Reduced rank regression in cointegrated models

T.W. Anderson*

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

Received 1 May 2001; revised 30 May 2001; accepted 1 June 2001

Abstract

The coefficient matrix of a cointegrated first-order autoregression is estimated by reduced rank regression (RRR), depending on the larger canonical correlations and vectors of the first difference of the observed series and the lagged variables. In a suitable coordinate system the components of the least-squares (LS) estimator associated with the lagged nonstationary variables are of order $1/T$, where $T$ is the sample size, and are asymptotically functionals of a Brownian motion process; the components associated with the lagged stationary variables are of the order $T^{-1/2}$ and are asymptotically normal. The components of the RRR estimator associated with the stationary part are asymptotically the same as for the LS estimator. Some components of the RRR estimator associated with nonstationary regressors have zero error to order $1/T$ and the other components have a more concentrated distribution than the corresponding components of the LS estimator. © 2002 Elsevier Science S.A. All rights reserved.

JEL classification: C32

Keywords: Nonstationary autoregressions; Least-squares estimators; Error-correction form; Asymptotic distributions

1. Introduction

The type of cointegrated model treated in this paper is a nonstationary autoregressive process of which some linear functions are stationary. Many statistical procedures for the cointegrated models are procedures used for stationary autoregressive processes but the sampling behavior of the statistics

* Tel.: +1-650-723-4732; fax: +1-650-725-8977.
E-mail address: twa@stat.stanford.edu (T.W. Anderson).

0304-4076/02 - see front matter © 2002 Elsevier Science S.A. All rights reserved.
PII: S0304-4076(01)00095-1
reflects the nonstationarity property. The linear autoregressive model of finite order is formally a kind of linear regression. The procedure studied in this paper is the reduced rank regression (RRR) estimator of the coefficient matrix (Anderson, 1951). This estimator is compared with the least-squares (LS) estimator of the matrix and the improvement in efficiency is evaluated.

The classical regression model is

$$Y_t = BX_t + Z_t,$$  \hskip 1cm (1.1)

where $Z_t$ is an unobserved random vector disturbance of $p$ components with $\mathbb{E} Z_t = 0$, $\mathbb{E} Z_t Z'_t = \Sigma_Z$, and $\mathbb{E} X_t Z'_t = 0$, and $X_t$ is an observed random vector of $q$ components. In this paper we specify $\mathbb{E} X_t = 0$ in order to concentrate on the properties of the covariances; in practice sample covariances are calculated in terms of deviations from sample means. Then $\mathbb{E} X_t X'_t = \Sigma_X$ and

$$\mathbb{E} Y_t Y'_t = \Sigma_Y = B \Sigma_X B' + \Sigma_Z.$$  \hskip 1cm (1.2)

Given a sample of $T$ observations $(Y_1, X_1), \ldots, (Y_T, X_T)$ the sample covariances are

$$S_{YY} = \frac{1}{T} \sum_{t=1}^{T} Y_t Y'_t, \quad S_{XX} = \frac{1}{T} \sum_{t=1}^{T} X_t X'_t, \quad S_{YX} = \frac{1}{T} \sum_{t=1}^{T} Y_t X'_t.$$  \hskip 1cm (1.3)

The LS estimator of the coefficient matrix $B$ is

$$\hat{B} = S_{YX} S_{XX}^{-1}.$$  \hskip 1cm (1.4)

Define the residuals as $\hat{Z}_t = Y_t - \hat{B} X_t$, $t = 1, \ldots, T$. The estimator of $\Sigma_Z$ is

$$S_{ZZ} = S_{YY} - \hat{B} S_{XX} \hat{B}'.$$  \hskip 1cm (1.5)

In many problems the rank of $B$ (or some submatrix of $B$) is specified to be $k < \min(p, q)$. For example, the limited information maximum likelihood (LIML) estimator of the vector $\alpha$ satisfying $\alpha' B = 0$ in a simultaneous equation model requires the rank of $B$ to be not greater than $p - 1$ (Anderson and Rubin, 1949). The reduced rank regression estimator of $B$ can be defined in terms of the solutions to the determinantal and vector equations

$$|\hat{B} S_{XX} \hat{B}' - t S_{ZZ}| = 0,$$  \hskip 1cm (1.6)

$$\hat{B} S_{XX} \hat{B}' f = t S_{ZZ} f, \quad f' S_{ZZ} f = 1.$$  \hskip 1cm (1.7)

Let the solutions to (1.6) be ordered $t_1 < \cdots < t_p$, and define the vector $f_i$ as the solution to (1.7) for $t = t_i$. Then define

$$F = (f_{n+1}, \ldots, f_p),$$  \hskip 1cm (1.8)

where $n = p - k$. Then the RRR estimator of $B$ is

$$\hat{B}_k = S_{ZZ} F'F \hat{B}.$$  \hskip 1cm (1.9)

Note that $\hat{B}_k$ is the product $AC$ of the $p \times k$ matrix $A = S_{ZZ} F$ and the $k \times p$ matrix $C = F' \hat{B}$. $A$ and $C$ are normalized by $A' S_{ZZ}^{-1} A = I$ and $\mathbb{C} S_{XX} C' = T =
This estimator was derived by Anderson (1951) as the maximum likelihood estimator when \( Z_1, \ldots, Z_T \) are normally independently distributed.

