\mathbf{S}_{YY} = \frac{1}{N}\sum_{\alpha=1}^{N} \mathbf{y}_\alpha\mathbf{y}_\alpha', \ \mathbf{S}_{XX} = \frac{1}{N}\sum_{\alpha=1}^{N} \mathbf{x}_\alpha\mathbf{x}_\alpha', \ \mathbf{S}_{YX} = \frac{1}{N}\sum_{\alpha=1}^{N} \mathbf{y}_\alpha\mathbf{x}_\alpha' \ . \tag{1.3}$$

The assumption $\mathcal{E}\mathbf{X} = \mathbf{0}$ is made for convenience. In practice the sample covariances would be calculated as $N^{-1}\Sigma_{\alpha=1}^{N}(\mathbf{y}_\alpha - \bar{\mathbf{y}})(\mathbf{y}_\alpha - \bar{\mathbf{y}})'$, etc., where $\bar{\mathbf{y}} = N^{-1}\Sigma_{\alpha=1}^{N}\mathbf{y}_\alpha$.

2. Canonical Correlations and Variates

The basis for reduced rank regression is canonical analysis. [See Anderson (1984), Chapter 12, for example.] The first canonical correlation between $\mathbf{Y}$ and $\mathbf{X}$ is the maximum correlation between linear combinations of $\mathbf{Y}$ and $\mathbf{X}$, $\rho = \text{correl}(\boldsymbol{\alpha}'\mathbf{Y}, \boldsymbol{\gamma}'\mathbf{X})$. The maximization procedure leads to the equation

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

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are normalized by

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{YY}\boldsymbol{\alpha} = 1 = \boldsymbol{\gamma}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\gamma}. \tag{2.2}$$

For a solution to (2.1) ρ must satisfy

$$\begin{vmatrix} -\rho\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & -\rho\boldsymbol{\Sigma}_{XX} \end{vmatrix} = 0. \tag{2.3}$$

Number the p largest solutions as $\rho_1 \geq \rho_2 \geq \ldots \geq \rho_p \geq 0$. Let the solutions to (2.1) and (2.2) constitute the matrices $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_p] = \mathbf{A}$, $[\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_q] = \boldsymbol{\Gamma}$, $\text{diag}(\rho_1, \rho_2, \ldots, \rho_p) = \mathbf{R}$.

We shall assume that the rank of $\boldsymbol{\Sigma}_{YX}$ is k $(\leq p)$ and $\rho_1 > \ldots > \rho_k$; then the solution of (2.1) and (2.2) is unique for such a value of ρ except for multiplication by -1. To eliminate this indeterminacy we shall require that $\alpha_{ii} > 0$, $i = 1, \ldots, k$. (Since $(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k)$ has rank k, the components can be numbered so $\alpha_{ii} \neq 0$, $i = 1, \ldots, k$.) We shall take $\boldsymbol{\alpha}_{k+1}, \ldots \boldsymbol{\alpha}_p$ and $\boldsymbol{\gamma}_{k+1}, \ldots, \boldsymbol{\gamma}_q$ so $\mathbf{A}'\boldsymbol{\Sigma}_{YY}\mathbf{A} = \mathbf{I}$, $\boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{XX}\boldsymbol{\Gamma} = \mathbf{I}$, $\mathbf{A}'\boldsymbol{\Sigma}_{YX}\boldsymbol{\Gamma} = \bar{\mathbf{R}} = (\mathbf{R}, \mathbf{0})$.

The sample canonical correlations and variates are defined by (2.1), (2.2), and (2.3) with $\boldsymbol{\Sigma}_{YY}, \boldsymbol{\Sigma}_{XY}$, and $\boldsymbol{\Sigma}_{XX}$ replaced by $\mathbf{S}_{YY}, \mathbf{S}_{XY}$, and $\mathbf{S}_{XX}$, respectively. Let the p largest roots satisfy $r_1 > r_2 > \ldots > r_p > 0$. Then $(r_1, \ldots, r_p)$ estimates $(\rho_1, \ldots, \rho_p)$ and $[\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \ldots, \hat{\boldsymbol{\alpha}}_p] = \hat{\mathbf{A}}$ and $\left[\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_q\right] = \hat{\boldsymbol{\Gamma}}$ estimate $\mathbf{A}$ and $\boldsymbol{\Gamma}$, respectively. Elimination of $\hat{\boldsymbol{\alpha}}$ from the sample version of (2.1) yields the equations for $\hat{\boldsymbol{\gamma}}$

$$\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\hat{\boldsymbol{\gamma}} = r^2\mathbf{S}_{XX}\hat{\boldsymbol{\gamma}}, \qquad \hat{\boldsymbol{\gamma}}'\mathbf{S}_{XX}\hat{\boldsymbol{\gamma}} = 1\ , \tag{2.4}$$

and r^2 satisfies $\left|\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX} - r^2\mathbf{S}_{XX}\right| = 0$.

3. Reduced Rank Regression and the Likelihood Ratio Criterion for Rank

The least squares (LS) estimator of $\mathbf{B}$ is $\widehat{\mathbf{B}} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}$. In order to define the reduced rank regression (RRR) estimator of rank k define $\widehat{\mathbf{\Gamma}}_1 = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_k)$. Then the RRR estimator is

$$\widehat{\mathbf{B}}_k = \mathbf{S}_{YX}\widehat{\mathbf{\Gamma}}_1\widehat{\mathbf{\Gamma}}_1' \ . \tag{3.1}$$

Anderson (1951) derived $\widehat{\mathbf{B}}_k$ as the maximum likelihood estimator of $\mathbf{B}$ when the rank of $\mathbf{B}$ is specified to be k $(0 \leq k \leq p)$, the $\mathbf{Z}_\alpha$ are independently distributed as $N(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$, and the $\mathbf{X}$'s are nonstochastic. The derivation is valid for the $\mathbf{X}$'s distributed independently of the $\mathbf{Z}$'s.

