



ELSEVIER

Journal of Econometrics 106 (2002) 203–216

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

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 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).

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

$$\mathbf{Y}_t = \mathbf{B}\mathbf{X}_t + \mathbf{Z}_t, \tag{1.1}$$

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

$$\mathcal{E}\mathbf{Y}_t\mathbf{Y}'_t = \Sigma_{YY} = \mathbf{B}\Sigma_{XX}\mathbf{B}' + \Sigma_{ZZ}. \tag{1.2}$$

Given a sample of T observations $(\mathbf{Y}_1, \mathbf{X}_1), \dots, (\mathbf{Y}_T, \mathbf{X}_T)$ the sample covariances are

$$\mathbf{S}_{YY} = \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_t\mathbf{Y}'_t, \quad \mathbf{S}_{XX} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t\mathbf{X}'_t, \quad \mathbf{S}_{YX} = \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_t\mathbf{X}'_t. \tag{1.3}$$

The LS estimator of the coefficient matrix \mathbf{B} is

$$\hat{\mathbf{B}} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}. \tag{1.4}$$

Define the residuals as $\hat{\mathbf{Z}}_t = \mathbf{Y}_t - \hat{\mathbf{B}}\mathbf{X}_t$, $t = 1, \dots, T$. The estimator of Σ_{ZZ} is

$$\mathbf{S}_{\hat{Z}\hat{Z}} = \mathbf{S}_{YY} - \hat{\mathbf{B}}\mathbf{S}_{XX}\hat{\mathbf{B}}'. \tag{1.5}$$

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

$$|\hat{\mathbf{B}}\mathbf{S}_{XX}\hat{\mathbf{B}}' - t\mathbf{S}_{\hat{Z}\hat{Z}}| = 0, \tag{1.6}$$

$$\hat{\mathbf{B}}\mathbf{S}_{XX}\hat{\mathbf{B}}'\mathbf{f} = t\mathbf{S}_{\hat{Z}\hat{Z}}\mathbf{f}, \quad \mathbf{f}'\mathbf{S}_{\hat{Z}\hat{Z}}\mathbf{f} = 1. \tag{1.7}$$

Let the solutions to (1.6) be ordered $t_1 < \dots < t_p$, and define the vector \mathbf{f}_i as the solution to (1.7) for $t = t_i$. Then define

$$\mathbf{F} = (\mathbf{f}_{n+1}, \dots, \mathbf{f}_p), \tag{1.8}$$

where $n = p - k$. Then the RRR estimator of \mathbf{B} is

$$\hat{\mathbf{B}}_k = \mathbf{S}_{\hat{Z}\hat{Z}}\mathbf{F}\mathbf{F}'\hat{\mathbf{B}}. \tag{1.9}$$

Note that $\hat{\mathbf{B}}_k$ is the product \mathbf{AC} of the $p \times k$ matrix $\mathbf{A} = \mathbf{S}_{\hat{Z}\hat{Z}}\mathbf{F}$ and the $k \times p$ matrix $\mathbf{C} = \mathbf{F}'\hat{\mathbf{B}}$. \mathbf{A} and \mathbf{C} are normalized by $\mathbf{A}'\mathbf{S}_{\hat{Z}\hat{Z}}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{C}\mathbf{S}_{XX}\mathbf{C}' = \mathbf{T} =$

diag(t_{n+1}, \dots, t_p). This estimator was derived by Anderson (1951) as the maximum likelihood estimator when $\mathbf{Z}_1, \dots, \mathbf{Z}_T$ are normally independently distributed.

Alternatively, the RRR estimator is

$$\hat{\mathbf{B}}_k = \mathbf{S}_{YX} \hat{\mathbf{\Gamma}} \hat{\mathbf{\Gamma}}', \tag{1.10}$$

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

$$\mathbf{S}_{XY} \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX} \hat{\gamma} = r^2 \mathbf{S}_{XX} \hat{\gamma}, \quad \hat{\gamma}' \mathbf{S}_{XX} \hat{\gamma} = 1 \tag{1.11}$$

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

An autoregressive process

$$\mathbf{Y}_t = \mathbf{B} \mathbf{Y}_{t-1} + \mathbf{Z}_t \tag{1.12}$$

has the form of (1.1) with \mathbf{X}_t replaced by \mathbf{Y}_{t-1} . The process is stationary if the roots of

$$|\mathbf{B} - \lambda \mathbf{I}| = 0 \tag{1.13}$$

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

$$\mathbf{Y}_t = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{Z}_{t-s} \tag{1.14}$$

with covariance matrix

$$\mathbf{\Sigma}_{YY} = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{\Sigma}_{ZZ} \mathbf{B}^{s'} \tag{1.15}$$