Alternatively, the RRR estimator is

\[
\hat{B}_k = S_{YX} \hat{\Gamma} \hat{\Gamma}' 
\]

where \( \hat{\Gamma} = (\hat{\gamma}_{n+1}, \ldots, \hat{\gamma}_p) \) and \( \hat{\gamma}_i \) is the solution to

\[
S_{XY} S_{YY}^{-1} S_{YX} \hat{\gamma} = r_i^2 S_{XX} \hat{\gamma}, \quad \hat{\gamma}' S_{XX} \hat{\gamma} = 1
\]

and \( r_i^2 = t_i/(1 + t_i) \) is the square of the \( i \)th canonical correlation between \( Y \) and \( X \). In the form (1.10) the matrices \( S_{YX} \hat{\Gamma} \) and \( \hat{\Gamma}' \) are normalized by \( \hat{\Gamma}' S_{XX} \hat{\Gamma} = I \) and \( (S_{YX} \hat{\Gamma})' S_{YY}^{-1} (S_{YX} \hat{\Gamma}) = \hat{\Gamma}' S_{XY} S_{YY}^{-1} S_{YX} \hat{\Gamma} = \hat{R}^2 = \text{diag}(r_{n+1}^2, \ldots, r_p^2) \). The asymptotic distribution of \( \hat{B}_k \) in this model has been given and discussed by Anderson (1999).

An autoregressive process

\[
Y_t = BY_{t-1} + Z_t
\]

has the form of (1.1) with \( X_t \) replaced by \( Y_{t-1} \). The process is stationary if the roots of

\[
|B - \lambda I| = 0
\]

satisfy \( |\lambda_i| < 1 \). In this case (1.12) can be solved repeatedly so that

\[
Y_t = \sum_{s=0}^{\infty} B^s Z_{t-s}
\]

with covariance matrix

\[
\Sigma_{YY} = \sum_{s=0}^{\infty} B^s \Sigma_{ZZ} B'^s.
\]

The LS estimator is (1.4) and the RRR estimator is (1.9) and (1.10) with the \( X_t \) replaced by \( Y_{t-1} \). In econometric models some components of \( X_t \) can be exogenous variables and some can be components of \( Y_{t-1} \) [as noted by Anderson (1951)].

The likelihood ratio criterion (LRC) for testing the null hypothesis that the rank of \( B \) is \( k \) is

\[
-2 \log \text{LRC} = -T \sum_{i=1}^{n} \log(1 - r_i^2),
\]

when \( n = p - k \). The limiting distribution of (1.16) is \( \chi^2 \) with \( (p - k)(q - k) \) degrees of freedom (Anderson, 1951) when the null hypothesis is true. \( (q = p \) for the AR model.)
2. Nonstationary models; cointegration

2.1. The cointegrated first-order autoregression

The AR process is nonstationary if $|\lambda_i| > 1$ for at least one $i$. A special case is $B = I$; then $\lambda_i = 1$, $i = 1, \ldots, p$, and the model is $Y_t = Y_{t-1} + Z_t$. For $Y_0 = 0$ we obtain $Y_t = \sum_{s=0}^{t-1} Z_{t-s}$ and $\delta Y_t Y_t' = t \Sigma ZZ$. We say \{\begin{align*} Y_t \end{align*}\} is integrated of order 1; in symbols \{\begin{align*} Y_t \end{align*}\} $\in I(1)$. Note that the first difference $\Delta Y_t = Y_t - Y_{t-1} = Z_t$ is integrated of order 0; \{\begin{align*} \Delta Y_t \end{align*}\} $\in I(0)$.

A model is “cointegrated of order $n$” if $|\lambda_i| = 1$, $i = 1, \ldots, n$; and $|\lambda_i| < 1$, $i = n + 1, \ldots, p$. The “error-correction form” of the model is

\begin{align*}
\Delta Y_t &= \Pi Y_{t-1} + Z_t, \tag{2.1}
\end{align*}

where $\Pi = B - I$. Then $n$ of the eigenvalues of $\Pi$ are 0. Note that (2.1) is of the form of (1.1) with $Y_t$ replaced by $\Delta Y_t$ and $X_t$ replaced by $Y_{t-1}$. The LS estimator of $\Pi$ and the RRR estimator have the forms of those given in Section 1.

Define

\begin{align*}
S_{\Delta Y, \Delta Y} &= \frac{1}{T} \sum_{t=1}^{T} \Delta Y_t \Delta Y_t', \\
S_{\Delta Y, \bar{Y}} &= \frac{1}{T} \sum_{t=1}^{T} \Delta Y_t Y_{t-1}', \\
S_{\bar{Y}, \bar{Y}} &= \frac{1}{T} \sum_{t=1}^{T} Y_{t-1} Y_{t-1}'.
\end{align*}

(2.2)

The LS estimator of $\Pi$ is

\begin{align*}
\hat{\Pi} = S_{\Delta Y, \bar{Y}} S_{\bar{Y}, \bar{Y}}^{-1}.
\end{align*}

(2.3)

To find the RRR estimator we solve

\begin{align*}
| S_{\bar{Y}, \Delta Y} S_{\Delta Y, \bar{Y}}^{-1} S_{\Delta Y, \bar{Y}} - r^2 S_{\bar{Y}, \bar{Y}} | = 0
\end{align*}

