ESTIMATION OF REDUCED RANK REGRESSION

by

T. W. ANDERSON
Department of Statistics
Stanford University

Technical Report No. 1998–4
April 1998

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Estimation of Reduced Rank Regression

T. W. Anderson

Department of Statistics, Stanford University, Stanford, CA 94305-4065

April 10, 1998

Abstract

In the regression model $Y = BX + Z$ with $Z$ unobserved, $EZ = 0$, and $EZZ = \Sigma_{ZZ}$, the least squares estimator of $B$ is $\hat{B} = S_{YX}S_{XX}^{-1}$. If the rank of $B$ is known to be $k$ less than the dimensions of $Y$ and $X$, the reduced rank regression estimator of $B$, say $\hat{B}_k$, depends on the first $k$ canonical variates of $Y$ and $X$ (Anderson 1951a). The limiting distribution of $\hat{B}_k$ is obtained and compared with the limiting distribution of $\hat{B}$. The advantage of $\hat{B}_k$ is characterized.

Key Words: Canonical variates, reduced rank regression, maximum likelihood estimators, test of rank.

1 Introduction

Reduced rank regression 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 \( Y_\alpha \) on a vector of \( q \) independent variables \( X_\alpha \) is

\[
Y_\alpha = BX_\alpha + Z_\alpha,
\]

(1.1)

where the unobservable disturbance or error \( Z_\alpha \) is distributed independently of \( X_\alpha \) with \( Z_\alpha = 0 \) and \( EZ_\alpha Z'_\alpha = \Sigma_{ZZ} \).

If the rank of \( B \) is \( k \), only \( k \leq \min(p, q) \) linear combinations of the components of \( X \) suffice to predict or "explain" \( Y \). These linear combinations are the canonical variates of \( X \) (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 \((y_1, x_1), ..., (y_N, x_N)\) an estimator of \( B \) is desired. Anderson (1951a) found the maximum likelihood estimator of \( B \) of preassigned rank \( k \) when \( x_1, ..., x_N \) are considered nonstochastic and \( z_1, ..., z_N \) are independently distributed according to \( N(0, \Sigma_{ZZ}), \alpha = 1, ..., N \).

If \((Y_\alpha', X_\alpha')'\) have a joint normal distribution with mean vector \( E(Y_\alpha', X_\alpha')' = (\mu_Y', \mu_X')' \) and covariance matrix

\[
E \begin{bmatrix}
Y_\alpha - \mu_Y \\
X_\alpha - \mu_X
\end{bmatrix}
(Y_\alpha' - \mu_Y', X_\alpha' - \mu_X') = 
\begin{bmatrix}
\Sigma_{YY} & \Sigma_{YX} \\
\Sigma_{XY} & \Sigma_{XX}
\end{bmatrix} = \Sigma,
\]

(1.2)

then the density of \((Y_\alpha', X_\alpha')'\) is

\[
n \begin{bmatrix}
(\mu_Y') \\
(\mu_X')
\end{bmatrix}, \begin{bmatrix}
\Sigma_{YY} & \Sigma_{YX} \\
\Sigma_{XY} & \Sigma_{XX}
\end{bmatrix}
\]

(1.3)

\[
= n \left[ y | \mu_Y + B(x - \mu_X), \Sigma_{ZZ} \right] n \left[ x | \mu_X, \Sigma_{XX} \right],
\]
where

\[ B = \Sigma_{YY} \Sigma_{XX}^{-1}, \]  

(1.4)

\[ \Sigma_{ZZ} = \Sigma_{YY} - B \Sigma_{XX} B' = \Sigma_{YY} - \Sigma_{YY} \Sigma_{XX}^{-1} \Sigma_{XY}. \]  

(1.5)

The maximum likelihood estimator of \( B \) 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 \( x_1, ..., x_N \) nonstochastic.

The maximum likelihood estimator can be defined in terms of the canonical variates. One form is \( \hat{B}_k = S_{YX} \hat{\Gamma}_1\hat{\Gamma}_1' \), where \( S_{YX} \) is the sample covariance between \( Y \) and \( X \) and \( \hat{\Gamma}_1 \) consists of the coefficients of the first \( k \) canonical variates of \( X \); other forms are given in (2.13) below.

The objectives of this paper are to obtain the asymptotic distribution of \( \hat{B}_k \) for \( x_1, ..., x_N \) nonstochastic and for \( x_1, ..., x_N \) observations on a random vector with \( \mathcal{E}X = \mu_X \) and \( \mathcal{E}X = \Sigma_{XX} \), and to relate the asymptotic distribution of \( \hat{B}_k \) to the asymptotic distribution of \( \hat{\Gamma}, \hat{\Lambda}, \) and \( \hat{\Gamma}_1 \), the sample canonical correlations and coefficients of the canonical variables. In fact, in asymptotics \( \hat{B}_k \) is a simple function of the sample covariance matrix of \( Y \) and \( X \) expressed in terms of the canonical variables. It is shown that the asymptotic distribution of \( \hat{B}_k \) does not require that the dependent variables are normally distributed.

Stoica and Viberg (1996) have shown that the asymptotic distribution of \( \hat{B}_k \) is normal if \( Z_\alpha \) is normal and have given an expression of the covariance of the limit distribution, but it is difficult to interpret this expression. In this paper the covariance matrix is given in terms that are apparent, and it is shown that the result does not depend on the normality of \( Z_\alpha \).

In Section 6 the effect of misspecification of the rank of \( B \) is considered. The asymptotic distribution of the reduced rank estimator is different when the rank is underestimated and when it is overestimated.
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) and Anderson (1997), for example. The equations defining the canonical correlations and variates (in the population) are

\[
\begin{pmatrix}
-\rho \Sigma_{YY} & \Sigma_{YX} \\
\Sigma_{XY} & -\rho \Sigma_{XX}
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\gamma
\end{pmatrix}
= 0,
\]

(2.1)

where \( \rho \) satisfies

\[
\begin{vmatrix}
-\rho \Sigma_{YY} & \Sigma_{YX} \\
\Sigma_{XY} & -\rho \Sigma_{XX}
\end{vmatrix}
= 0,
\]

(2.2)

and

\[
\alpha' \Sigma_{YY} \alpha = 1, \quad \gamma' \Sigma_{XX} \gamma = 1.
\]

(2.3)

The number of positive canonical correlations is the rank of \( \Sigma_{YX} \), which is the rank of \( \mathbf{B} \). The roots of (2.2) 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 \( q > p \). To eliminate the indeterminacy of a solution \( \alpha, \gamma \) in (2.1) we require \( \alpha_{ii} > 0 \). (Since the matrix \( \mathbf{A} = (\alpha_1, \ldots, \alpha_p) \) is nonsingular, the components of \( \mathbf{Y} \) can be numbered in such a way that the \( i \)th component of \( \alpha_i \) is nonzero.