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

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

$$-2 \log \text{LRC} = -T \sum_{i=1}^n \log(1 - r_i^2), \tag{1.16}$$

when $n = p - k$. The limiting distribution of (1.16) is χ^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 $|\lambda_i| \geq 1$ for at least one i . A special case is $\mathbf{B} = \mathbf{I}$; then $\lambda_i = 1, i = 1, \dots, p$, and the model is $\mathbf{Y}_t = \mathbf{Y}_{t-1} + \mathbf{Z}_t$. For $\mathbf{Y}_0 = \mathbf{0}$ we obtain $\mathbf{Y}_t = \sum_{s=0}^{t-1} \mathbf{Z}_{t-s}$ and $\mathcal{E} \mathbf{Y}_t \mathbf{Y}'_t = t \boldsymbol{\Sigma}_{ZZ}$. We say $\{\mathbf{Y}_t\}$ is integrated of order 1; in symbols $\{\mathbf{Y}_t\} \in I(1)$. Note that the first difference $\Delta \mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1} = \mathbf{Z}_t$ is integrated of order 0; $\{\Delta \mathbf{Y}_t\} \in I(0)$.

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

$$\Delta \mathbf{Y}_t = \boldsymbol{\Pi} \mathbf{Y}_{t-1} + \mathbf{Z}_t, \tag{2.1}$$

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

Define

$$\begin{aligned} \mathbf{S}_{\Delta Y, \Delta Y} &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{Y}_t \Delta \mathbf{Y}'_t, & \mathbf{S}_{\Delta Y, \bar{Y}} &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{Y}_t \mathbf{Y}'_{t-1}, \\ \mathbf{S}_{\bar{Y} \bar{Y}} &= \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_{t-1} \mathbf{Y}'_{t-1}. \end{aligned} \tag{2.2}$$

The LS estimator of $\boldsymbol{\Pi}$ is

$$\hat{\boldsymbol{\Pi}} = \mathbf{S}_{\Delta Y, \bar{Y}} \mathbf{S}_{\bar{Y} \bar{Y}}^{-1}. \tag{2.3}$$

To find the RRR estimator we solve

$$|\mathbf{S}_{\bar{Y}, \Delta Y} \mathbf{S}_{\Delta Y, \Delta Y}^{-1} \mathbf{S}_{\Delta Y, \bar{Y}} - r^2 \mathbf{S}_{\bar{Y} \bar{Y}}| = 0 \tag{2.4}$$

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

$$\mathbf{S}_{\bar{Y}, \Delta Y} \mathbf{S}_{\Delta Y, \Delta Y}^{-1} \mathbf{S}_{\Delta Y, \bar{Y}} \hat{\gamma} = r^2 \mathbf{S}_{\bar{Y} \bar{Y}} \hat{\gamma}, \quad \hat{\gamma}' \mathbf{S}_{\bar{Y} \bar{Y}} \hat{\gamma} = 1. \tag{2.5}$$

Then the RRR estimator of $\boldsymbol{\Pi}$ is

$$\hat{\boldsymbol{\Pi}}_k = \mathbf{S}_{\Delta Y, \bar{Y}} \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Gamma}}'. \tag{2.6}$$

2.2. Transformation

In order to study the behavior of $\hat{\boldsymbol{\Pi}}$ and $\hat{\boldsymbol{\Pi}}_k$ 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

$$\omega' \mathbf{\Pi} = \mathbf{0}, \tag{2.7}$$

where n is the multiplicity of $\lambda = 1$ as a root of the characteristic equation $|\mathbf{B} - \lambda \mathbf{I}| = 0$.

Let the solutions of (2.7) be assembled into the matrix $\mathbf{\Omega}_1 = (\omega_1, \dots, \omega_n)$; then $\mathbf{\Omega}'_1 \mathbf{\Pi} = \mathbf{0}$ and the rank of $\mathbf{\Omega}_1$ is n . Note that (2.7) is equivalent to $\omega' \mathbf{B} = \omega'$. This assumption implies that $\{\mathbf{Y}_t\}$ is $I(1)$.