(2.4)

for $r_1^2 < \cdots < r_p^2$. Then $\hat{r}_i$ is the solution to

\begin{align*}
S_{\bar{Y}, \Delta Y} S_{\Delta Y, \bar{Y}}^{-1} S_{\Delta Y, \bar{Y}} \hat{r}_i = r_i^2 S_{\bar{Y}, \bar{Y}} \hat{r}_i, \quad S_{\bar{Y}, \bar{Y}} \hat{r}_i = 1.
\end{align*}

(2.5)

Then the RRR estimator of $\Pi$ is

\begin{align*}
\hat{\Pi}_k = S_{\Delta Y, \bar{Y}} \hat{r} \hat{r}'.
\end{align*}

(2.6)

2.2. Transformation

In order to study the behavior of $\hat{\Pi}$ and $\hat{\Pi}_k$ we want to change coordinates so as to distinguish the nonstationary and stationary dimensions. We shall assume the following condition.
Condition A. There are \( n \) linearly independent solutions to
\[
\omega' \Pi = 0, \tag{2.7}
\]
where \( n \) is the multiplicity of \( \lambda = 1 \) as a root of the characteristic equation
\[ |B - \lambda I| = 0. \]

Let the solutions of (2.7) be assembled into the matrix \( \Omega_i = (\omega_1, \ldots, \omega_n) \); then \( \Omega_i' \Pi = 0 \) and the rank of \( \Omega_i \) is \( n \). Note that (2.7) is equivalent to \( \omega' B = \omega' \). This assumption implies that \( \{Y_t\} \) is \( I(1) \).

Condition A implies that the rank of \( \Pi \) is \( k = p - n \), and there exists a \( p \times k \) matrix \( \Omega_2 \) such that
\[
\Omega_2' \Pi = \Upsilon_{22} \Omega_2', \tag{2.8}
\]
\( \Upsilon_{22} \) is nonsingular, and \( \Omega = (\Omega_1, \Omega_2) \) is nonsingular. Define
\[
X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_1' Y_t \\ \Omega_2' Y_t \end{bmatrix}, \quad W_t = \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_1' Z_t \\ \Omega_2' Z_t \end{bmatrix}. \tag{2.9}
\]
Then \( (\Delta X'_t, X'_t) \) constitutes a stationary process. These results have been shown in Anderson (2000, 2001b), see also Johansen (1995).

The transformed process \( X_t \) satisfies the autoregressive model
\[
X_t = \Psi X_{t-1} + W_t, \tag{2.10}
\]
\[
\Delta X_t = \Upsilon X_{t-1} + W_t, \tag{2.11}
\]
where
\[
\Psi = \Omega' B (\Omega')^{-1}, \tag{2.12}
\]
\[
\Upsilon = \Omega' \Pi (\Omega')^{-1} = \Psi - I = \begin{bmatrix} 0 & 0 \\ 0 & \Upsilon_{22} \end{bmatrix}. \tag{2.13}
\]

In terms of subvectors the process \( \{X_t\} \) is generated by
\[
X_{1t} = X_{1,t-1} + W_{1t}, \tag{2.14}
\]
\[
X_{2t} = \Psi_{22} X_{2,t-1} + W_{2t}. \tag{2.15}
\]
Let \( X_{10} = X_{1,-1} = \cdots = 0 \) and \( W_{10} = W_{1,-1} = \cdots = 0 \). Then \( X_{1t} = \sum_{s=0}^{t-1} W_{1,t-s} \) constitutes a random walk, and \( \{X_{2t}\} \) is a stationary process.

3. Asymptotic distribution of the LS estimator

3.1. Asymptotic distribution of the sample covariances

The LS estimator of \( \Upsilon \) in (2.11) is
\[
\hat{\Upsilon} = S_{\Delta X, \Delta X} S_{X, \hat{X}}^{-1}, \tag{3.1}
\]
where
\[
S_{\Delta X, \Delta X} = \frac{1}{T} \sum_{t=1}^{T} \Delta X_t \Delta X_t', \quad S_{\Delta X, \tilde{X}} = \frac{1}{T} \sum_{t=1}^{T} \Delta X_t X_{t-1}',
\]
\[
S_{\tilde{X}, \tilde{X}} = \frac{1}{T} \sum_{t=1}^{T} X_{t-1} X_{t-1}'.
\] (3.2)

The deviation of the estimator from the process parameter is
\[
\gamma - \gamma = \Psi - \Psi = S_{\tilde{X}, \tilde{X}}^{-1}.
\] (3.3)

To study the behavior of this statistic we need to distinguish the random walk dimensions from the stationary process dimensions. When \( X_t \) is partitioned into subvectors of \( n \) and \( k \) components \( X_t = (X'_1, X'_2)' \), respectively, we may use the notation
\[
S_{\tilde{X}, \tilde{X}} = (S_{\tilde{X}, \tilde{X}}^{-1}, S_{\tilde{X}, \tilde{X}}^2) \begin{bmatrix} S_{11, \tilde{X}, \tilde{X}} & S_{12, \tilde{X}, \tilde{X}} \\ S_{21, \tilde{X}, \tilde{X}} & S_{22, \tilde{X}, \tilde{X}} \end{bmatrix}.
\] (3.4)