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

\[
\rho^2 \Sigma_{YY} \alpha = \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \alpha = \mathbf{B} \Sigma_{XX} \mathbf{B}' \alpha,
\]

(2.4)

\[
\rho^2 \Sigma_{XX} \gamma = \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY} \gamma.
\]

(2.5)
The solutions of (2.1) corresponding to \( \rho_1, \ldots, \rho_p \) can be assembled as \( \mathbf{A} = (\alpha_1, \ldots, \alpha_p) \) and 
\((\gamma_1, \ldots, \gamma_p)\). If \( q > p \), there are \( q - p \) additional solutions \((\gamma_{p+1}, \ldots, \gamma_q)\) to (2.1) with \( \rho = 0 \).

Let \( \mathbf{R} = (\gamma_1, \ldots, \gamma_q) \) and let \( \mathbf{R} = \text{diag}(\rho_1^2, \ldots, \rho_q^2) \), \( \mathbf{R} = (\mathbf{R}, \mathbf{0}) \). Then the solutions can be chosen to satisfy

\[
\begin{pmatrix}
\mathbf{A}' & 0 \\
0 & \mathbf{R}'
\end{pmatrix}
\begin{pmatrix}
\Sigma_{YY} & \Sigma_{YX} \\
\Sigma_{XY} & \Sigma_{XX}
\end{pmatrix}
\begin{pmatrix}
\mathbf{A}'
\\
\mathbf{R}'
\end{pmatrix}
= \begin{pmatrix}
\mathbf{I} & \mathbf{R}'
\end{pmatrix}.
\]

This is the covariance matrix of the canonical variates \( \mathbf{U} = \mathbf{A}' \mathbf{Y} \) and \( \mathbf{V} = \mathbf{R}' \mathbf{Y} \).

If the sample is from a normal distribution \( N(\mu, \Sigma) \), the unrestricted maximum likelihood estimators of the means and covariances are \( \hat{\mu} = (\bar{y}', \bar{x}')' \) and

\[
\hat{\Sigma} = \frac{1}{N} \sum_{\alpha=1}^{N} \begin{pmatrix}
(y_\alpha - \bar{y}) \\
x_\alpha - \bar{x}
\end{pmatrix} (y_\alpha' - \bar{y}', x_\alpha' - \bar{x}')
\]

\[
= \begin{pmatrix}
\hat{\Sigma}_{YY} & \hat{\Sigma}_{YX} \\
\hat{\Sigma}_{XY} & \hat{\Sigma}_{XX}
\end{pmatrix} = \frac{n}{N} \begin{pmatrix}
\mathbf{S}_{YY} & \mathbf{S}_{YX} \\
\mathbf{S}_{XY} & \mathbf{S}_{XX}
\end{pmatrix} = \frac{n}{N} \mathbf{S},
\]

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

\[
\begin{pmatrix}
-r \mathbf{S}_{YY} & \mathbf{S}_{YX} \\
\mathbf{S}_{XY} & -r \mathbf{S}_{XX}
\end{pmatrix}
\begin{pmatrix}
\mathbf{a}' \\
\mathbf{c}'
\end{pmatrix} = 0,
\]

\[
a' \mathbf{S}_{YY} \mathbf{a} = 1, \quad c' \mathbf{S}_{XX} \mathbf{c} = 1.
\]

The solutions with \( a_{ii} > 0, \quad i = 1, \ldots, p, \) and \( r_1 > r_2 > \ldots > r_p > 0 \) define the estimators \( \hat{\mathbf{A}} = (a_1, \ldots, a_p), \quad \hat{\mathbf{r}} = (c_1, \ldots, c_q), \quad \hat{\mathbf{R}} = \text{diag}(r_1, \ldots, r_p) \). These are uniquely defined except that

if \( q > p, c_{p+1}, \ldots, c_q \) satisfy \( c' \mathbf{S}_{XX} \mathbf{c}_j = 0, \quad j = 1, \ldots, p, \) and some other \((q-p)(q-p-1)\) arbitrary
conditions. From (2.8) and (2.9) we obtain $c = (1/\tau)S_{XX}^{-1}S_{XY}a$, $a = (1/\tau)S_{YY}^{-1}S_{YX}c$

\[ S_{YX}S_{XX}^{-1}S_{XY}a = \tau^2 S_{YY}a, \quad (2.10) \]

\[ S_{XY}S_{YY}^{-1}S_{YX}c = \tau^2 S_{XX}c. \quad (2.11) \]

Let $\hat{A}_1 = a_1, \ldots, a_k$, $\hat{R}_1 = (c_1, \ldots, c_k)$, $R_1 = \text{diag}(r_1, \ldots, r_k)$, and $\hat{\Phi}_1 = \hat{A}_1(I - R_1^2)^{-1/2} = S_{YY}^{-1}S_{YX}\hat{R}_1R_1^{-1}(I - R_1^2)^{-1/2}$. The unrestricted maximum likelihood estimator is

\[ B = S_{YX}S_{XX}^{-1}. \quad (2.12) \]

The maximum likelihood estimator of $B$ of rank $k$ found by Anderson (1951a) is

\[ \hat{B}_r = S_{ZZ}\hat{\Phi}_1\hat{\Phi}_1'B = S_{YY}\hat{A}_1\hat{R}_1\hat{R}_1' = S_{YX}\hat{R}_1\hat{R}_1', \quad (2.13) \]

where $S_{ZZ}$ is the sample covariance matrix of the residual $z = y - Bx$. A column of $\hat{\Phi}_1$ satisfies

\[ S_{YX}S_{XX}^{-1}S_{XY}\hat{\phi} = tS_{ZZ}\hat{\phi}, \quad \hat{\phi}'S_{ZZ}\hat{\phi} = 1, \quad (2.14) \]

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

The distribution of $nS_{YY}$, $nS_{YX}$, $nS_{XX}$ is the same as the distribution of $\Sigma_{\alpha=1}^n Y_\alpha Y'_\alpha$, $\Sigma_{\alpha=1}^n Y_\alpha X'_\alpha$, $\Sigma_{\alpha=1}^n X_\alpha X'_\alpha$ when $\mu_Y = 0$, $\mu_X = 0$ and $(Y', X')'$ is normally distributed. In any case the limiting distribution of $\sqrt{n}(S_{YY} - \Sigma_{YY})$, $\sqrt{n}(S_{YX} - \Sigma_{YX})$, $\sqrt{n}(S_{XX} - \Sigma_{XX})$ does not depend
on \((\mu_{Y}', \mu_X')\). Hence we shall consider the model as

\[
\begin{bmatrix}
S_{YY} & S_{YX} \\
S_{XY} & S_{XX}
\end{bmatrix}
= \frac{1}{n} \sum_{\alpha=1}^{n} \begin{bmatrix} Y_{\alpha} \\ X_{\alpha} \end{bmatrix} \begin{bmatrix} Y_{\alpha}' & X_{\alpha}' \end{bmatrix}
\]

(2.15)

and \(\varepsilon Y_\alpha = 0, \varepsilon X_\alpha = 0\).