Condition A implies that the rank of $\mathbf{\Pi}$ is $k = p - n$, and there exists a $p \times k$ matrix $\mathbf{\Omega}_2$ such that

$$\mathbf{\Omega}'_2 \mathbf{\Pi} = \mathbf{\Upsilon}_{22} \mathbf{\Omega}'_2, \tag{2.8}$$

$\mathbf{\Upsilon}_{22}$ is nonsingular, and $\mathbf{\Omega} = (\mathbf{\Omega}_1, \mathbf{\Omega}_2)$ is nonsingular. Define

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{X}_{1t} \\ \mathbf{X}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}'_1 \mathbf{Y}_t \\ \mathbf{\Omega}'_2 \mathbf{Y}_t \end{bmatrix}, \quad \mathbf{W}_t = \begin{bmatrix} \mathbf{W}_{1t} \\ \mathbf{W}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}'_1 \mathbf{Z}_t \\ \mathbf{\Omega}'_2 \mathbf{Z}_t \end{bmatrix}. \tag{2.9}$$

Then $(\Delta \mathbf{X}'_{1t}, \mathbf{X}'_{2t})$ constitutes a stationary process. These results have been shown in Anderson (2000, 2001b), see also Johansen (1995).

The transformed process \mathbf{X}_t satisfies the autoregressive model

$$\mathbf{X}_t = \mathbf{\Psi} \mathbf{X}_{t-1} + \mathbf{W}_t, \tag{2.10}$$

$$\Delta \mathbf{X}_t = \mathbf{\Upsilon} \mathbf{X}_{t-1} + \mathbf{W}_t, \tag{2.11}$$

where

$$\mathbf{\Psi} = \mathbf{\Omega}' \mathbf{B} (\mathbf{\Omega}')^{-1}, \tag{2.12}$$

$$\mathbf{\Upsilon} = \mathbf{\Omega}' \mathbf{\Pi} (\mathbf{\Omega}')^{-1} = \mathbf{\Psi} - \mathbf{I} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_{22} \end{bmatrix}. \tag{2.13}$$

In terms of subvectors the process $\{\mathbf{X}_t\}$ is generated by

$$\mathbf{X}_{1t} = \mathbf{X}_{1,t-1} + \mathbf{W}_{1t}, \tag{2.14}$$

$$\mathbf{X}_{2t} = \mathbf{\Psi}_{22} \mathbf{X}_{2,t-1} + \mathbf{W}_{2t}. \tag{2.15}$$

Let $\mathbf{X}_{10} = \mathbf{X}_{1,-1} = \dots = \mathbf{0}$ and $\mathbf{W}_{10} = \mathbf{W}_{1,-1} = \dots = \mathbf{0}$. Then $\mathbf{X}_{1t} = \sum_{s=0}^{t-1} \mathbf{W}_{1,t-s}$ constitutes a random walk, and $\{\mathbf{X}_{2t}\}$ is a stationary process.

3. Asymptotic distribution of the LS estimator

3.1. Asymptotic distribution of the sample covariances

The LS estimator of $\mathbf{\Upsilon}$ in (2.11) is

$$\hat{\mathbf{\Upsilon}} = \mathbf{S}_{\Delta \mathbf{X}, \mathbf{X}} \mathbf{S}_{\mathbf{X}, \mathbf{X}}^{-1}, \tag{3.1}$$

where

$$\begin{aligned}
 \mathbf{S}_{\Delta X, \Delta X} &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_t \Delta \mathbf{X}'_t, & \mathbf{S}_{\Delta X, \bar{X}} &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_t \mathbf{X}'_{t-1}, \\
 \mathbf{S}_{\bar{X} \bar{X}} &= \frac{1}{T} \sum_{t=1}^T \mathbf{X}_{t-1} \mathbf{X}'_{t-1}.
 \end{aligned} \tag{3.2}$$

The deviation of the estimator from the process parameter is

$$\Upsilon - \Upsilon = \Psi - \Psi = \mathbf{S}_{W\bar{X}} \mathbf{S}_{\bar{X}\bar{X}}^{-1}. \tag{3.3}$$