Since \( \{X_t\} \) is a random walk, we need to define a Brownian motion in order to describe the asymptotic distribution of some covariance submatrices. Consider a sequence of random vectors \( \{W_t\} \) with \( \mathbb{E} W_t = 0 \) and \( \mathbb{E} W_t W'_t = \Sigma_{WW} = \Omega' \Sigma_{ZZ} \Omega \). Define \( X(u) \) as the weak limit of
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} W_t = \sqrt{u} \frac{1}{\sqrt{T u}} \sum_{t=1}^{[Tu]} W_t, \quad 0 \leq u \leq 1.
\] (3.5)

For fixed \( u \) the limiting distribution of (3.5) is \( N(0, u \Sigma_{WW}) \) and increments are independent. We partition the vector \( X(u) \) into \( n \) and \( k \) components as \( X(u) = (X'_1(u), X'_2(u))' = X(u) \). Then
\[
\frac{1}{T} S_{X, X}^{11} = \frac{1}{T^2} \sum_{t=1}^{T} X_{1,t-1} X'_{1,t-1} = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sqrt{T}} \sum_{r=0}^{t-2} W_{1,t-1-r} \frac{1}{\sqrt{T}} \sum_{s=0}^{t-2} W'_{1,t-1-s} \xrightarrow{d} \int_0^1 X_1(u) X'_1(u) \, du = I_{11}.
\] (3.6)

See Billingsley (1968) or Johansen (1995), for example.

The second subvector \( X_{2t} = \Psi_2 X_{2,t-1} + W_t, \quad t = \ldots, -1, 0, 1, \ldots \) generates a stationary process \( X_{2t} = \sum_{s=0}^{\infty} \Psi_2^s W_{2,t-s} \). Hence
\[
S_{X, X}^{22} = \frac{1}{T} \sum_{t=1}^{T} X_{2,t-1} X'_{2,t-1} = \sum_{s=0}^{\infty} \Psi_2^s S_{XX}^{22} \Psi_2^s.
\] (3.7)
Before treating $S^{21}$ we consider

$$S_{\Delta X \bar{X}} = \begin{bmatrix} S_{\Delta X \bar{X}}^{11} & S_{\Delta X \bar{X}}^{12} \\ S_{\Delta X \bar{X}}^{21} & S_{\Delta X \bar{X}}^{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \Upsilon_{22} S_{XX}^{21} & \Upsilon_{22} S_{XX}^{22} \end{bmatrix} + \begin{bmatrix} S_{W \bar{X}}^{11} & S_{W \bar{X}}^{12} \\ S_{W \bar{X}}^{21} & S_{W \bar{X}}^{22} \end{bmatrix}. \tag{3.8}$$

We have

$$S_{W \bar{X}}^{1} = \frac{1}{T} \sum_{t=1}^{T} W_{t} X_{1,t-1} \quad \mathbb{d} \to \int_{0}^{1} \text{d} \mathbf{X}(u) \mathbf{X}'(u) = \mathbf{J}_{1} = \begin{bmatrix} \mathbf{J}_{11} \\ \mathbf{J}_{21} \end{bmatrix}, \tag{3.9}$$

$$S_{W \bar{X}}^{2} = \frac{1}{T} \sum_{t=1}^{T} W_{t} X_{2,t-1} \quad \mathbb{p} \to \Sigma_{W \bar{X}}^{2} = \mathbf{0}. \tag{3.10}$$

Since $\{W_{t}, X_{2t}\}$ is stationary

$$S_{W \bar{X}}^{\ast 2} = \sqrt{T} S_{W \bar{X}}^{2} \mathbb{d} \to N(0, ) \tag{3.11}$$

We use summation by parts to evaluate

$$S_{\Delta X \bar{X}}^{21} = \frac{1}{T} \sum_{t=1}^{T} \Delta X_{2t} X_{1,t-1}'$$

$$= \frac{1}{T} \left[ X_{2T} X_{1T}' - \sum_{t=1}^{T-1} X_{2t} \Delta X_{1t}' \right]$$

$$= \frac{1}{T} \left[ X_{2T} X_{1T}' - \sum_{t=1}^{T} (\Psi_{2} X_{2,t-1} + W_{2t}) W_{1t}' \right]$$

$$\mathbb{p} \to -\Sigma_{WW}^{21}. \tag{3.12}$$

Note that

$$\text{tr}(T^{-1} X_{2T} X_{1T}'(T^{-1} X_{2T} X_{1T}')') = (T^{-3/2} X_{1T}' X_{1T})(\text{tr} T^{-1/2} X_{2T} X_{2T}') \mathbb{p} \to 0. \tag{3.13}$$

Then from (3.8), $|\Upsilon_{22}| \neq 0$, and $S_{W \bar{X}}^{21} \mathbb{p} \to \mathbf{J}_{21}$, we conclude that

$$S_{X \bar{X}}^{21} = \Upsilon_{22}^{-1} (S_{\Delta X \bar{X}}^{21} - S_{W \bar{X}}^{21}) \tag{3.14}$$

$$\mathbb{d} \to -\Upsilon_{22}^{-1} (\mathbf{J}_{21} + \Sigma_{WW}^{21}). \tag{3.15}$$
Hence $T^{-1/2}S_{X\tilde{X}}^{21} \xrightarrow{p} 0$. We summarize as
\[
\begin{bmatrix}
\frac{1}{T}S_{X\tilde{X}}^{11} & \frac{1}{\sqrt{T}}S_{X\tilde{X}}^{12} \\
\frac{1}{\sqrt{T}}S_{X\tilde{X}}^{21} & S_{X\tilde{X}}^{22}
\end{bmatrix} \xrightarrow{d} \left[ \begin{array}{cc}
I_{11} & 0 \\
0 & \Sigma_{XX}^{22}
\end{array} \right].
\] (3.16)

More details are given in Anderson (2000).