3. Asymptotic distribution of the reduced rank regression when the independent variables are stochastic

The development of the asymptotic distributions relies heavily on Anderson (1997). We want to find the asymptotic distribution of \(\hat{B}_k = S_{YX} \hat{\Gamma}_1 \hat{\Gamma}_1'\). Note that the fact that the number of columns of \(\hat{\Gamma}_1\) is the rank of \(B\) implies that the rank of \(B\) is known to the statistician. The transformation to canonical variables \(U_\alpha = A'Y_\alpha, V_\alpha = \Gamma'X_\alpha\), and \(W_\alpha = A'Z_\alpha\) transforms (1.1) to

\[
U_\alpha = \Psi V_\alpha + W_\alpha,
\]

(3.1)

\(\Sigma_{UU} = A'\Sigma_{YY} A = I, \Sigma_{VV} = \Gamma'\Sigma_{XX} \Gamma = I, \Sigma_{UV} = A'\Sigma_{YX} \Gamma = (R, 0), \Sigma_{VW} = 0, \Sigma_{WW} = A'\Sigma_{ZZ} A = I - R^2\), and \(\Psi = A'B(\Gamma')^{-1} = \hat{R}\). Also \(S_{UU} = A'S_{YY} A, S_{UV} = A'S_{YX} \Gamma, S_{VV} = \Gamma S_{XX} \Gamma', \hat{\Psi} = S_{UV} S_{VV}^{-1} A' \hat{B}(\Gamma')^{-1}\) (the unrestricted estimator of \(\hat{\Psi}\)), and

\[
\hat{\Psi}_k = S_{UV} H_1 H_1',
\]

(3.2)
where $H_1 = \Gamma_1^{-1} \hat{\Gamma}_1$ satisfies

$$S_{UV} S_{UU}^{-1} S_{UV} H_1 = S_{VV} H_1 \hat{R}_1^2, \quad H_1^* S_{VV} H_1 = I_k. \quad (3.3)$$

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

Define $S_{UU}^* = \sqrt{n}(S_{UU} - I_p)$, $S_{UV}^* = \sqrt{n}(S_{UV} - \hat{R})$, $S_{VV}^* = \sqrt{n}(S_{VV} - I_q)$, $H_1^* = \sqrt{n}(H_1 - I_{(k)})$, and $\hat{R}^* = [\sqrt{n}(\hat{R} - R), 0]$, where $I_{(k)} = (I_k, 0)'$. Then substitution of these quantities into (3.3) yields

$$\begin{align*}
\bar{R}' \bar{R} I_{(k)} + \frac{1}{\sqrt{n}} \left[ S_{UU}^* \bar{R} I_{(k)} + \bar{R}' S_{UV}^* I_{(k)} - \bar{R}' S_{UU}^* \bar{R} I_{(k)} + \bar{R}' \bar{R} H_1^* \right] \\
= I_{(k)} R_1^2 + \frac{1}{\sqrt{n}} \left[ S_{VV}^* I_{(k)} R_1^2 + 2I_{(k)} R_1 R_1^* + H_1^* R_1^2 \right] + o_p \left( \frac{1}{\sqrt{n}} \right)
\end{align*} \quad (3.4)$$

or equivalently

$$\begin{align*}
S_{UU}^* I_{(k)} R_1 + \bar{R}' S_{UV}^* I_{(k)} - \bar{R}' S_{UU}^* I_{(k)} R_1 - S_{VV}^* I_{(k)} R_1^2 \\
= 2I_{(k)} R_1 R_1^* + H_1^* R_1^2 - \bar{R}' \bar{R} H_1^* + o_p(1).
\end{align*} \quad (3.5)$$

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

$$\begin{align*}
\begin{bmatrix}
S_{UU}^{11} R_1 + R_1 S_{UV}^{11} - R_1 S_{UU}^{11} R_1 - S_{VV}^{11} R_1^2 \\
S_{UU}^{21} R_1 - S_{VV}^{21} R_1^2 \\
2R_1 R_1^* + H_1^* R_1^2 - R_1^2 H_1^* \\
H_1^* R_1^2
\end{bmatrix}
+ o_p(1), \quad (3.6)
\end{align*}$$
where $H^* = (H_{11}', H_{21}')$. From (3.6) we obtain

$$H_{21}' R_1 + o_p(1) = S_{VV}' S_{UV}' R_1 = \left( S_{VV}' - R_1 S_{VV}' \right)' = \left( S_{WV}' \right)' . \quad (3.7)$$

From the second part of (3.3) we obtain

$$I'_{(k)} I_{(k)} + \frac{1}{\sqrt{n}} \left[ H_1'I_{(k)} + I'_{(k)} H_1' + I'_{(k)} S_{VV} I_{(k)} \right] = I_k + o_p \left( \frac{1}{\sqrt{n}} \right) , \quad (3.8)$$

from which we obtain

$$H_{11}' + H_{11} = -S_{VV} + o_p(1) . \quad (3.9)$$

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

$$\hat{\Psi}_k^* = S_{UV}' I_{(k)}' I_{(k)} + \bar{R} H_1'I_{(k)} + \bar{R} I_{(k)} H_1' + o_p(1)$$

$$= \begin{bmatrix}
R_1 (H_{11}' + H_{11}') + S_{UV}' & R_1 H_{21}' \\
S_{VV}' & 0
\end{bmatrix} + o_p(1)$$

$$= \begin{bmatrix}
S_{WV}' & S_{VV}' \\
S_{WV}' & 0
\end{bmatrix} + o_p(1) , \quad (3.10)$$

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 $S_{UV}' - \bar{R} S_{VV}' = S_{WV}^*$. The maximum likelihood estimator of $B$ unrestricted with respect to rank is $\hat{B} = S_{YX} S_{X}^{-1}$, the unrestricted estimator in terms of canonical variables is $\hat{\psi} = S_{UV} S_{VV}^{-1}$, and $\hat{\psi}^* = \sqrt{n}(\hat{\psi} - \psi)$.
\[ \Psi = S_{\Psi}^* \]. The effect of the rank restriction is to replace the lower right-hand corner of \( S_{\Psi}^* \) by 0.