The likelihood ratio criterion for testing the null hypothesis that the rank of $\mathbf{B}$ is k against the alternative that it is greater than k when the $\mathbf{Z}_t$ are independently distributed as $N(\mathbf{0}, \mathbf{\Sigma}_{ZZ})$, is

$$-2\log\lambda = -N\sum_{i=k+1}^{p}\log(1-r_i^2) \sim N\sum_{i=k+1}^{p} r_i^2 \tag{3.2}$$

[Anderson (1951)]. The null hypothesis is rejected if $-2\log\lambda$ is large. Under the null hypothesis $-2\log\lambda$ has a limiting χ^2-distribution with $(p-k)(q-k)$ degrees of freedom.

In order to describe the asymptotic distributions of $\widehat{\mathbf{B}}_k$ and $\widehat{\mathbf{B}}$ we shall use the so-called vec notation. For $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_n)$ define $\operatorname{vec}\mathbf{C} = (\mathbf{c}_1', \ldots, \mathbf{c}_n')'$; define the Kronecker product of two matrices $\mathbf{C}$ and $\mathbf{D}$ as $\mathbf{C}\otimes\mathbf{D} = (c_{ij}\mathbf{D})$. The Kronecker product satisfies $(\mathbf{C}\otimes\mathbf{D})(\mathbf{F}\otimes\mathbf{G}) = \mathbf{CF}\otimes\mathbf{DG}$, $\operatorname{vec}(\mathbf{CDF}) = (\mathbf{F}'\otimes\mathbf{C})\operatorname{vec}\mathbf{D}$, $\operatorname{vec}\mathbf{x}\mathbf{y}' = \mathbf{y}\otimes\mathbf{x}$.

We can write $\widehat{\mathbf{B}}=\mathbf{B}+\mathbf{S}_{ZX}\mathbf{S}_{XX}^{-1}$. Then $\operatorname{vec}\left(\widehat{\mathbf{B}}-\mathbf{B}\right) = \left(\mathbf{S}_{XX}^{-1}\otimes\mathbf{I}_p\right)\operatorname{vec}\mathbf{S}_{ZX}$. Define

$$\mathbf{S}_{ZX}^{*} = \sqrt{N}\mathbf{S}_{ZX}, \qquad \widehat{\mathbf{B}}^{*} = \sqrt{N}(\widehat{\mathbf{B}}-\mathbf{B}) = \mathbf{S}_{ZX}^{*}\mathbf{S}_{XX}^{-1}. \tag{3.3}$$

By the law of large numbers $\mathbf{S}_{XX} \xrightarrow{p} \mathbf{\Sigma}_{XX}$. By a central limit theorem $\operatorname{vec}\mathbf{S}_{ZX}^{*} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}_{XX}\otimes\mathbf{\Sigma}_{ZZ})$. These facts imply

$$\operatorname{vec}\widehat{\mathbf{B}}^{*} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}_{XX}^{-1}\otimes\mathbf{\Sigma}_{ZZ}). \tag{3.4}$$

To study the distribution of $\widehat{\mathbf{B}}_k$ we transform the model and statistics to a canonical form. Let $\mathbf{U}_\alpha=\mathbf{A}'\mathbf{Y}_\alpha, \mathbf{W}_\alpha=\mathbf{A}'\mathbf{Z}_\alpha, \mathbf{V}_\alpha=\mathbf{\Gamma}'\mathbf{X}_\alpha$. Then

$\mathbf{\Sigma}_{UU}$=$\mathbf{I}$, $\mathbf{\Sigma}_{VV}$=$\mathbf{I}$, $\mathbf{\Sigma}_{UV} = \mathbf{A}'\mathbf{\Sigma}_{YX}\mathbf{\Gamma} = (\mathbf{R}, \mathbf{0})$. The transformed observations constitute a sample from the model

$$\mathbf{U} = \mathbf{\Psi}\mathbf{V} + \mathbf{W}, \tag{3.5}$$

where $\mathbf{\Psi} = \mathbf{A}'\mathbf{B}(\mathbf{\Gamma}')^{-1} = (\mathbf{R}, \mathbf{0})$. The LS and RRR estimators of $\mathbf{\Psi}$ are

$$\widehat{\mathbf{\Psi}} = \mathbf{S}_{UV}\mathbf{S}_{VV}^{-1}, \qquad \widehat{\mathbf{\Psi}}_k = \mathbf{S}_{UV}\mathbf{H}_1\mathbf{H}_1', \tag{3.6}$$

respectively, where $\mathbf{H}_1 = \mathbf{\Gamma}^{-1}\widehat{\mathbf{\Gamma}}_1$ satisfies

$$\mathbf{S}_{VU}\mathbf{S}_{UU}^{-1}\mathbf{S}_{UV}\mathbf{H}_1 = \mathbf{S}_{VV}\mathbf{H}_1\widehat{\mathbf{R}}_1^2, \qquad \mathbf{H}_1'\mathbf{S}_{VV}\mathbf{H}_1 = \mathbf{I}\,. \tag{3.7}$$

Here $\widehat{\mathbf{R}}_1 = \mathrm{diag}(r_1, \ldots, r_k)$.

We need the limiting distribution of $\widehat{\mathbf{\Psi}}_k^* = \sqrt{N}(\widehat{\mathbf{\Psi}}_k - \mathbf{\Psi})$ [Anderson (1999b)] when the rank is correctly specified. Define $\mathbf{S}_{UU}^* = \sqrt{N}(\mathbf{S}_{UU} - \mathbf{I})$, $\mathbf{S}_{VV}^* = \sqrt{N}\,(\mathbf{S}_{VV} - \mathbf{\Sigma}_{VV})$, $\mathbf{S}_{UV}^* = \sqrt{N}\,(\mathbf{S}_{UV} - \mathbf{\Sigma}_{UV})$, $\mathbf{R}_1^* = \sqrt{N}(\widehat{\mathbf{R}}_1 - \mathbf{R}_1)$, $\mathbf{R}_1 = \mathrm{diag}(\rho_1, \ldots, \rho_k)$, and $\mathbf{H}_1^* = \sqrt{N}(\mathbf{H}_1 - \mathbf{I}_{(k)}) = \left[\mathbf{H}_{11}^{*\prime}, \mathbf{H}_{21}^{*\prime}\right]'$, where $\mathbf{I}_{(k)} = [\mathbf{I}_k, \mathbf{0}]'$. The expansion of (3.7) in terms of $\mathbf{S}_{VU}^* = \sqrt{N}\,(\mathbf{S}_{VU} - \mathbf{\Sigma}_{UV})$, $\mathbf{H}_1^*$ and $\mathbf{R}_1^*$, partitioned into k and $q-k$ rows, is