To study the behavior of this statistic we need to distinguish the random walk dimensions from the stationary process dimensions. When \mathbf{X}_t is partitioned into subvectors of n and k components $\mathbf{X}_t = (\mathbf{X}'_{1t}, \mathbf{X}'_{2t})'$, respectively, we may use the notation

$$\mathbf{S}_{W\bar{X}} = (\mathbf{S}_{W\bar{X}}^1, \mathbf{S}_{W\bar{X}}^2) \begin{bmatrix} \mathbf{S}_{W\bar{X}}^{11} & \mathbf{S}_{W\bar{X}}^{12} \\ \mathbf{S}_{W\bar{X}}^{21} & \mathbf{S}_{W\bar{X}}^{22} \end{bmatrix}. \tag{3.4}$$

Since $\{\mathbf{X}_{1t}\}$ 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 $\{\mathbf{W}_t\}$ with $\mathcal{E}\mathbf{W}_t = \mathbf{0}$ and $\mathcal{E}\mathbf{W}_t \mathbf{W}'_t = \Sigma_{WW} = \Omega' \Sigma_{ZZ} \Omega$. Define $\mathbf{X}(u)$ as the weak limit of

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \mathbf{W}_t = \sqrt{u} \frac{1}{\sqrt{Tu}} \sum_{t=1}^{[Tu]} \mathbf{W}_t, \quad 0 \leq u \leq 1. \tag{3.5}$$

For fixed u the limiting distribution of (3.5) is $N(\mathbf{0}, u\Sigma_{WW})$ and increments are independent. We partition the vector $\mathbf{X}(u)$ into n and k components as $\mathbf{X}(u) = (\mathbf{X}'_1(u), \mathbf{X}'_2(u))' = \mathbf{X}(u)$. Then

$$\begin{aligned}
 \frac{1}{T} \mathbf{S}_{\bar{X}\bar{X}}^{11} &= \frac{1}{T^2} \sum_{t=1}^T \mathbf{X}_{1,t-1} \mathbf{X}'_{1,t-1} = \frac{1}{T} \sum_{t=1}^T \frac{1}{\sqrt{T}} \sum_{r=0}^{t-2} \mathbf{W}_{1,t-1-r} \frac{1}{\sqrt{T}} \sum_{s=0}^{t-2} \mathbf{W}'_{1,t-1-s} \\
 &\stackrel{d}{\rightarrow} \int_0^1 \mathbf{X}_1(u) \mathbf{X}'_1(u) du = \mathbf{I}_{11}.
 \end{aligned} \tag{3.6}$$

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

The second subvector $\mathbf{X}_{2t} = \Psi_2 \mathbf{X}_{2,t-1} + \mathbf{W}_t$, $t = \dots, -1, 0, 1, \dots$ generates a stationary process $\mathbf{X}_{2t} = \sum_{s=0}^{\infty} \Psi_2^s \mathbf{W}_{2,t-s}$. Hence

$$\mathbf{S}_{\bar{X}\bar{X}}^{22} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_{2,t-1} \mathbf{X}'_{2,t-1} \xrightarrow{p} \Sigma_{XX}^{22} = \sum_{s=0}^{\infty} \Psi_2^s \Sigma_{WW}^{22} \Psi_2^{s'}. \tag{3.7}$$

Before treating \mathbf{S}^{21} we consider

$$\mathbf{S}_{\Delta X, \bar{X}} = \begin{bmatrix} \mathbf{S}_{\Delta X, \bar{X}}^{11} & \mathbf{S}_{\Delta X, \bar{X}}^{12} \\ \mathbf{S}_{\Delta X, \bar{X}}^{21} & \mathbf{S}_{\Delta X, \bar{X}}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Upsilon}_{22} \mathbf{S}_{\bar{X}\bar{X}}^{21} & \boldsymbol{\Upsilon}_{22} \mathbf{S}_{\bar{X}\bar{X}}^{22} \end{bmatrix} + \begin{bmatrix} \mathbf{S}_{W\bar{X}}^{11} & \mathbf{S}_{W\bar{X}}^{12} \\ \mathbf{S}_{W\bar{X}}^{21} & \mathbf{S}_{W\bar{X}}^{22} \end{bmatrix}. \tag{3.8}$$