To characterize the distribution of \( \hat{\Psi}_k \) and \( \hat{B}_k \) we use the notation \( \text{vec } A = \text{vec } (a_1, \ldots, a_m) = (a'_1, \ldots, a'_m)' \) and \( A \otimes B = (a_i B) \) and the property \( \text{vec } ABC = (C' \otimes A) \text{vec } B \), which implies \( \text{vec } xy' = \text{vec } x'y' = (y \otimes x) \text{vec } 1 = y \otimes x \). Then

\[
\text{vec } \hat{\Psi}_r^* = \text{vec } \frac{1}{\sqrt{n}} \left[ \sum_{\alpha=1}^{n} w_{\alpha} v_{\alpha}^{(1)'}, \sum_{\alpha=1}^{n} \left( \begin{array}{c} w_{\alpha}^{(1)} \\ 0 \end{array} \right) v_{\alpha}^{(2)'} \right] = \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} \left[ \begin{array}{c} v_{\alpha}^{(1)} \otimes w_{\alpha} \\ v_{\alpha}^{(2)} \otimes \left( \begin{array}{c} w_{\alpha}^{(1)} \\ 0 \end{array} \right) \end{array} \right].
\]

(3.11)

When \( v_{\alpha} \) and \( w_{\alpha} \) are independent, we obtain

\[
\mathcal{E} \text{ vec } \hat{\Psi}_k^* \left( \text{vec } \hat{\Psi}_k^* \right)' \rightarrow \left[ \begin{array}{c} I_k \otimes (I_p - R^2) & 0 \\ 0 & I_{q-k} \otimes \left[ I_{k-R_2^2} 0 \\ 0 \right] \end{array} \right]
\]

(3.12)

\[
= \text{diag} \left( I_p - R^2, \ldots, I_p - R^2, I_k - R_1^2, 0, \ldots, I_k - R_1^2, 0 \right),
\]

where there are \( k \) blocks of \( I_p - R^2 \) and \( q - k \) blocks of \( \text{diag}(I_k - R_1^2, 0) \). The rank of (3.12) is \( kp + (q - k)k = k(p + q - k) \).

**Theorem 1.** Let \((u'_\alpha, v'_\alpha), \alpha = 1, \ldots, n,\) be observations on the random vector \((U', V')'\) with mean 0 and covariance matrix (2.6). Let \( \Psi = \Sigma_{UV} \Sigma_{VV}^{-1}, \) \( S_{UU} = n^{-1} \sum_{\alpha=1}^{n} u_{\alpha} u_{\alpha}', \) \( S_{UV} = n^{-1} \sum_{\alpha=1}^{n} u_{\alpha} v_{\alpha}', \) \( S_{VV} = n^{-1} \sum_{\alpha=1}^{n} v_{\alpha} v_{\alpha}' \). Let \( H_{1i} \) satisfy (3.3) and \( h_{ii} > 0 \). Suppose \( U - \Psi V \) is independent of \( V \). Then the limiting distribution of \( \text{vec} \sqrt{n} (\hat{\Psi}_k - \Psi) \), where \( \hat{\Psi}_k \) is defined by (3.2), is normal with mean 0 and covariance matrix (3.12).

In the original coordinate system we obtain

\[
\text{vec } (\hat{B}_k - B) = \text{vec } [(A')^{-1}(\hat{\Psi}_k - \Psi) \Gamma']
\]

(3.13)
\[
= \left[ \Gamma \otimes (\mathbf{A}')^{-1} \right] \text{vec}(\hat{\Psi}_k - \Psi)
\]
\[
= \left[ (\Gamma_1, \Gamma_2) \otimes \Sigma_{ZZ}(A_1, A_2)(I - R^2)^{-1} \right] \text{vec}(\hat{\Psi}_k - \Psi).
\]

From (3.12) and (3.13) we obtain

\[
\begin{align*}
\mathcal{E} \text{ vec } n (\hat{B}_k - B) \left[ \text{vec} (\hat{B}_k - B) \right]' & \rightarrow \left[ (\Gamma_1, \Gamma_1') \otimes \Sigma_{ZZ} A_1 (I_p - R_2)^{-1} A' \Sigma_{ZZ} \right] \\
& \quad + \left[ (\Gamma_2, \Gamma_2') \otimes \Sigma_{ZZ} A_1 (I_k - R_1)^{-1} A_2' \Sigma_{ZZ} \right] \\
& = \left[ (\Gamma_1, \Gamma_1') \otimes \Sigma_{ZZ} \right] + \left[ (\Gamma_2, \Gamma_2') \otimes \Sigma_{ZZ} A_1 (I_k - R_1)^{-1} A_2' \Sigma_{ZZ} \right].
\end{align*}
\]

(3.14)

**Corollary 1.** Let \((y'_\alpha, x'_\alpha)', \alpha = 1, \ldots, n\), be observations on the random vector \((Y', X')'\) with mean 0 and covariance matrix (1.2). Let \(B = \Sigma_{YY} \Sigma_{XX}^{-1}, S_{YY} = n^{-1} \sum_{\alpha=1}^{n} y_\alpha y'_\alpha, S_{XY} = n^{-1} \sum_{\alpha=1}^{n} x_\alpha y'_\alpha, S_{XX} = n^{-1} \sum_{\alpha=1}^{n} x_\alpha x'_\alpha\). Let the columns of \(\hat{\Gamma}_1\) satisfy (2.9), (2.11), and \(\hat{\gamma}_{ii} > 0\). Suppose that \(Y - BX\) is independent of \(X\). Then the limiting distribution of \(\frac{n}{\sqrt{n}} \text{vec}(\hat{B}_k - B)\), with \(\hat{B}_k = S_{YY} \hat{\Gamma}_1 \hat{\Gamma}'_1\), is normal with mean 0 and covariance matrix (3.14).

It is curious that \(\hat{\Psi}_k^*\) depends asymptotically only on \(S_{YY} S_{XX}^{-1} = (\mathbf{A}')^{-1} S_{YY} - \hat{R} S_{YY} S_{YY}^{-1} \hat{\Gamma}\), but not further on \(\hat{\Gamma}_1 = (c_1, \ldots, c_k)\). The limiting distribution of \(\sqrt{n} (c_1 - \gamma_1), \ldots, \sqrt{n} (c_k - \gamma_k)\) is normal with covariances

\[
n \mathcal{E} (c_j - \gamma_j) (c_j - \gamma_j)' \rightarrow \frac{1}{2} \gamma_j \gamma_j' + (1 - \rho_j^2) \sum_{k \neq j} \frac{\rho_k^2 + \rho_j^2 - 2 \rho_k^2 \rho_j^2}{(\rho_j^2 - \rho_k^2)^2} \gamma_k \gamma_k', \quad j = 1, \ldots, k, \tag{3.15}
\]

\[
n \mathcal{E} (c_j - \gamma_j) (c_\ell - \gamma_\ell) \rightarrow -\frac{(1 - \rho_j^2)(1 - \rho_\ell^2)}{(\rho_j^2 - \rho_\ell^2)^2}, \quad j \neq \ell. \tag{3.16}
\]

11
See Anderson (1997). These follow from (3.6)

\[ h^*_i (\rho_j^2 - \rho_i^2) = s^{*UV}_i \rho_j + \rho_i s^{*UU}_i j \rho_j - s^{*VV}_i \rho_j^2 + o_p(1), i \neq j, \quad i = 1, \ldots, q, j = 1, \ldots, k. \]

(3.17)

These covariances are valid for \((Y, X)\) normally distributed and \(\rho_1 > \ldots > \rho_k\); they depend on the second-order moments of the sample covariance matrices (hence, on the fourth-order moments of the observed variable). However, the limiting distribution of \(\text{vec} \tilde{B}^*\) is \(N(0, \Sigma^{-1}_{XX} \otimes \Sigma_{ZZ})\), irrespective of whether \(Y\) and \(X\) are normal.