$$\begin{aligned}\begin{bmatrix} 2\mathbf{R}_1\mathbf{R}_1^* + \mathbf{H}_{11}^*\mathbf{R}_1^2 - \mathbf{R}_1^2\mathbf{H}_{11}^* \\ \mathbf{H}_{21}^*\mathbf{R}_1^2 \end{bmatrix} &= \begin{bmatrix} \mathbf{S}_{VU}^{*11}\mathbf{R}_1 + \mathbf{R}_1\mathbf{S}_{UV}^{*11} - \mathbf{R}_1\mathbf{S}_{UU}^{*11}\mathbf{R}_1 - \mathbf{S}_{VV}^{*11}\mathbf{R}_1^2 \\ \mathbf{S}_{VU}^{*21}\mathbf{R}_1 - \mathbf{S}_{VV}^{*21}\mathbf{R}_1^2 \end{bmatrix} \\ &\quad + o_p(1), \end{aligned} \tag{3.8}$$

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

Then expansion of $\widehat{\mathbf{\Psi}}_k^*$ in terms of $\mathbf{S}_{UV}^*$ and $\mathbf{H}_1^*$ gives

$$\widehat{\mathbf{\Psi}}_k^* = \sqrt{N}(\widehat{\mathbf{\Psi}}_k - \mathbf{\Psi}) = \begin{bmatrix} \mathbf{S}_{WV}^{*11} & \mathbf{S}_{WV}^{*12} \\ \mathbf{S}_{WV}^{*21} & \mathbf{0} \end{bmatrix} + o_p(1). \tag{3.10}$$

Note that the lower right-hand corner of $\widehat{\mathbf{\Psi}}_k^*$ is $\mathbf{0}$; that is, there is no error (of order $1/\sqrt{N}$) in the estimation of the lower right-hand corner of $\mathbf{\Psi}$. In contrast the least squares estimator is

$$\widehat{\mathbf{\Psi}}^* = \begin{bmatrix} \mathbf{S}_{WV}^{*11} & \mathbf{S}_{WV}^{*12} \\ \mathbf{S}_{WV}^{*21} & \mathbf{S}_{WV}^{*22} \end{bmatrix} + o_p(1). \tag{3.11}$$

Because the estimator $\widehat{\mathbf{\Psi}}_k$ depends asymptotically only on $\mathbf{S}_{WV}^*$, the limiting normal distribution of $\widehat{\mathbf{\Psi}}_k^*$ is valid under the conditions for which

the least squares estimator is asymptotically normally distributed. One such set of conditions is that $\mathbf{X}$ and $\mathbf{Z}$ are independently distributed with finite second-order moments.

When the estimators in the canonical form are transformed back to the original coordinates, we obtain the following theorem:

THEOREM 1 *Let* $\mathbf{Y}$ *and* $\mathbf{X}$ *be related by* (1.1)*, where* $\mathcal{E}\mathbf{X}\mathbf{X}' = \mathbf{\Sigma}_{XX}, \mathcal{E}\mathbf{Z}\mathbf{Z}' = \mathbf{\Sigma}_{ZZ}$*, and* $\mathcal{E}\mathbf{X}\mathbf{Z}' = \mathbf{0}$*. Then* $\hat{\mathbf{B}}_k^* = \sqrt{N}(\widehat{\mathbf{B}}_k - \mathbf{B})$ *has a limiting normal distribution with mean* $\mathbf{0}$*; the covariance of vec*$\widehat{\mathbf{B}}_k^*$ *in the limiting distribution is*

$$\mathbf{\Sigma}_{XX}^{-1} \otimes \mathbf{\Sigma}_{ZZ} - \left[\mathbf{\Sigma}_{XX}^{-1} - \mathbf{\Pi}(\mathbf{\Pi}'\mathbf{\Sigma}_{XX}\mathbf{\Pi})^{-1}\mathbf{\Pi}'\right] \otimes \left[\mathbf{\Sigma}_{ZZ} - \mathbf{\Lambda}(\mathbf{\Lambda}'\mathbf{\Sigma}_{ZZ}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\right], \tag{3.12}$$

where $\mathbf{B} = \mathbf{\Lambda}\mathbf{\Pi}'$ *is any factorization.*

A particular factorization is $\mathbf{B} = (\mathbf{\Sigma}_{YX}\mathbf{\Gamma}_1)\,\mathbf{\Gamma}_1'$, in which case $\mathbf{\Gamma}_1'\mathbf{\Sigma}_{XX}\mathbf{\Gamma}_1 = \mathbf{I}_k$ and $(\mathbf{\Sigma}_{YX}\mathbf{\Gamma}_1)'\,\mathbf{\Sigma}_{ZZ}^{-1}\,(\mathbf{\Sigma}_{YX}\mathbf{\Gamma}_1) = \mathbf{R}_1^2\,(\mathbf{I} - \mathbf{R}_1^2)^{-1}$.

COROLLARY. *The conclusion of the theorem holds if* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *are nonstochastic and* $\mathbf{S}_{XX} \to \mathbf{\Sigma}_{XX}$ *or if* $\mathbf{\Sigma}_{ZZ}$ *is replaced by* $\operatorname{plim}_{N\to\infty}\mathbf{S}_{ZZ}$.