### 3.2. Asymptotic distribution of the LS estimator

It will be convenient to use the notation $\text{vec} A = (a_1^\prime, \ldots, a_p^\prime)^\prime$ for $A = (a_1, \ldots, a_p)$. A useful relation is $\text{vec} ABC = (C^\prime \otimes A)\text{vec} B$.

**Theorem 1.** Suppose $W_1, W_2, \ldots$ are independently distributed with $E W_t = 0$, $E W_t W_t^\prime = \Sigma_{WW}$, and $E W_t X_{t-s} = 0$, $s = 1, 2, \ldots$. Then
\[
\left[ \sqrt{T}(\hat{\mathbf{y}}_1 - \mathbf{y}_1), \hat{\mathbf{y}}_2 - \mathbf{y}_2 \right] \xrightarrow{p} 0,
\] (3.17)
\[
\left[ T(\hat{\mathbf{y}}_1 - \mathbf{y}_1), \sqrt{T}(\hat{\mathbf{y}}_2 - \mathbf{y}_2) \right] \xrightarrow{d} [J_1 I_{11}, S_{W\tilde{X}}^{*2} (\Sigma_{XX}^{22})^{-1}]
\] (3.18)
and
\[
\text{vec} S_{W\tilde{X}}^{*2} \xrightarrow{d} N(0, \Sigma_{XX}^{22} \otimes \Sigma_{WW}).
\] (3.19)

**Proof.** To demonstrate (3.17) write
\[
\left[ \sqrt{T}(\hat{\mathbf{y}}_1 - \mathbf{y}_1), \hat{\mathbf{y}}_2 - \mathbf{y}_2 \right] = \begin{bmatrix}
\frac{1}{T}S_{W\tilde{X}}^{11} & S_{W\tilde{X}}^{21} \\
\frac{1}{\sqrt{T}}S_{W\tilde{X}}^{21} & S_{W\tilde{X}}^{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
1/T S_{W\tilde{X}}^{11} & 1/\sqrt{T} S_{W\tilde{X}}^{12} \\
1/\sqrt{T} S_{W\tilde{X}}^{21} & S_{W\tilde{X}}^{22}
\end{bmatrix}
\] (3.20)
and use $T^{-1/2}S_{W\tilde{X}}^{11} \xrightarrow{p} 0$, $S_{W\tilde{X}}^{22} \xrightarrow{p} 0$, and (3.16). The limit in distribution (3.18) follows from (3.20), (3.9), (3.10), and (3.16). The limit in distribution (3.19) follows from $\text{vec} S_{W\tilde{X}}^{*2} = T^{-1/2} \sum_{t=1}^T (X_{2,t-1} \otimes W_t)$ and (3.11).

The conditions on \{W_t\} could be weakened. The main point here is that $W_t$ does not have to be normal.

To convert (3.18) to the original terms of $Y_t$ and $Z_t$ we define $S_{Z\tilde{X}}^{ij} = T^{-1} \sum_{t=1}^T Z_{it} X_{i,t-1}$. Then
\[
\hat{B} - B = \left[ \frac{1}{T} (\Omega')^{-1} J_1 I_{11} \right] \Omega_1' + \frac{1}{\sqrt{T}} S_{Z\tilde{X}}^{*2} (\Sigma_{XX}^{22})^{-1} \Omega_2' + [o(T^{-1}), o(T^{-2})].
\] (3.21)
This result is interpreted to mean that the discrepancy $\hat{B} - B$ in the random walk direction multiplied by $T$ is approximately $(\Omega')^{-1}J_1I_1'$ and in the stationary direction multiplied by $\sqrt{T}$ is approximately normal with covariances given by elements of $(\Sigma_{XX}^{22})(\Omega_2')^{-1} \Sigma_{ZZ}$. From (3.21) we derive

$$
\sqrt{T}(\hat{B} - B) \xrightarrow{d} S_2^*\Omega_2(\Omega_2')^{-1} \Sigma_{ZZ}. $$

(3.22)

4. Asymptotic distribution of the RRR estimator

In the $X$-coordinate system the RRR estimator of $Y$ is

$$
\hat{Y}_k = S_{\Delta X,\hat{X}}G_2G_2',
$$

(4.1)

where $G_2 = (g_{n+1}, \ldots, g_p)$. $g_i$ satisfies

$$
S_{\hat{X},\Delta X}S^{-1}_{\Delta X,\Delta X}S_{\Delta X,\hat{X}}g - r^2S_{\hat{X}\hat{X}}g = 1
$$

(4.2)

for $r = r_i$ and $r_i$ (for $r_1 < \cdots < r_p$) satisfies

$$
|S_{\hat{X},\Delta X}S^{-1}_{\Delta X,\Delta X}S_{\Delta X,\hat{X}} - r^2S_{\hat{X}\hat{X}}| = 0.
$$

(4.3)