We have

$$\begin{aligned} \mathbf{S}_{W\bar{X}}^{11} &= \frac{1}{T} \sum_{t=1}^T \mathbf{W}_t \mathbf{X}'_{1,t-1} \\ &\xrightarrow{d} \int_0^1 d\mathbf{X}(u) \mathbf{X}'_1(u) = \mathbf{J}_{.1} = \begin{bmatrix} \mathbf{J}_{11} \\ \mathbf{J}_{21} \end{bmatrix}, \end{aligned} \tag{3.9}$$

$$\begin{aligned} \mathbf{S}_{W\bar{X}}^{22} &= \frac{1}{T} \sum_{t=1}^T \mathbf{W}_t \mathbf{X}'_{2,t-1} \\ &\xrightarrow{p} \boldsymbol{\Sigma}_{W\bar{X}}^{22} = \mathbf{0}. \end{aligned} \tag{3.10}$$

Since $\{\mathbf{W}_t, \mathbf{X}_{2t}\}$ is stationary

$$\mathbf{S}_{W\bar{X}}^{*2} = \sqrt{T} \mathbf{S}_{W\bar{X}}^{22} \xrightarrow{d} N(\mathbf{0}, \cdot). \tag{3.11}$$

We use summation by parts to evaluate

$$\begin{aligned} \mathbf{S}_{\Delta X, \bar{X}}^{21} &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{2t} \mathbf{X}'_{1,t-1} \\ &= \frac{1}{T} \left[\mathbf{X}_{2T} \mathbf{X}'_{1T} - \sum_{t=1}^T \mathbf{X}_{2t} \Delta \mathbf{X}'_{1t} \right] \\ &= \frac{1}{T} \left[\mathbf{X}_{2T} \mathbf{X}'_{1T} - \sum_{t=1}^T (\boldsymbol{\Psi}_2 \mathbf{X}_{2,t-1} + \mathbf{W}_{2t}) \mathbf{W}'_{1t} \right] \\ &\xrightarrow{p} -\boldsymbol{\Sigma}_{WW}^{21}. \end{aligned} \tag{3.12}$$

Note that

$$\text{tr}(T^{-1} \mathbf{X}_{2T} \mathbf{X}'_{1T})(T^{-1} \mathbf{X}_{2T} \mathbf{X}'_{1T})' = (T^{-3/2} \mathbf{X}'_{1T} \mathbf{X}_{1T})(\text{tr } T^{-1/2} \mathbf{X}_{2T} \mathbf{X}'_{2T}) \xrightarrow{p} \mathbf{0}. \tag{3.13}$$

Then from (3.8), $|\boldsymbol{\Upsilon}_{22}| \neq 0$, and $\mathbf{S}_{W\bar{X}}^{21} \xrightarrow{p} \mathbf{J}_{21}$, we conclude that

$$\mathbf{S}_{\bar{X}\bar{X}}^{21} = \boldsymbol{\Upsilon}_{22}^{-1} (\mathbf{S}_{\Delta X, \bar{X}}^{21} - \mathbf{S}_{W\bar{X}}^{21}) \tag{3.14}$$

$$\xrightarrow{d} -\boldsymbol{\Upsilon}_{22}^{-1} (\mathbf{J}_{21} + \boldsymbol{\Sigma}_{WW}^{21}). \tag{3.15}$$

Hence $T^{-1/2}S_{\bar{X}\bar{X}}^{21} \xrightarrow{P} \mathbf{0}$. We summarize as

$$\begin{bmatrix} \frac{1}{T}S_{\bar{X}\bar{X}}^{11} & \frac{1}{\sqrt{T}}S_{\bar{X}\bar{X}}^{12} \\ \frac{1}{\sqrt{T}}S_{\bar{X}\bar{X}}^{21} & S_{\bar{X}\bar{X}}^{22} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{XX}^{22} \end{bmatrix}. \tag{3.16}$$

More details are given in Anderson (2000).

3.2. Asymptotic distribution of the LS estimator

It will be convenient to use the notation $\text{vec } \mathbf{A} = (\mathbf{a}'_1, \dots, \mathbf{a}'_p)'$ for $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_p)$. A useful relation is $\text{vec } \mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A})\text{vec } \mathbf{B}$.