4. Misspecification: Underestimation of Regression Rank

In Sections 2 and 3 it was supposed that the rank of $\mathbf{B}$ was known and that the number of columns of $\mathbf{\Gamma}_1$ was that rank. In this section we suppose that the number of columns used to define the reduced rank regression is $s\,(< k)$. Let $\widehat{\mathbf{\Gamma}} = (\widehat{\mathbf{\Gamma}}_a, \widehat{\mathbf{\Gamma}}_b, \widehat{\mathbf{\Gamma}}_2) = \mathbf{\Gamma}(\mathbf{H}_a, \mathbf{H}_b, \mathbf{H}_2)$ of s, $k - s$, and $q - k$ columns, respectively. Let $\mathbf{R} = \operatorname{diag}(\mathbf{R}_a, \mathbf{R}_b, \mathbf{0})$, where $\mathbf{R}_a$ is $s \times s$ and $\mathbf{R}_b$ is $(k - s) \times (k - s)$. The estimator of $\mathbf{B}$ is $\widehat{\mathbf{B}}_s = \mathbf{S}_{YX}\widehat{\mathbf{\Gamma}}_a\widehat{\mathbf{\Gamma}}_a'$ and of $\mathbf{\Psi}$ is $\widehat{\mathbf{\Psi}}_s = \mathbf{S}_{UV}\mathbf{H}_a\mathbf{H}_a'$, where $\widehat{\mathbf{\Gamma}}_a$ and $\mathbf{H}_a$ have s columns. The probability limit of the replacement of the first part of (3.7) and the distinctness of the k larger ρ's implies $\mathbf{H}_a$ converges stochastically to a diagonal matrix, and the probability limit of the replacement of the second part of (3.7) and $h_{ii} > 0$ implies $\mathbf{H}_a \xrightarrow{p} \mathbf{I}_{(s)}$. Then

$$\widehat{\mathbf{\Psi}}_s \xrightarrow{p} \mathbf{\Psi}_s = \begin{bmatrix} \mathbf{R}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{4.1}$$

The asymptotic bias of $\widehat{\mathbf{\Psi}}_s$ is

$$\mathbf{\Psi}_s - \mathbf{\Psi} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{4.2}$$

which depends only on the consistency of the estimators. In the original coordinates the asymptotic bias is $(\mathbf{A}')^{-1}(\mathbf{\Psi}_s - \mathbf{\Psi})\mathbf{\Gamma}' = -\mathbf{\Sigma}_{YY}\mathbf{A}_b\mathbf{R}_b\mathbf{\Gamma}_b'$, where $\mathbf{A}$ and $\mathbf{\Gamma}$ have been partitioned as $\mathbf{A} = (\mathbf{A}_a, \mathbf{A}_b, \mathbf{A}_2)$ and $\mathbf{\Gamma} = (\mathbf{\Gamma}_a, \mathbf{\Gamma}_b, \mathbf{\Gamma}_2)$. Note that the diagonal elements of (diagonal) $\mathbf{R}_b$ are the $k-s$ smaller nonzero roots.

Define $\mathbf{H}_a^* = \sqrt{N}\,(\mathbf{H}_a - \mathbf{I}_{(s)}) = (\mathbf{H}_{aa}^{*\prime}, \mathbf{H}_{ba}^{*\prime}, \mathbf{H}_{2a}^{*\prime})'$. The substitution in (3.7) yields (3.8) with k replaced by s and 1 by a. The partitioning of the $q \times s$ replacement of the first part of (3.7) into s, $k-s$, and $q-k$ rows is

$$\begin{bmatrix} 2\mathbf{R}_a\mathbf{R}_a^* + \mathbf{H}_{aa}^*\mathbf{R}_a^2 - \mathbf{R}_a^2\mathbf{H}_{aa}^* \\ \mathbf{H}_{ba}^*\mathbf{R}_a^2 - \mathbf{R}_b^2\mathbf{H}_{ba}^* \\ \mathbf{H}_{2a}^*\mathbf{R}_a^2 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{VU}^{*aa}\mathbf{R}_a + \mathbf{R}_a\mathbf{S}_{UV}^{*aa} - \mathbf{R}_a\mathbf{S}_{UU}^{*aa}\mathbf{R}_a - \mathbf{S}_{VV}^{*aa}\mathbf{R}_a^2 \\ \mathbf{S}_{VU}^{*ba}\mathbf{R}_a + \mathbf{R}_b\mathbf{S}_{UV}^{*ba} - \mathbf{R}_b\mathbf{S}_{UU}^{*ba}\mathbf{R}_a - \mathbf{S}_{VV}^{*ba}\mathbf{R}_a^2 \\ \mathbf{S}_{VU}^{*2a}\mathbf{R}_a - \mathbf{S}_{VV}^{*2a}\mathbf{R}_a^2 \end{bmatrix} + o_p(1). \tag{4.3}$$

We obtain from (4.3)

$$\mathbf{H}_{2a}^*\mathbf{R}_a = \mathbf{S}_{VU}^{*2a} - \mathbf{S}_{VV}^{*2a}\mathbf{R}_a + o_p(1) = (\mathbf{S}_{WV}^{*a2})' + o_p(1). \tag{4.4}$$

From the upper left-hand corner of (3.9) we find $\mathbf{H}_{aa}^* + \mathbf{H}_{aa}^{*\prime} = -\mathbf{S}_{VV}^{*aa} + o_p(1)$. Then from $\widehat{\mathbf{\Psi}}_s = \mathbf{S}_{UV}\mathbf{H}_a\mathbf{H}_a'$ we obtain

$$\begin{aligned} \sqrt{N}\left(\widehat{\mathbf{\Psi}}_s - \mathbf{\Psi}_s\right) &= \begin{bmatrix} \mathbf{S}_{UV}^{*aa} + \mathbf{R}_a(\mathbf{H}_{aa}^* + \mathbf{H}_{aa}^{*\prime}) & \mathbf{R}_a\mathbf{H}_{ba}^{*\prime} & \mathbf{R}_a\mathbf{H}_{2a}^{*\prime} \\ \mathbf{S}_{UV}^{*ba} + \mathbf{R}_b\mathbf{H}_{ba}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{S}_{UV}^{*2a} & \mathbf{0} & \mathbf{0} \end{bmatrix} + o_p(1) \\ &= \begin{bmatrix} \mathbf{S}_{WV}^{*aa} & \mathbf{R}_a\mathbf{H}_{ba}^{*\prime} & \mathbf{S}_{WV}^{*a2} \\ \mathbf{S}_{UV}^{*ba} + \mathbf{R}_b\mathbf{H}_{ba}^* & \mathbf{0} & \mathbf{0} \\ \mathbf{S}_{WV}^{*2a} & \mathbf{0} & \mathbf{0} \end{bmatrix} + o_p(1). \end{aligned} \tag{4.5}$$