4 Asymptotic distribution when the independent variable is non-stochastic

Now suppose that \(X_\alpha = x_\alpha\), \(\alpha = 1, \ldots, n\), is nonstochastic. We assume that

\[ S_{XX} = \frac{1}{n} \sum_{\alpha=1}^{n} x_\alpha x'_\alpha \rightarrow \Sigma_{XX}. \]

(4.1)

The model is

\[ Y_\alpha = B x_\alpha + Z_\alpha, \]

(4.2)

where \(\varepsilon Z_\alpha = 0\) and \(\varepsilon Z_\alpha Z'_\alpha = \Sigma_{ZZ}\).

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

\[
\begin{bmatrix}
-\rho (\Sigma_{ZZ} + BS_{XX}B') & BS_{XX} \\
S_{XX}B' & -\rho S_{XX}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\gamma
\end{bmatrix}
= 0,
\]

(4.3)
\( \alpha' (\Sigma_{ZZ} + BS_{XX}B') \alpha = 1, \quad \gamma' S_{XX} \gamma = 1. \)  \hspace{1cm} (4.4)

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

\[
BS_{XX}B' \alpha = \rho^2 (\Sigma_{ZZ} + BS_{XX}B') \alpha. \hspace{1cm} (4.5)
\]

This equation and (4.4) imply \( \alpha' \Sigma_{ZZ} \alpha = 1 - \rho^2 \), \( \theta = \rho^2/(1 - \rho^2) \), and \( \phi = \alpha(1 - \rho^2)^{-\frac{1}{2}} \). The solutions to (4.3) and (4.4) and \( \alpha_i > 0 \), \( \rho_1 > \ldots > \rho_p \) define the matrices \( A_n = (\alpha_1, \ldots, \alpha_p) \), \( \Gamma_n = (\gamma_1, \ldots, \gamma_q) \), \( R_n = \text{diag}(\rho_1, \ldots, \rho_p) \). Where convenient, the subscript \( n \) is used to emphasize that the matrices of transformed parameters depend on \( n \) through \( S_{XX} \). Now define \( U = A'_n Y \), \( V_\alpha = \Gamma_n x_\alpha \), \( W = A'_n Z \). Then

\[
\Sigma_{WW} = A'_n \Sigma_{ZZ} A_n = I - R_n^2, \hspace{1cm} (4.6)
\]

\[
S_{VV} = \Gamma'_n S_{XX} \Gamma_n = I, \hspace{1cm} (4.7)
\]

\[
A'_n B S_{XX} \Gamma_n = \bar{R}_n = A'_n B(\Gamma'_n)^{-1}. \hspace{1cm} (4.8)
\]

We write the model for \( U \) in terms of \( v \) and \( W \) as

\[
U = \Psi v + W, \hspace{1cm} (4.9)
\]

where \( \bar{R}_n = A'_n B(\Gamma'_n)^{-1} \) has been replaced by \( \Psi \).
The unrestricted maximum likelihood estimators of $\mathbf{B}$ and $\Sigma_{ZZ}$ are given by (2.12) and

$$S_{ZZ} = \frac{1}{n} \sum_{a=1}^{n} \left( \mathbf{y}_a - \hat{\mathbf{B}} \mathbf{x}_a \right) \left( \mathbf{y}_a - \hat{\mathbf{B}} \mathbf{x}_a \right)' = S_{YY} - \hat{\mathbf{B}} S_{XX} \hat{\mathbf{B}}'. \quad (4.10)$$

The estimators of $\mathbf{A}_n$, $\Gamma_n$, 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 $\Psi$ and $\Sigma_{WW}$ are

$$\hat{\Psi} = S_{UV} S_{VV}^{-1} = S_{UV}, \quad (4.11)$$

$$S_{WW} = \frac{1}{n} \sum_{a=1}^{n} \left( \mathbf{u}_a - \hat{\Psi} \mathbf{v}_a \right) \left( \mathbf{u}_a - \hat{\Psi} \mathbf{v}_a \right)' \quad (4.12)$$

$$= S_{UU} - \hat{\Psi} S_{VV} \hat{\Psi}' = S_{UU} - \hat{\Psi} \hat{\Psi}'$$

$$= S_{UU} - S_{UV} S_{VV}^{-1} S_{VV} = S_{UU} - S_{UV} S_{VV}. \quad (4.13)$$

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

$$S_{VV} S_{UU}^{-1} S_{UV} \mathbf{H}_1 = S_{VV} \mathbf{H}_1 \mathbf{R}_1^2 = \mathbf{H}_1 \mathbf{R}_1^2, \quad (4.13)$$

$$\mathbf{I} = \mathbf{H}_1' S_{VV} \mathbf{H}_1 = \mathbf{H}_1' \mathbf{H}_1. \quad (4.14)$$

Substitution for $S_{VV}, S_{UU}, \mathbf{H}_1, \mathbf{R}_1^2$ in (4.13) yields

$$S_{VV} I_{(k)} \mathbf{R}_1 + \bar{R}' S_{UV} I_{(k)} - \bar{R}' S_{UU} I_{(k)} \mathbf{R}_1 = 2 I_{(k)} \mathbf{R}_1 R_1^* + \mathbf{H}_1^* \mathbf{R}_1^2 - \bar{R}' \bar{R} \mathbf{H}_1^* + o_p(1), \quad (4.15)$$

14
which is (3.5) with $S_{VV} I_{(k)} R_1^2$ omitted. As in Section 3, (4.15) implies
\[
H_{11}^* R_1 + o_p(1) = (S_{WW}^{*12})'.
\] (4.16)
Substitution in (4.14) yields
\[
H_{11}^{*1} + H_{11}^* = 0 + o_p(1).
\] (4.17)
Then again
\[
\Psi_k^* = \begin{bmatrix}
S_{WW}^{*11} & S_{WW}^{*12} \\
S_{WW}^{*21} & 0
\end{bmatrix} + o_p(1).
\] (4.18)
In (3.11) $v_\alpha$ is nonstochastic and (3.12) holds.