Theorem 1. Suppose $\mathbf{W}_1, \mathbf{W}_2, \dots$ are independently distributed with $\mathcal{E}\mathbf{W}_t = \mathbf{0}$, $\mathcal{E}\mathbf{W}_t\mathbf{W}'_t = \Sigma_{WW}$, and $\mathcal{E}\mathbf{W}_t\mathbf{X}'_{t-s} = \mathbf{0}$, $s = 1, 2, \dots$. Then

$$[\sqrt{T}(\hat{\mathbf{Y}}_{\cdot 1} - \mathbf{Y}_{\cdot 1}), \mathbf{Y}_{\cdot 2} - \mathbf{Y}_{\cdot 2}] \xrightarrow{P} \mathbf{0}, \tag{3.17}$$

$$[T(\hat{\mathbf{Y}}_{\cdot 1} - \mathbf{Y}_{\cdot 1}), \sqrt{T}(\hat{\mathbf{Y}}_{\cdot 2} - \mathbf{Y}_{\cdot 2})] \xrightarrow{d} [\mathbf{J}_{\cdot 1}\mathbf{I}_{11}^{-1}, \mathbf{S}_{W\bar{X}}^{* \cdot 2}(\Sigma_{XX}^{22})^{-1}] \tag{3.18}$$

and

$$\text{vec } \mathbf{S}_{W\bar{X}}^{* \cdot 2} \xrightarrow{d} N(\mathbf{0}, \Sigma_{XX}^{22} \otimes \Sigma_{WW}). \tag{3.19}$$

Proof. To demonstrate (3.17) write

$$[\sqrt{T}(\hat{\mathbf{Y}}_{\cdot 1} - \mathbf{Y}_{\cdot 1}), \hat{\mathbf{Y}}_{\cdot 2} - \mathbf{Y}_{\cdot 2}] = \left[\frac{1}{\sqrt{T}}\mathbf{S}_{W\bar{X}}^1, \mathbf{S}_{W\bar{X}}^2 \right] \begin{bmatrix} \frac{1}{T}S_{\bar{X}\bar{X}}^{11} & \frac{1}{\sqrt{T}}S_{\bar{X}\bar{X}}^{12} \\ \frac{1}{\sqrt{T}}S_{\bar{X}\bar{X}}^{21} & S_{\bar{X}\bar{X}}^{22} \end{bmatrix}^{-1} \tag{3.20}$$

and use $T^{-1/2}S_{W\bar{X}}^1 \xrightarrow{P} \mathbf{0}$, $S_{W\bar{X}}^2 \xrightarrow{P} \mathbf{0}$, 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 $\text{vec } \mathbf{S}_{W\bar{X}}^{* \cdot 2} = T^{-1/2} \sum_{t=1}^T (\mathbf{X}_{2,t-1} \otimes \mathbf{W}_t)$ and (3.11). \square

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

To convert (3.18) to the original terms of \mathbf{Y}_t and \mathbf{Z}_t we define $\mathbf{S}_{Z\bar{X}}^j = T^{-1} \sum_{t=1}^T \mathbf{Z}_t\mathbf{X}'_{j,t-1}$. Then

$$\hat{\mathbf{B}} - \mathbf{B} = \left[\frac{1}{T}(\mathbf{\Omega}'_1)^{-1}\mathbf{J}_{\cdot 1}\mathbf{I}_{11}^{-1}\mathbf{\Omega}'_1 + \frac{1}{\sqrt{T}}\mathbf{S}_{Z\bar{X}}^{* \cdot 2}(\Sigma_{XX}^{22})^{-1}\mathbf{\Omega}'_2 \right] + [o(T^{-1}), o(T^{-\frac{1}{2}})]. \tag{3.21}$$

This result is interpreted to mean that the discrepancy $\hat{\mathbf{B}} - \mathbf{B}$ in the random walk direction multiplied by T is approximately $(\mathbf{\Omega}')^{-1} \mathbf{J}_1 \mathbf{I}'_1$ and in the stationary direction multiplied by \sqrt{T} is approximately normal with covariances given by elements of $(\mathbf{\Sigma}_{XX}^{22})^{-1} \otimes \mathbf{\Sigma}_{ZZ} = (\mathbf{\Omega}'_2 \mathbf{\Sigma}_{YY} \mathbf{\Omega}_2)^{-1} \otimes \mathbf{\Sigma}_{ZZ}$. From (3.21) we derive

$$\begin{aligned} \sqrt{T}(\hat{\mathbf{B}} - \mathbf{B}) &\stackrel{d}{\rightarrow} \mathbf{S}_{Z\bar{X}}^{*,2} (\mathbf{\Sigma}_{XX}^{22})^{-1} \mathbf{\Omega}'_2 \\ &= \mathbf{S}_{Z\bar{Y}}^* \mathbf{\Omega}_2 (\mathbf{\Omega}'_2 \mathbf{\Sigma}_{YY} \mathbf{\Omega}_2)^{-1} \mathbf{\Omega}'_2. \end{aligned} \tag{3.22}$$