The difference of the estimators $\widehat{\mathbf{\Psi}}_s$ and $\widehat{\mathbf{\Psi}}_k$ minus the bias multiplied by $\sqrt{N}$ is

$$\sqrt{N}\left[\widehat{\mathbf{\Psi}}_s - \widehat{\mathbf{\Psi}}_k - (\mathbf{\Psi}_s - \mathbf{\Psi})\right] = \begin{bmatrix} \mathbf{0} & \mathbf{R}_a\mathbf{H}_{ba}^{*\prime} - \mathbf{S}_{WV}^{*ba} & \mathbf{0} \\ \mathbf{R}_b(\mathbf{H}_{ba}^* - \mathbf{S}_{VV}^{*ba}) & -\mathbf{S}_{WV}^{*bb} & -\mathbf{S}_{WV}^{*b2} \\ \mathbf{0} & -\mathbf{S}_{WV}^{*2b} & \mathbf{0} \end{bmatrix} + o_p(1). \tag{4.6}$$

The estimator $\widehat{\mathbf{\Psi}}_k$ has no asymptotic bias; whereas the estimator $\widehat{\mathbf{\Psi}}_s$ has asymptotic bias in the middle $k-s$ rows and columns, but no residual random error in that location. If the canonical correlations $\rho_{s+1}, \ldots, \rho_k$ are small, the bias will not be important.

The limiting distribution of $\widehat{\mathbf{\Psi}}_k^*$ does not require normality of $\mathbf{Z}$ or $\mathbf{X}$ because it depends only on $\mathbf{S}_{WV}^*$. However, the limiting distribution of $\widehat{\mathbf{\Psi}}_s^*$ might depend on higher-order moments of $\mathbf{Z}$ and $\mathbf{X}$ (or $\mathbf{W}$ and $\mathbf{V}$) because the elements of $\mathbf{H}_{ba}^*$ are linear combinations of elements of $\mathbf{S}_{VW}^*$, $\mathbf{S}_{UV}^*$, and $\mathbf{\Sigma}_{VV}^*$.

In the original coordinates the bias is $\mathbf{B}_s - \mathbf{B} = -\mathbf{\Sigma}_{YY}\mathbf{A}_b\mathbf{R}_b\mathbf{\Gamma}_b'$.

5. Misspecification: Overestimation of Regression Rank

Partition $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_c, \mathbf{H}_d)$ of k, t, and $q-(k+t)$ columns. Suppose that the investigator estimates $\mathbf{B}$ of rank k by an estimator of rank $k+t$,

$$\widehat{\mathbf{B}}_{k+t} = \mathbf{S}_{YX}\widehat{\mathbf{\Gamma}}_{1,c}\widehat{\mathbf{\Gamma}}_{1,c}' = (\mathbf{A}')^{-1}\widehat{\mathbf{\Psi}}_{k+t}\mathbf{\Gamma}' = \mathbf{\Sigma}_{YY}\mathbf{A}\widehat{\mathbf{\Psi}}_{k+t}\mathbf{\Gamma}', \tag{5.1}$$

where $\widehat{\mathbf{\Gamma}}_{1,c}$ has $k+t$ columns,

$$\widehat{\mathbf{\Psi}}_{k+t} = \mathbf{S}_{UV}\mathbf{H}_{1,c}\mathbf{H}_{1,c}', \tag{5.2}$$

and $\mathbf{H}_{1,c} = (\mathbf{H}_1, \mathbf{H}_c)$ has $k+t$ columns. We suppose that $\rho_1 > \cdots > \rho_k > \rho_{k+1} = \cdots = \rho_p = 0$, and that $(\mathbf{Y}', \mathbf{X}')$ has a normal distribution. The probability limits of $\mathbf{I} = \mathbf{H}'\mathbf{S}_{VV}\mathbf{H}$, $\mathbf{S}_{VU}\mathbf{S}_{UU}^{-1}\mathbf{S}_{UV}\mathbf{H} = \mathbf{S}_{VV}\mathbf{H}(\widehat{\mathbf{R}}, \mathbf{0})'(\widehat{\mathbf{R}}, \mathbf{0})$, and $h_{ii} > 0$ imply $\mathbf{H}_{11} \xrightarrow{p} \mathbf{I}_k$, $\mathbf{H}_{21} \xrightarrow{p} \mathbf{0}$, $\mathbf{H}_{12} \xrightarrow{p} \mathbf{0}$. (Here $\mathbf{H}$ is partitioned into k and $q-k$ rows and k and $q-k$ columns.) However, $\mathbf{H}_{22}$ does not converge in probability to a constant matrix; that fact reflects the fact that the solutions of (2.1) for $\rho^2 = 0$ and (2.2) are indeterminate. From $\mathbf{I} = \mathbf{H}'\mathbf{S}_{VV}\mathbf{H}$ and $\mathbf{S}_{VV} \xrightarrow{p} \mathbf{I}$, we find that $\mathbf{H}_{22}'\mathbf{H}_{22} \xrightarrow{p} \mathbf{I}$. Let the singular value decomposition of $\mathbf{H}_{22}$ be $\mathbf{H}_{22} = \mathbf{EDF}$, where $\mathbf{D}$ is diagonal (with positive diagonal elements) and $\mathbf{E}$ and $\mathbf{F}$ are orthogonal. Then $\mathbf{L}_2 = \mathbf{EF}$ is orthogonal; define $\mathbf{H}_{22}^* = \sqrt{N}\,(\mathbf{H}_{22} - \mathbf{EF})$. (See Anderson, 1989 and/or 1999a, for details and justification.) Then

$$\mathbf{H}_{22} = \mathbf{L}_2 + \frac{1}{\sqrt{N}}\mathbf{H}_{22}^*. \tag{5.3}$$

Note that $\mathbf{L}_2'\mathbf{H}_{22}^* = \mathbf{H}_{22}^{*\prime}\mathbf{L}_2$.