5 Discussion

5.1 Other approaches to the asymptotic covariances

Stoica and Viberg (1996) have obtained an expression for the covariance of the limiting distribution of $\sqrt{n}(\hat{B}_k - B)$ when the $x$'s are nonstochastic satisfying (4.1) by assuming that the $y$'s (or $z$'s) are normally distributed and calculating the Fisher information matrix. The expression is
\[
\mathcal{E}_{n \text{vec} (\hat{B}_k - B) \left( \text{vec} (\hat{B}_k - B) \right)'} \rightarrow \left( \Gamma_1 \otimes I_p, (I_q \otimes \Sigma_{YX} \Gamma_1) \right)
\times \left\{ \begin{bmatrix}
\Gamma_1 \otimes I_p \\
I_q \otimes \Gamma_1' \Sigma_{XY}
\end{bmatrix}
\left[ \Sigma_{XX} \otimes \Sigma_{ZZ}^{-1} \right]
\times \left( (\Gamma_1 \otimes I_p, (I_q \otimes \Sigma_{YX} \Gamma_1) \right)\right\}^+.
\]
\[ \times \left[ \begin{array}{c} \Gamma_1' \otimes I_p \\ I_q \otimes \Gamma_1' \Sigma_{XY} \end{array} \right], \tag{5.1} \]

where \([ \cdot ]^+\) denotes the Moore-Penrose inverse. It is not immediately obvious that (5.1) is the same as (3.14), but when the two expressions are transformed to the canonical variable framework the calculation of equality can be carried out. However, the treatment of Stoica and Viberg does not show that these covariances hold when the \(z\)'s are not normal or when the \(x\)'s are stochastic.

Ahn and Reinsel (1988) consider a general autoregressive model

\[ Y_t = \sum_{j=1}^{m} \Phi_j Y_{t-j} + Z_t, \tag{5.2} \]

where \(\Phi_j = A_j B_j\) and \(A_j\) and \(B_j\) are \(p \times r_j\) matrices with restricted patterns. For \(m = 1\) the model is similar to the one treated here with \(X_\alpha\) replace by \(Y_{t-1}\). They also give an expression that involves the Fisher information matrix. See also Reinsel (1993).

Heldal (1993) has considered a more general setup and obtained a more general expression for the covariance matrix of the asymptotic distribution of \(\text{vec} \tilde{B}_k\). However, it is not clear how the result would be applied here.

### 5.2 Other estimators of the regression matrix.

The unrestricted estimator of \(B\) is the usual regression matrix \(\hat{B} = S_{YX} S_{XX}^{-1}\); in canonical terms the unrestricted estimator of \(\Psi\) is \(\hat{\Psi} = S_{UV} S_{VV}^{-1}\). The normalized error of estimation of \(\Psi\) by \(\hat{\Psi}\) is

\[ \hat{\Psi}^* = \sqrt{n} \left( \hat{\Psi} - \Psi \right) = S_{\hat{\Psi}V} + o_p(1) = \begin{bmatrix} S_{\hat{\Psi}V1}^{11} & S_{\hat{\Psi}V1}^{12} \\ S_{\hat{\Psi}V2}^{11} & S_{\hat{\Psi}V2}^{12} \end{bmatrix} + o_p(1); \tag{5.3} \]
instead of 0 in the lower right-hand corner of \( \hat{\Psi}^* \), there is \( S_{WV}^{22} \), which is independent of the rest of \( \hat{\Psi}^* \). The difference is

\[
\hat{\Psi}^* - \hat{\Psi}_k^* = \begin{bmatrix}
0 & 0 \\
0 & S_{WV}^{22}
\end{bmatrix} + o_p(1). \tag{5.4}
\]

The asymptotic covariance of \( \text{vec} \hat{\Psi}^* \) is

\[
I_q \otimes (I_p - R^2) = \begin{bmatrix}
I_k \otimes (I_p - R^2) & 0 \\
0 & I_{q-k} \otimes (I_p - R^2)
\end{bmatrix}
\]

\[
= \text{diag} \left[ I_k \otimes (I_p - R^2), I_{q-k} \otimes (I_p - R^2) \right], \tag{5.5}
\]

which is to be compared with (3.12). Note the difference between (3.12) and (5.5) in the lower right-hand corner. When the transformation to \( \hat{B} = (A')^{-1} \hat{\Psi} \Gamma' \) is made as in (3.13), we find the asymptotic covariance matrix as \( \Sigma_{XX}^{-1} \otimes \Sigma_{ZZ} \).

The difference of the asymptotic covariance matrix of \( \hat{\Psi} \) given by (5.5) and that of \( \hat{\Psi}_k \) given by (3.12) is positive semi-definite. Similarly, the asymptotic covariance matrix of \( \hat{B} \) is greater than that of \( \hat{B}_k \), indicating that the reduced rank regression is a better estimator than the least squares estimator when \( k \) is known.

The asymptotic covariance matrix of \( \text{vec} S_{WV}^{22} \) is \( I_{q-k} \otimes I_{p-k} \). The asymptotic covariance matrix of \( \text{vec} (\hat{B} - \hat{B}_k) \) is \( \Gamma_2 \Gamma_2' \otimes \Sigma_{ZZ} A_2 A_2' \Sigma_{ZZ} \). The sum of the asymptotic covariance matrices of \( \text{vec} \hat{B}_k \) and \( \text{vec} (\hat{B} - \hat{B}_k) \) is the asymptotic covariance matrix of \( \hat{B} \); This is a kind of analysis of variance.

If it were known that the rank of \( \Psi_k \) was \( k \) and \( U^{(1)}_a = (U_{1a}, \ldots, U_{ka})' \) and \( V^{(1)}_a = (V_{1a}, \ldots, V_{ka})' \)
were the canonical variables with positive canonical correlations, the estimator would be

\[
\tilde{\Psi} = \begin{bmatrix}
S_{UV}^{11} (S_{VV}^{11})^{-1} & 0 \\
0 & 0
\end{bmatrix}.
\]  \tag{5.6}

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

\[
S_{UV}^{11} (S_{VV}^{11})^{-1} - \tilde{\mathbf{R}} = (S_{UV}^{11} - \mathbf{R} S_{VV}^{11}) (S_{VV}^{11})^{-1}
= S_{WW}^{11} (S_{VV}^{11})^{-1}
= \frac{1}{\sqrt{n}} S_{WW}^{11} (I_r + \frac{1}{\sqrt{n}} S_{VV}^{11})^{-1}
= \frac{1}{\sqrt{n}} S_{WW}^{11} + o_p \left( \frac{1}{\sqrt{n}} \right). \tag{5.7}
\]

The error in the other three submatrices in (5.6) would be zero. Of course, this estimator is not feasible, but it shows what use could be made of prior information.