Theorem 2. Under the conditions of Theorem 1

$$
\begin{bmatrix}
T(\hat{Y}_{k1} - Y_{11})
\sqrt{T}(\hat{Y}_{k2} - Y_{12})
\end{bmatrix}
\xrightarrow{d}
\begin{bmatrix}
0
S_{Y,\hat{X}X}^{21}(\Sigma_{XX}^{22})^{-1}
\end{bmatrix}
$$

(4.4)

and

$$
J_{21} = J_{21} - \Sigma_{Y,\hat{X}X}(\Sigma_{XX}^{11})^{-1}J_{11}
$$

$$
= \int_0^1 d[X_2(u) - \Sigma_{Y,\hat{X}X}(\Sigma_{XX}^{11})^{-1}X_1(u)]X'_1(u).
$$

(4.5)

Proof. The equations (4.2) for $g = g_{n+1}, \ldots, g_p$ are summarized in

$$
QG_2 = S_{\hat{X}\hat{X}}G_2\tilde{G}_2^2, 
G'_2S_{\hat{X}\hat{X}}G_2 = I,
$$

(4.6)
where \( Q = S_{XX,XX} S_{XX,XX}^{-1} S_{XX,XX} S_{XX,XX}^{-1} \) and \( \hat{R}_2 = \text{diag}(\gamma_{n+1}, \ldots, \gamma_{n}) \). Let \( G_2 = (G'_{12}, G'_{22})' \). Then the Eqs. (4.6) can be written as

\[
\frac{1}{T} Q_{11} T G_{12} + Q_{12} G_{22} = \left( \frac{1}{T} S_{XX,XX}^{11} T G_{12} + \frac{1}{T} S_{XX,XX}^{22} G_{22} \right) \hat{R}_2, \tag{4.7}
\]

\[
\frac{1}{T} Q_{21} T G_{12} + Q_{22} G_{22} = \left( \frac{1}{T} S_{XX,XX}^{21} T G_{12} + \frac{1}{T} S_{XX,XX}^{22} G_{22} \right) \hat{R}_2, \tag{4.8}
\]

\[
(TG'_{12}, G'_{22}) = \left[ \begin{array}{cc}
\frac{1}{T} S_{XX,XX}^{11} & \frac{1}{T} S_{XX,XX}^{12} \\
\frac{1}{T} S_{XX,XX}^{21} & S_{XX,XX}^{22}
\end{array} \right] \left[ \begin{array}{c}
TG_{12} \\
G_{22}
\end{array} \right] = I. \tag{4.9}
\]

Define \( H \) by \( TG_{12} = HG_{22} \). Since \( T^{-1} Q_{11} \xrightarrow{p} 0 \), \( T^{-1} Q_{21} \xrightarrow{p} 0 \), and \( T^{-1} S_{XX,XX}^{11} \xrightarrow{p} 0 \), (4.7), (4.8), and (4.9) are asymptotically equivalent to

\[
Q_{12} G_{22} = (I_{11} H + S_{XX,XX}^{12}) G_{22} \hat{R}_2, \tag{4.10}
\]

\[
Q_{22} G_{22} = S_{XX,XX}^{22} G_{22} \hat{R}_2^2, \tag{4.11}
\]

\[
G'_{22} S_{XX,XX}^{22} G_{22} = I. \tag{4.12}
\]

We can solve (4.10) for

\[
H = I_{11}^{-1} (Q_{12} G_{22} \hat{R}_2^{-2} G_{22}^{-1} - S_{XX,XX}^{12})
= I_{11}^{-1} (Q_{12} Q_{22}^{-1} S_{XX,XX}^{22} - S_{XX,XX}^{12}). \tag{4.13}
\]

From

\[
S_{XX,XX} \xrightarrow{d} \begin{bmatrix}
J_{11} & 0 \\
-\Sigma_{XX,XX} & \gamma_{22} \Sigma_{XX,XX}^{22}
\end{bmatrix} \tag{4.14}
\]

and

\[
S_{XX,XX} \xrightarrow{p} \Sigma_{XX,XX} = \begin{bmatrix}
\Sigma_{WW,WW} & \Sigma_{WW,XX}^{12} \\
\Sigma_{WW,XX}^{21} & \Sigma_{WW,XX}^{22} + \gamma_{22} \Sigma_{XX,XX}^{22} \gamma_{22}'
\end{bmatrix}, \tag{4.15}
\]

we calculate

\[
\begin{bmatrix}
Q_{12} \\
Q_{22}
\end{bmatrix} \xrightarrow{d} \begin{bmatrix}
[J_{11}'(\Sigma_{XX,XX}^{-1} )_{12} - S_{WW,WW}^{-1} (\Sigma_{XX,XX}^{-1} )_{22}] \gamma_{22} \Sigma_{XX,XX}^{22} \\
\Sigma_{XX,XX}^{22} \gamma_{22}'(\Sigma_{XX,XX}^{-1} )_{22} \gamma_{22} \Sigma_{XX,XX}^{22}
\end{bmatrix}. \tag{4.16}
\]