Let the first t columns of $\mathbf{H}_{22}$, $\mathbf{L}_2$, $\mathbf{H}_{22}^*$ be $\mathbf{H}_{2c}$, $\mathbf{L}_c$, $\mathbf{H}_{2c}^*$, respectively. Then (5.2) is

$$\widehat{\mathbf{\Psi}}_{k+t} = \left[\overline{\mathbf{R}} + \frac{1}{\sqrt{N}}\mathbf{S}_{UV}^*\right]\left[\begin{pmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{L}_c\end{pmatrix} + \frac{1}{\sqrt{N}}(\mathbf{H}_1^*, \mathbf{H}_c^*)\right]\left[\begin{pmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{L}_c'\end{pmatrix} + \frac{1}{\sqrt{N}}\begin{pmatrix}\mathbf{H}_1^{*\prime}\\ \mathbf{H}_c^{*\prime}\end{pmatrix}\right]$$

$$= \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{\sqrt{N}} \Bigg\{ \mathbf{S}^*_{UV} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_c\mathbf{L}_c' \end{bmatrix} + \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} (\mathbf{H}^*_1, \mathbf{H}^*_c) \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_c' \end{bmatrix} + \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{H}^{*\prime}_1 \\ \mathbf{L}_c\mathbf{H}^{*\prime}_c \end{bmatrix} \Bigg\} + o_p\Big(\frac{1}{\sqrt{N}}\Big). \tag{5.4}$$

We obtain from (5.4)

$$\begin{aligned} \widehat{\mathbf{\Psi}}^*_{k+t} &= \sqrt{N}\,(\widehat{\mathbf{\Psi}}_{k+t} - \mathbf{\Psi}_{k+t}) = \sqrt{N}\,(\widehat{\mathbf{\Psi}}_{k+t} - \mathbf{\Psi}) \qquad (5.5) \\ &= \begin{bmatrix} \mathbf{S}^{*11}_{UV} & \mathbf{S}^{*12}_{UV}\mathbf{L}_c\mathbf{L}_c' \\ \mathbf{S}^{*21}_{UV} & \mathbf{S}^{*22}_{UV}\mathbf{L}_c\mathbf{L}_c' \end{bmatrix} + \begin{bmatrix} \mathbf{R}_1\mathbf{H}^*_{11} & \mathbf{R}_1\mathbf{H}^*_{1c}\mathbf{L}_c' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ &\quad + \begin{bmatrix} \mathbf{R}_1\mathbf{H}^{*\prime}_{11} & \mathbf{R}_1\mathbf{H}^{*\prime}_{21} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + o_p(1). \end{aligned}$$

The equation $\mathbf{H}'\mathbf{S}_{VV}\mathbf{H} = \mathbf{I}$ implies

$$\begin{bmatrix} \mathbf{S}^{*11}_{VV}+\mathbf{H}^*_{11}+\mathbf{H}^{*\prime}_{11} & \mathbf{S}^{*12}_{VV}\mathbf{L}_2+\mathbf{H}^*_{12}+\mathbf{H}^{*\prime}_{21}\mathbf{L}_2 \\ \mathbf{L}_2'\mathbf{S}^{*21}_{VV}+\mathbf{L}_2'\mathbf{H}^*_{21}+\mathbf{H}^{*\prime}_{12} & \mathbf{L}_2'\mathbf{S}^{*22}_{VV}\mathbf{L}_2+\mathbf{L}_2'\mathbf{H}^*_{22}+\mathbf{H}^{*\prime}_{22}\mathbf{L}_2 \end{bmatrix} = \mathbf{0} + o_p(1); \tag{5.6}$$

in particular, $\mathbf{H}^*_{11} + \mathbf{H}^{*\prime}_{11} = -\mathbf{S}^{*11}_{VV} + o_p(1)$. From $\mathbf{S}_{VU}\mathbf{S}^{-1}_{UU}\mathbf{S}_{UV}\mathbf{H} = \mathbf{S}_{VV}\mathbf{H}(\widehat{\mathbf{R}}, \mathbf{0})'(\widehat{\mathbf{R}}, \mathbf{0})$, we obtain

$$\begin{aligned} &\begin{bmatrix} 2\mathbf{R}_1\mathbf{R}^*_1 + \mathbf{H}^*_{11}\mathbf{R}^2_1 - \mathbf{R}^2_1\mathbf{H}^*_{11} & -\mathbf{R}^2_1\mathbf{H}^*_{12} \\ \mathbf{H}^*_{21}\mathbf{R}^2_1 & \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{S}^{*11}_{VU}\mathbf{R}_1 - \mathbf{R}_1\mathbf{S}^{*11}_{UU}\mathbf{R}_1 + \mathbf{R}_1\mathbf{S}^{*11}_{UV} - \mathbf{S}^{*11}_{VV}\mathbf{R}^2_1 & \mathbf{R}_1\mathbf{S}^{*12}_{UV}\mathbf{L}_2 \\ \mathbf{S}^{*21}_{VU}\mathbf{R}_1 - \mathbf{S}^{*21}_{VV}\mathbf{R}^2_1 & \mathbf{0} \end{bmatrix} + o_p(1). \end{aligned} \tag{5.7}$$

This equation implies

$$\mathbf{R}_1\mathbf{H}^{*\prime}_{21} + o_p(1) = (\mathbf{S}^{*21}_{VU} - \mathbf{S}^{*21}_{VV}\mathbf{R}_1)' = (\mathbf{S}^{*21}_{VW})' = \mathbf{S}^{*12}_{WV}, \tag{5.8}$$

$$\mathbf{R}_1\mathbf{H}^*_{1c} = -\mathbf{S}^{*12}_{UV}\mathbf{L}_c + o_p(1). \tag{5.9}$$

When the upper left-hand corners of (5.6), (5.8), and (5.9) are substituted into (5.5), we find

$$\widehat{\mathbf{\Psi}}_{k+t} = \begin{bmatrix} \mathbf{S}^{*11}_{WV} & \mathbf{S}^{*12}_{WV} \\ \mathbf{S}^{*21}_{WV} & \mathbf{S}^{*22}_{WV}\mathbf{L}_c\mathbf{L}_c' \end{bmatrix} + o_p(1). \tag{5.10}$$

The difference between $\widehat{\mathbf{\Psi}}^*_{k+t}$ and $\widehat{\mathbf{\Psi}}^*_k$ is

$$\widehat{\mathbf{\Psi}}^*_{k+t} - \widehat{\mathbf{\Psi}}^*_k = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{*22}_{WV}\mathbf{L}_c\mathbf{L}_c' \end{bmatrix} + o_p(1). \tag{5.11}$$