6 Misspecification

6.1 Underestimation of regression rank

In Sections 3 and 4 it was supposed that the rank of \( \mathbf{B} \) was known and that the number of columns of \( \Gamma_1 \) was that rank. In this section we suppose that the number of columns used to define the reduced rank regression is \( s < k \). Let \( \hat{\Gamma} = (\hat{\Gamma}_a, \hat{\Gamma}_b, \hat{\Gamma}_2) = \Gamma(H_a, H_b, H_2) \) of \( s, \, r - s, \) and \( q - k \) columns, respectively. Suppose \( \mathbf{R} = \text{diag}(\mathbf{R}_a, \mathbf{R}_b, \mathbf{0}) \), where \( \mathbf{R}_a \) is \( s \times s \) and \( \mathbf{R}_b \) is \( (k - s) \times (k - s) \). The estimator of \( \mathbf{B} \) is \( \hat{\mathbf{B}}_s = S_{XY} \hat{\Gamma}_a \hat{\Gamma}'_a \) and of \( \Psi \) is \( \hat{\Psi}_s = S_{UV} H_a H'_a \), where \( \hat{\Gamma}_a \) and \( H_a \) have \( s \) columns. Since \( H_a \xrightarrow{p} I(s) \), we define \( H^*_a = \sqrt{n} (H_a - I(s)) \).
The substitution in (3.3) yields (3.4) and (3.5) with \( k \) replaced by \( s \). The partitioning of (3.5) into \( s, k - s, \) and \( q - k \) rows is

\[
\begin{bmatrix}
S_{\nu \bar{\nu}}^{aa} R_a + R_a S_{U \bar{U}}^{aa} - R_a S_{U \bar{U}}^{ba} R_a - S_{\nu \bar{\nu}}^{aa} R_a^2 \\
S_{\nu \bar{\nu}}^{ba} R_a + R_b S_{U \bar{U}}^{ba} - R_b S_{U \bar{U}}^{ba} R_a - S_{\nu \bar{\nu}}^{bb} R_a^2 \\
S_{U \bar{U}}^{ba} - S_{\nu \bar{\nu}}^{2a} R_a R_a^2
\end{bmatrix}
\]

(6.1)

\[
= \begin{bmatrix}
2R_a R_a^* + H_{aa}^* R_a^2 - R_a^2 H_{aa}^* \\
H_{ba}^* R_a^2 - R_b^2 H_{ba}^* \\
H_{2a}^* R_a^2
\end{bmatrix} + o_p(1).
\]

From (6.1) we obtain

\[
H_{2a}^* R_a = S_{\nu \bar{\nu}}^{2a} - S_{U \bar{U}}^{2a} R_a + o_p(1) = \left(S_{\nu \bar{\nu}}^{aa}\right)' + o_p(1),
\]

(6.2)

and (3.17) for \( j = 1, \cdots, s \). The partitioning of the second part of (3.3) gives (3.8) and (3.9) with \( k \) replaced by \( s \). Then from \( \tilde{\Psi}_s = S_{U \bar{U}} H_a H_a' \) we obtain

\[
\tilde{\Psi}_s^* = \begin{bmatrix}
S_{U \bar{U}}^{aa} + R_a (H_{aa}^* + H_{aa}'') & R_a H_{ba}^* & R_a H_{2a}^* \\
S_{U \bar{U}}^{ba} + R_b H_{ba}^* & 0 & 0 \\
S_{U \bar{U}}^{2a} & 0 & 0
\end{bmatrix} + o_p(1)
\]

(6.3)

\[
= \begin{bmatrix}
S_{\nu \bar{\nu}}^{aa} & R_a H_{ba}^* & S_{\nu \bar{\nu}}^{aa} \\
S_{U \bar{U}}^{ba} + R_b H_{ba}^* & 0 & 0 \\
S_{U \bar{U}}^{2a} & 0 & 0
\end{bmatrix} + o_p(1).
\]
It will be seen from (3.17) and (6.3) that $\mathbf{H}_{ba}^*$ and hence $\hat{\mathbf{V}}_s^*$ cannot be expressed simply in terms of $\mathbf{S}_{VV}^s$. Linear combinations of individual elements of $\mathbf{S}_{UU}^s$ and $\mathbf{S}_{VV}^s$ are involved; the asymptotic covariances of $\hat{\mathbf{V}}^*$ will involve second-order moments of $\mathbf{S}_{UU}^s$ and $\mathbf{S}_{VV}^s$.

6.2 Overestimation of regression rank

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

$$
\hat{\mathbf{B}}_{k+t} = \mathbf{S}_{YY} \hat{\mathbf{r}}_{1,c} \hat{\mathbf{r}}_{1,c}' = (\mathbf{A})^{-1} \hat{\mathbf{V}}_{k+t} \mathbf{r}',
$$

(6.4)

where $\hat{\mathbf{r}}_{1,c}$ has $k + t$ columns, and

$$
\hat{\mathbf{V}}_{k+t} = \mathbf{S}_{UU} \mathbf{H}_{1,c} \mathbf{H}_{1,c}',
$$

(6.5)

where $\mathbf{H}_{1,c} = (\mathbf{H}_1, \mathbf{H}_c)$ has $k + t$ columns. We shall study the properties of this estimator. For convenience, we suppose that $q = p$ and that $(\mathbf{Y}', \mathbf{X}')$ has a normal distribution. The probability limits of $\mathbf{I} = \mathbf{H}' \mathbf{S}_{VV} \mathbf{H}$ and

$$
\mathbf{S}_{VV} \mathbf{S}_{UU}^{-1} \mathbf{S}_{UU} \mathbf{H} = \mathbf{S}_{VV} \mathbf{H} \hat{\mathbf{R}}^2,
$$

(6.6)

imply $\mathbf{H}_{11} \xrightarrow{p} \mathbf{I}_r$, $\mathbf{H}_{21} \xrightarrow{p} \mathbf{0}$, $\mathbf{H}_{12} \xrightarrow{p} \mathbf{0}$. (Here $\mathbf{H}$ is partitioned into $k$ and $p - k$ rows and columns.) However, $\mathbf{H}_{22}$ does not converge in probability to a constant matrix; that fact reflects the fact that the solutions of (2.5) for $\rho^2 = 0$ and (2.3) are indeterminate. The matrix $\mathbf{H}_{22}$ can be represented
as

\[ H_{22} = L_2 + \frac{1}{\sqrt{n}} H_{22}' . \]  

(6.7)

Let the singular value decompositon of \( H_{22} \) be \( H_{22} = EDF \) where \( D \) is diagonal and \( E \) and \( F \) are orthogonal; then \( L_2 = EF \) is orthogonal and \( H_{22}' = \sqrt{n}(H_{22} - EF) \). (See Anderson, 1989 and/or 1997 for details and justification.)