Hence

\[
Q_{12} Q_{22}^{-1} S_{XX,XX}^{22} \xrightarrow{d} [J_{11}'(\Sigma_{XX,XX}^{-1} )_{12} (\Sigma_{XX,XX}^{-1} )_{22}^{-1} - \Sigma_{WW,WW}^{-1}] (\gamma_{22}')^{-1}
= -[J_{11}'(\Sigma_{WW,WW}^{-1} )_{12} - \Sigma_{WW,WW}^{-1} (\gamma_{22}')^{-1} \tag{4.17}
\]

and

\[
H \xrightarrow{d} I_{11}^{-1} J_{21}'(\gamma_{22}')^{-1}, \tag{4.18}
\]
where $J_{2:1}$ is given by (4.5). Now we calculate $\hat{\Upsilon}_k = S_{\Delta X, \hat{X}} G_2 G'_2$, where

$$G_2 = \begin{bmatrix} G_{12} \\ G_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ \Gamma_{22} \end{bmatrix} + \begin{bmatrix} \frac{1}{T} H G_{22} \\ \frac{1}{\sqrt{T}} G_{22}^* \end{bmatrix},$$

(4.19)

$$S_{\Delta X, \hat{X}} = \begin{bmatrix} 0 & 0 \\ -\Sigma_{WW}^{21} & \Sigma_{XX}^{22} \end{bmatrix} + \begin{bmatrix} S_{W \hat{X}}^{11} & \frac{1}{\sqrt{T}} S_{W \hat{X}}^{12} \\ \Sigma_{WW}^{21} & \frac{1}{\sqrt{T}} \{ S_{W \hat{X}}^{22} \Sigma_{XX}^{22} + S_{W \hat{X}}^{22} \} \end{bmatrix},$$

(4.20)

Then $\hat{\Upsilon}_k$ is

$$S_{\Delta X, \hat{X}} G_2 G'_2 = \begin{bmatrix} 0 & 0 \\ 0 & \Upsilon_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{T} \Upsilon_{22} \Sigma_{XX}^{22} G_{22} G'_2 H' \\ \frac{1}{T} \Sigma_{WW}^{21} H G_{22} G'_2 + \frac{1}{\sqrt{T}} \{ S_{W \hat{X}}^{22} \Sigma_{XX}^{22} \} \\ \frac{1}{T} \Sigma_{WW}^{21} \Sigma_{XX}^{22} G_{22} G'_2 + \frac{1}{\sqrt{T}} \{ S_{W \hat{X}}^{22} \Sigma_{XX}^{22} + S_{W \hat{X}}^{22} \} \end{bmatrix}$$

+ \begin{bmatrix} 0 & o_p(T^{-1/2}) \\ o_p(T^{-1}) & o_p(T^{-1/2}) \end{bmatrix}. \tag{4.21}

From (4.12) we have

$$I = \Gamma'_{22} \Sigma_{XX}^{22} \Gamma_{22} + \frac{1}{\sqrt{T}} \{ \Gamma'_{22} S_{W \hat{X}}^{22} \Gamma_{22} + \Gamma'_{22} \Sigma_{XX}^{22} G_{22}^* \Sigma_{XX}^{22} \}$$

+ $o_p(T^{-1/2}), \tag{4.22}$

implying

$$0 = (\Sigma_{XX}^{22})^{-1} S_{W \hat{X}}^{22} (\Sigma_{XX}^{22})^{-1} + G_{22}^* \Gamma_{22} + \Gamma_{22} G_{22}^* + o_p(1). \tag{4.23}$$

When (4.23) is used in (4.21), we obtain

$$S_{\Delta X, \hat{X}} G_2 G'_2 = \begin{bmatrix} 0 & 0 \\ 0 & \Upsilon_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{T} \Upsilon_{22} H' \end{bmatrix} + \begin{bmatrix} 0 & o_p(T^{-1/2}) \\ o_p(T^{-1}) & o_p(T^{-1/2}) \end{bmatrix}. \tag{4.24}$$

This gives us (4.4). \qed
5. Discussion

Let us compare the asymptotic distributions of the LS and RRR estimators. The asymptotic distribution of the LS estimator can be expressed as

\[
\begin{bmatrix}
T(\hat{\Sigma}_{11} - \Sigma_{11}) & \sqrt{T}(\hat{\Sigma}_{12} - \Sigma_{12}) \\
T(\hat{\Sigma}_{21} - \Sigma_{21}) & \sqrt{T}(\hat{\Sigma}_{22} - \Sigma_{22})
\end{bmatrix}
\begin{bmatrix}
J_{11} I_{11}^{-1} & S_{12W}^{*} (\Sigma_{xx})^{-1} \\
J_{21} I_{11}^{-1} & S_{22W}^{*} (\Sigma_{xx})^{-1}
\end{bmatrix},
\]

(5.1)

The limiting distribution of \(T(\hat{\Sigma}_{11} - \Sigma_{11})\) is the distribution of \(J_{11} I_{11}^{-1}\), while the limiting distribution of \(T(\hat{\Sigma}_{21} - \Sigma_{21})\) is \(0\). The limiting distribution of \(T(\hat{\Sigma}_{21} - \Sigma_{21})\) is that of \(J_{21} I_{11}^{-1}\) while the limiting distribution of \(T(\hat{\Sigma}_{21} - \Sigma_{21})\) is \(J_{21} I_{11}^{-1} \cdot \frac{1}{T} \Omega_2^{-1} + o(T^{-1})\). The distribution of \(J_{21} I_{11}^{-1} \cdot \frac{1}{T} \Omega_2^{-1} + o(T^{-1})\) is more concentrated around \(0\) than that of \(J_{21}\). For each method the estimator of \(\Sigma_{xx}\) is superefficient.