The matrix $\mathbf{L}_2$ has a limiting distribution depending only on $\mathbf{S}_{WV}^{*22}$. [In fact, it has the limiting distribution of the set of t characteristic vectors of $\mathbf{S}_{WV}^{*22}\mathbf{S}_{VW}^{*22}$ corresponding to the largest characteristic roots of $\mathbf{S}_{WV}^{*22}\mathbf{S}_{VW}^{*22}$; see Anderson (1999a). The matrix $\mathbf{S}_{WV}^{*22}\mathbf{S}_{VW}^{*22}$ has a limiting Wishart distribution of dimension $p-k$ and with $q-k$ degrees of freedom; $\left(\mathbf{S}_{WV}^{*22}\mathbf{L}_c\mathbf{L}_c'\right)\left(\mathbf{S}_{WV}^{*22}\mathbf{L}_c\mathbf{L}_c'\right)' = \mathbf{S}_{WV}^{*22}\mathbf{L}_c\mathbf{L}_c'\mathbf{S}_{VW}^{*22}$ has a limiting Wishart distribution of dimension $p-k$ with $q-(k+t)$ degrees of freedom.] Thus $\mathbf{S}_{WV}^{*22}\mathbf{L}_c\mathbf{L}_c'$ is independent of $\widehat{\mathbf{\Psi}}_k$; it adds an independent error to $\widehat{\mathbf{\Psi}}_k^*$; $\widehat{\mathbf{\Psi}}_{k+t}$ has more variability (asymptotically) than $\widehat{\mathbf{\Psi}}_k$.

We see from (5.3) that $\mathbf{H}_{22}-\mathbf{L}_2 \xrightarrow{p} \mathbf{0}$; a part of this statement is that $\mathbf{H}_{2c}-\mathbf{L}_c \xrightarrow{p} \mathbf{0}$. The limiting distribution of $\mathbf{L}_2$ is the Haar invariant measure on orthogonal matrices. $\mathbf{H}_{22}$ is not a consistent estimator, but, nevertheless, $\widehat{\mathbf{\Psi}}_{k+t}$ is a consistent estimator of $\mathbf{\Psi}$. In a sense the factor $\mathbf{L}_c\mathbf{L}_c'$ represents a matrix fraction of $t/(p-k)$.

In the original coordinates we obtain

$$\widehat{\mathbf{B}}_{k+t}^* - \widehat{\mathbf{B}}_k^* = \mathbf{\Sigma}_{YY}\mathbf{A}\left(\widehat{\mathbf{\Psi}}_{k+t}^* - \widehat{\mathbf{\Psi}}_k^*\right)\mathbf{\Gamma}' = \mathbf{\Sigma}_{YY}\mathbf{A}_2\mathbf{S}_{WV}^{*22}\mathbf{L}_c\mathbf{L}_c'\mathbf{\Gamma}_2' \ . \qquad (5.12)$$

6. Testing Rank

The likelihood ratio criterion (under normality) for testing the null hypothesis that the rank of $\mathbf{B}$ is m against alternative that the rank is greater than m is given by (3.2) for $m=k$. The null hypothesis is rejected if the observed value of λ is smaller than a specified value determined by the desired significance level. Under the null hypothesis $(m=k)$ $-2\log\lambda$ has a limiting χ^2-distribution with $(p-m)(q-m)$ degrees of freedom.

If k, the rank of $\mathbf{B}$, is equal to m, the hypothesized rank, the probability of rejecting the hypothesis is the significance level. If $m=s<k$, the probability of rejection will be greater than the significance level. From (3.8) we find $r_i = \rho_i + (1/\sqrt{N})r_i^*$, where

$$2r_i^* = 2s_{ii}^{*VU} - \rho_i s_{ii}^{*UU} - \rho_i s_{ii}^{*VV} + o_p(1), \qquad i=1,\cdots,k. \qquad (6.1)$$

For large N the probability of rejection is close to 1; the investigator will avoid underestimating the rank. If $m=k+t>k$, the probability of rejection is not exactly the significance level, but will be small; the investigator will then be overestimating the rank of $\mathbf{B}$.

The limiting distribution of $g_{k+1} = n\,r_{k+1}^2, \ldots,\ g_p = n\,r_p^2$ has density

$$\text{const} \prod_{i=k+1}^{p} g_i^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i=k+1}^{p} g_i\right) \prod_{\substack{i,j=k+1 \\ i<j}}^{p} (g_i - g_j) \tag{6.2}$$

for $g_{k+1} > \ldots > g_p > 0$ [Anderson (1999a)]. The density of the limiting distribution of $-2\log\lambda = -N\sum_{i=m+1}^{p}\log(1-r_i^2)$ can be found from (6.2) by integration.

There are several methods of determining the appropriate rank, such as minimizing a criterion, possibly AIC or BIC, or carrying out a sequence of tests (testing $m = p-1$, $m = p-2$, etc.). If the rank is overestimated, the penalty is an increased variability in the reduced rank regression estimator. If the rank is underestimated, there will be a bias.

If the procedure is to test in sequence the hypotheses $m = p-1,\ p-2, \ldots, 0$ with the hypothesis $m = s$ accepted and the hypothesis $m = s+1$ rejected, the bias would tend to be small; since the bias depends on the smaller canonical correlations, large values of these smaller roots would tend to make acceptances of hypotheses $m = p-1, \ldots, s$ unlikely.

An example is given in Section 8.

7. Reduced Rank Regression Estimator of a Submatrix

For nonstochastic observations on $\mathbf{X}$ the model (1.2) can be transformed into a model like (1.1). Let

$$\mathbf{X}_1^+ = \mathbf{X}_1 - \mathbf{S}_{XX}^{12}\left(\mathbf{S}_{XX}^{22}\right)^{-1}\mathbf{X}_2, \tag{7.1}$$

$$\mathbf{B}_2^+ = \mathbf{B}_2 + \mathbf{B}_1\mathbf{S}_{XX}^{12}\left(\mathbf{S}_{XX}^{22}\right)^{-1}, \tag{7.2}$$

where $\mathbf{S}_{XX}^{ij} = N^{-1}\Sigma_{\alpha=1}^{N}\mathbf{x}_{i\alpha}\mathbf{x}_{j\alpha}$ and $\mathbf{x}_{i\alpha}$ is the α-th observation on $\mathbf{X}_i$. Then (1.2) is equivalent to

$$\mathbf{Y} = \mathbf{B}_1\mathbf{X}_1^+ + \mathbf{B}_2^+\mathbf{X}_2 + \mathbf{Z}\ . \tag{7.3}$$

The variables $\mathbf{X}_1^+$ and $\mathbf{X}_2$ are uncorrelated in the sample. Usually one component of $\mathbf{X}_2$ would be 1.