Let the first \( t \) columns of \( H_{22} \), \( L_2 \), \( H_{22}' \) be \( H_{2c} \), \( L_c \), \( H_{2c}' \), respectively. Then (6.5) is

\[
\hat{\Psi}_{k+s} = \left[ R + \frac{1}{\sqrt{n}} S_{UV}^* \right] \left[ \begin{pmatrix} I_k & 0 \\ 0 & L_c \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} H_{11}' & H_{1c}' \\ 0 & L_c' \end{pmatrix} \right] \left[ \begin{pmatrix} I_k & 0 \\ 0 & L_c' \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} H_{11}' & H_{1c}' \\ 0 & L_c' \end{pmatrix} \right] \\
+ \left[ \begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} I_k & 0 \\ 0 & L_c' \end{pmatrix} \right] \left[ \begin{pmatrix} R_1 & 0 \\ 0 & 0 \end{pmatrix} \right] + o_p \left( \frac{1}{\sqrt{n}} \right). \tag{6.8}
\]

From (6.8) we obtain

\[
\hat{\Psi}_{k+s}^* = \sqrt{n} \left( \hat{\Psi}_{k+t} - \Psi_{k+t} \right) \tag{6.9}
\]

\[
= \begin{bmatrix} S_{UV}^{*1} & S_{UV}^{*2}L_cL_c' \\ S_{UV}^{*21} & S_{UV}^{*22}L_cL_c' \end{bmatrix} \left[ \begin{array}{cc} R_1H_{11}' & R_1H_{1c}L_c' \\ 0 & 0 \end{array} \right] + \left[ \begin{array}{cc} R_1H_{11}' & R_1H_{12}' \\ 0 & 0 \end{array} \right] + o_p \left( \frac{1}{\sqrt{n}} \right). 
\]

From \( H'S_{VV}H = I \) we obtain

\[
\begin{bmatrix} S_{VV}^{*11} + H_{11}' + H_{11}' & S_{VV}^{*12}L_2 + H_{12}' + H_{21}'L_2 \\ L_2'S_{VV}^{*21} + L_2'H_{21}' + H_{12}' & L_2'S_{VV}^{*22}L_2 + L_2'H_{22}' + H_{22}'L_2 \end{bmatrix} = 0 + o_p(1); \tag{6.10}
\]

21
in particular, $H_{11}^* + H_{11}' = -S_{VV}^{*11} + o_p(1)$. From $S_{UU} S_{UV}^{-1} S_{VV} H = S_{VV} H \tilde{R}^2$ we obtain

$$
\begin{bmatrix}
S_{VV}^{*11}R_1 - R_1 S_{UU}^{*11} R_1 + R_1 S_{VV}^{*11} - S_{VV}^{*11} R_1^2 & R_1 S_{UU}^{*11} L_2 \\
S_{VV}^{*21} R_1 - S_{VV}^{*11} R_1^2 & 0 \\
2R_1 R_1^* + H_{11}^* R_1^2 - R_1^2 H_{11}^* - R_1^2 H_{12}^* \\
H_{21}^* R_1^2 & 0
\end{bmatrix} + o_p(1).
$$

(6.11)

From (6.11) we derive

$$
R_1 H_{21}^{*'} + o_p(1) = \left(S_{VV}^{*21} - S_{VV}^{*11} R_1 \right)' = \left(S_{WW}^{*21} \right)' = S_{VV}^{*12},
$$

(6.12)

$$
R_1 H_{c}^{*} = -S_{UU}^{*12} L_c + o_p(1). 
$$

(6.13)

When the upper left-hand corner of (6.10), (6.12), and (6.13) are substituted into (6.9), we find

$$
\hat{\Psi}_k^{*} = \begin{bmatrix}
S_{VV}^{*11} & S_{VV}^{*12} \\
S_{WW}^{*21} & S_{WW}^{*22} L_c L_c'
\end{bmatrix} + o_p(1).
$$

(6.14)

The difference between $\hat{\Psi}_{k+\ell}^{*}$ and $\hat{\Psi}_k^{*}$ is

$$
\hat{\Psi}_{k+\ell}^{*} - \hat{\Psi}_k^{*} = \begin{bmatrix}
0 & 0 \\
0 & S_{WW}^{*22} L_c L_c'
\end{bmatrix} + o_p(1).
$$

(6.15)

The matrix $L_c$ has a limiting distribution independent of $S_{VV}^{*11}$, $S_{VV}^{*12}$, and $S_{WW}^{*21}$; see Anderson (1997). Thus $S_{WW}^{*22} L_c L_c'$ is independent of $\hat{\Psi}_r$; (6.15) adds an independent error to $\hat{\Psi}_r$. 

22
From (6.8) we see that \( H_{22} - L_2 \xrightarrow{P} 0 \); a part of this statement is that \( H_{2C} - L_c \xrightarrow{P} 0 \). When \( V \) is normal, the limiting distribution of \( L_2 \) is the Haar invariant measure on orthogonal matrices. \( H_{22} \) is not a consistent estimator, but, nevertheless, \( \hat{\Psi}_{k+t} \) is a consistent estimator of \( \Psi \).

### 6.3 Testing rank.

The likelihood ratio criterion (under normality) for testing the null hypothesis that of the rank of \( B \) is \( m \) against alternative that the rank is greater than \( m \) is

\[
\lambda = \prod_{i=m+1}^{p} \frac{1}{(1 + t_i)^{-n/2}} = \prod_{i=m+1}^{p} (1 - \tau_i^2)^{n/2} \tag{6.16}
\]

(Anderson 1951a). The null hypothesis is rejected if the observed value of \( \lambda \) is smaller than a specified value, which is determined by the desired significance level and the fact that under the null hypothesis

\[
-2 \log \lambda = n \sum_{i=m+1}^{p} \log(1 + t_i) = -n \sum_{i=m+1}^{p} \log(1 - \tau_i^2) \tag{6.17}
\]

has a limiting \( \chi^2 \)-distribution with \( (p - m)^2 \) degrees of freedom.

If \( k \), the rank of \( B \) is equal to \( m \), the hypothesized rank, the probability of rejecting the hypothesis is the significance level. If \( k < m \), the probability of rejection is not exactly the significance level but will be small; the investigator will then be overestimating the rank of \( B \). If \( k > m \), the probability of rejection will be greater than the significance level. From (3.6) we find

\[
r_i = \rho_i + (1/\sqrt{n})\tau_i^*, \quad i = 1, \ldots, k.
\]

\[
2r_i^* = 2s_{ii}^{*VU} - \rho_i s_{ii}^{*UU} - \rho_i s_{ii}^{*VV} + o_p(1), \quad i = 1, \ldots, k. \tag{6.18}
\]
For large $n$ the probability of rejection is close to 1; the investigator will avoid underestimating the rank.

Acknowledgements

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

References