In the original coordinate system we have

\[
\hat{\Sigma}_{k} - \Sigma = \frac{1}{T} (\Omega')^{-1} \begin{bmatrix}
0 \\
-\Sigma_{WW} (\Sigma_{WW}^{-1})^{-1}, I]J_{11} I_{11}^{-1} \Omega_1
\end{bmatrix}
\]

\[+ \frac{1}{\sqrt{T}} S_{Zx}^{*} (\Sigma_{xx}^{22})^{-1} \Omega_2^{-1} + o(T^{-1}), o(T^{-1/2})
\]

(5.2)

and

\[
\sqrt{T}(\hat{\Sigma}_{k} - \Sigma) \overset{d}{\rightarrow} S_{Zx}^{*} (\Sigma_{xx}^{22})^{-1} \Omega_2
\]

\[= S_{Zx}^{*} \Omega_2 (\Omega_2^{*} \Sigma_{yy} \Omega_2)^{-1} \Omega_2^{-1}.
\]

(5.3)

Note that the limiting distribution of \(\sqrt{T}(\hat{\Sigma}_{k} - \Sigma)\) is the same as the limiting distribution of \(\sqrt{T}(\hat{\Sigma}_{k} - \Sigma)\), but the terms of higher order are different.

The asymptotic distributions developed here do not require normality of the disturbances. In this regard the RRR estimator does not require stronger conditions than the LS estimator. This is in contrast to the fact that the asymptotic distribution of \(G_2\), a factor of \(\hat{\Sigma}_{k}\), does depend on normality of the disturbances.

Johansen (1995) gave (5.3) in Theorem 13.7, but he did not evaluate the term of order \(1/T\) in (5.2).

In the usual regression model (1.1) the two matrices \(\sqrt{T}(\hat{\Psi}_{k} - \Psi)\) and \(\sqrt{T}(\hat{\Psi} - \Psi)\) agree entirely except for the upper left-hand corner of \((\hat{\Psi}_{k} - \Psi_{k})\) being \(\Sigma_{ww}\). However, in the usual regression model the transformation to canonical variables \(U = A'Y\), \(V = \Gamma'X\) yields an error vector \(W = A'Z = U - \Psi V\) with uncorrelated components whereas here the transformation \(\Delta X_t = \Delta Y_t\), \(X_{t-1} = \Delta Y_{t-1}\) yields an error vector \(W = \Omega'Z_t = \Delta X_t - \Psi X_{t-1}\) in which \(\Sigma_{WW}\) may not be \(\Sigma_{ww}\).
6. Higher-order processes

Now we consider an AR($m$) process

$$Y_t = \sum_{j=1}^{m} B_j Y_{t-j} + Z_t,$$  \hspace{1cm} (6.1)

If the roots $\lambda_1, \ldots, \lambda_m$ of

$$|\lambda^m I - \lambda^{m-1}B_1 - \cdots - B_m| = 0$$  \hspace{1cm} (6.2)

satisfy $|\lambda_j| < 1$, $j = 1, \ldots, pm$, (6.1) defines a stationary process. If $\lambda_j = 1$ for one or more values of $j$, the process is nonstationary. The model (6.1) can be put in an “error-correction form”

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{m-1} \Pi_j \Delta Y_{t-j} + Z_t,$$  \hspace{1cm} (6.3)

where $\Pi = \sum_{j=1}^{m} B_j - I$ and $\Pi_i = - \sum_{j=i+1}^{m} B_j$, $i = 1, \ldots, m - 1$. Let the multiplicity of the root $\lambda_j = 1$ be $n$, and define $k = p - n$, so $\lambda_1 = \cdots = \lambda_n = 1$ and $|\lambda_j| < 1$, $j = n + 1, \ldots, pm$. Assume Condition A. Then $\Delta Y_t$ and $\Omega_2^2 Y_t$ can be given initial distributions so that $(\Delta Y_t, \Omega_2^2 Y_t)$ is stationary (Anderson, 2000), and the rank of $\Pi$ is $k$. See also Johansen (1995).

The model (6.3) is of the form $Y = A_1 V_1 + A_2 V_2 + Z$ with $Y$ replaced by $\Delta Y_t$, $A_1$ by $\Pi$, $V_1$ by $Y_{t-1}$, $A_2$ by $(\Pi_1, \ldots, \Pi_{m-1}) = \Pi$ and $V_2$ by $(\Delta Y_t, \ldots, \Delta Y'_{t-m+1})' = \Delta Y_{t-1}$. Then the parameters can be estimated by reduced rank regression (Anderson, 1951). In the definitions of the LS estimator of $\Pi$ (2.3) and the RRR estimator (2.6) $\Delta Y_t$ and $Y_{t-1}$ are replaced by

$$\Delta Y_t - S_{\Delta Y, \Delta Y}^{-1} \Delta Y_{t-1},$$  \hspace{1cm} (6.4)

$$Y_{t-1} - S_{\Delta Y, \Delta Y}^{-1} \Delta Y_{t-1}.$$  \hspace{1cm} (6.5)

The asymptotic distribution of the canonical correlations and vectors has been given by Anderson (2001a) and of the correlations by Hansen and Johansen (1999).

References