The LS estimator of $(\mathbf{B}_1^+, \mathbf{B}_2)$ is

$$\left(\widehat{\mathbf{B}}_1^+, \widehat{\mathbf{B}}_2\right) = \left[\mathbf{S}_{YX}^{+\cdot 1}\left(\mathbf{S}_{XX}^{+11}\right)^{-1}, \mathbf{S}_{YX}^{\cdot 2}\left(\mathbf{S}_{XX}^{22}\right)^{-1}\right]\ , \tag{7.4}$$

where $\mathbf{S}_{YX}^{+\cdot 1} = N^{-1}\Sigma_{\alpha=1}^{N}\mathbf{y}_\alpha\mathbf{x}_{1\alpha}^{+\prime}$ and $\mathbf{S}_{XX}^{+11} = N^{-1}\Sigma_{\alpha=1}^{n}\mathbf{x}_{1\alpha}^{+}\mathbf{x}_{1\alpha}^{+\prime} = \mathbf{S}_{XX}^{11} - \mathbf{S}_{XX}^{12}(\mathbf{S}_{XX}^{22})^{-1}\mathbf{S}_{XX}^{21}$. The RRR estimator of $\mathbf{B}_1^{+}$ is $\mathbf{S}_{YX}^{+\cdot 1}\widehat{\mathbf{\Gamma}}_1\widehat{\mathbf{\Gamma}}_1'$, where $\widehat{\mathbf{\Gamma}}_1 = (\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_k)$ and $\hat{\boldsymbol{\gamma}}_i$ is found from (2.15) with $\mathbf{S}_{XY}$. For the Corollary it is assumed that $\mathbf{S}_{XX} \to \mathbf{\Sigma}_{XX}$. In Theorem 1 $\mathbf{\Sigma}_{XX}$ and $\mathbf{B}$ are replaced by $\mathbf{\Sigma}_{XX}^{+11}$ and $\mathbf{B}_1 = \mathbf{\Lambda\Pi}'$.

If the observations on $\mathbf{X}$ are stochastic with $\mathcal{E}\mathbf{X} = \mathbf{0}$ and $\mathcal{E}\mathbf{X}\mathbf{X}' = \mathbf{\Sigma}_{XX}$ Theorem 1 holds with the above definition of $\widehat{\mathbf{B}}_{1k}$.

8. An Example

Reinsel and Velu (1998) give a chemometric example from Skagerberg, MacGregor, and Kiparissedes (1992) using (1.2) with $p = 6$ characteristics of a polymer $\mathbf{Y}$ and $q = 23$ process measurements $\mathbf{X}$ with $N = 56$ observations. Of the 23 process variables $q_2 = 3$ measurements $\mathbf{X}_2$ are included in each model considered. Since the $q_1 = 20$ measurements $\mathbf{X}_1$ are correlated and impact the polymer characteristics somewhat similarly, the question is whether $\mathbf{X}_1$ can be replaced by m linear combinations of $\mathbf{X}_1$ (rank $\mathbf{B}_1$ of m).

Let $M_m = -\left[N - q + (q_1 - p - 1)/2\right]\sum_{i=m+1}^{p}\log\left(1 - r_i^2\right)$. [The factor $\left[N - q + (q_1 - p - 1)/2\right]/N = .705$ is the "Bartlett correction."] The canonical correlations r_i, $i = 1, \ldots, 6$, are given on the left of the table below, and the observed values of the criterion M_m, $m = 0, 1, \ldots, 5$, are given on the right. At the 5% level of significance the null hypothesis that the rank of $\mathbf{B}_1$ is 1 (or 0) is rejected; the hypothesis that the rank is 2 (or $3, \ldots, 5$) is accepted. If the decision procedure is a sequence of 5% significance tests $m = 5, 4, \ldots, 1$, the conclusion is that the 20 components of $\mathbf{X}_1$ can be replaced by the first two canonical variables. This conclusion could be an underestimation of the rank; the values of $r_3^2, \ldots, r_6^2$ give an idea of the effects lost by using only 2 canonical variables. The conclusion of rank 2 could be an overestimation of rank, but the value of $r_2^2 = .866$ suggests that the use of the second canonical variable is needed.

Canonical correlations			Test criteria			
i	r_i	r_i^2	m	M_m	d.f.	5% significance point
1	0.991	0.983	0	310.53	120	146.57
2	0.930	0.866	1	150.00	95	118.75
3	0.727	0.529	2	70.80	72	92.81
4	0.633	0.400	3	41.14	51	68.67
5	0.533	0.284	4	20.95	32	46.19
6	0.422	0.178	5	7.76	15	25.00

9. Acknowledgements

The author wishes to thank Yasuo Amemiya, Naoto Kunitomo, and Tore Schweder for helpful suggestions.

References

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

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

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis,* Second edition, John Wiley & Sons, Inc., New York.

Anderson, T.W. (1989). The asymptotic distribution of characteristic roots and vectors in multivariate components of variance, *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin* (Leon Jay Gleser, Michael D. Perlman, S. James Press, and Allan R. Sampson, eds.), Springer Verlag, New York, 177-196.

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

Anderson, T.W. (1999b). Asymptotic distribution of the reduced rank regression estimator under general conditions, *Ann. Statist.*, **27**, 1141-1154.

Anderson, T.W., and Herman Rubin. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, *Ann. Math. Statist.*, **20**, 46-63.

Johansen, Soren. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models,* Oxford University Press, Oxford.

Reinsel, Gregory C. and R.V. Velu. (1998). *Multivariate Reduced-rank Regression*, Springer, New York.

Skagerberg, B., J. MacGregor and C. Kiparissedes. (1992). Multivariate data analysis applied to low-density polyethylene reactors. *Chemometrics and Intelligent Laboratory Systems*, **14**, 341–356.

Stoica, Petre and Mats Viberg (1996). Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Trans. Signal Process.*, **44**, 3069-3078.

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

T.W.Anderson

Department of Statistics

Stanford University

Stanford, CA 94305-4065

USA

E-mail: twa@stat.stanford.edu