4. Asymptotic distribution of the RRR estimator

In the X -coordinate system the RRR estimator of Υ is

$$\hat{\Upsilon}_k = \mathbf{S}_{\Delta X, \bar{X}} \mathbf{G}_2 \mathbf{G}'_2, \tag{4.1}$$

where $\mathbf{G}_2 = (\mathbf{g}_{n+1}, \dots, \mathbf{g}_p)$, \mathbf{g}_i satisfies

$$\mathbf{S}_{\bar{X}, \Delta X} \mathbf{S}_{\Delta X, \Delta X}^{-1} \mathbf{S}_{\Delta X, \bar{X}} \mathbf{g} = r^2 \mathbf{S}_{\bar{X}\bar{X}} \mathbf{g}, \quad \mathbf{g}' \mathbf{S}_{\bar{X}\bar{X}} \mathbf{g} = 1 \tag{4.2}$$

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

$$|\mathbf{S}_{\bar{X}, \Delta X} \mathbf{S}_{\Delta X, \Delta X}^{-1} \mathbf{S}_{\Delta X, \bar{X}} - r^2 \mathbf{S}_{\bar{X}\bar{X}}| = 0. \tag{4.3}$$

Theorem 2. Under the conditions of Theorem 1

$$\begin{bmatrix} T(\hat{\Upsilon}_k^{11} - \Upsilon_{11}) & \sqrt{T}(\hat{\Upsilon}_k^{12} - \Upsilon_{12}) \\ T(\hat{\Upsilon}_k^{21} - \Upsilon_{21}) & \sqrt{T}(\hat{\Upsilon}_k^{22} - \Upsilon_{22}) \end{bmatrix} \stackrel{d}{\rightarrow} \begin{bmatrix} \mathbf{0} & \mathbf{S}_{W\bar{X}}^{*12} (\mathbf{\Sigma}_{XX}^{22})^{-1} \\ \mathbf{J}_{2,1} \mathbf{I}_{11}^{-1} & \mathbf{S}_{W\bar{X}}^{*22} (\mathbf{\Sigma}_{XX}^{22})^{-1} \end{bmatrix} \tag{4.4}$$

and

$$\begin{aligned} \mathbf{J}_{2,1} &= \mathbf{J}_{21} - \mathbf{\Sigma}_{WW}^{21} (\mathbf{\Sigma}_{WW}^{11})^{-1} \mathbf{J}_{11} \\ &= \int_0^1 d[\mathbf{X}_2(u) - \mathbf{\Sigma}_{WW}^{21} (\mathbf{\Sigma}_{WW}^{11})^{-1} \mathbf{X}_1(u)] \mathbf{X}'_1(u). \end{aligned} \tag{4.5}$$

Proof. The equations (4.2) for $\mathbf{g} = \mathbf{g}_{n+1}, \dots, \mathbf{g}_p$ are summarized in

$$\mathbf{Q} \mathbf{G}_2 = \mathbf{S}_{\bar{X}\bar{X}} \mathbf{G}_2 \hat{\mathbf{R}}_2^2, \quad \mathbf{G}'_2 \mathbf{S}_{\bar{X}\bar{X}} \mathbf{G}_2 = \mathbf{I}, \tag{4.6}$$

where $\mathbf{Q} = \mathbf{S}_{\bar{X}, \Delta X} \mathbf{S}_{\Delta X, \Delta X}^{-1} \mathbf{S}_{\Delta X, \bar{X}}$ and $\hat{\mathbf{R}}_2 = \text{diag}(\Upsilon_{n+1}, \dots, \Upsilon_n)$. Let $\mathbf{G}_2 = (\mathbf{G}'_{12}, \mathbf{G}'_{22})'$. Then the Eqs. (4.6) can be written as

$$\frac{1}{T} \mathbf{Q}_{11} T \mathbf{G}_{12} + \mathbf{Q}_{12} \mathbf{G}_{22} = \left(\frac{1}{T} \mathbf{S}_{\bar{X}\bar{X}}^{11} T \mathbf{G}_{12} + \mathbf{S}_{\bar{X}\bar{X}}^{22} \mathbf{G}_{22} \right) \hat{\mathbf{R}}_2^2, \tag{4.7}$$

$$\frac{1}{T} \mathbf{Q}_{21} T \mathbf{G}_{12} + \mathbf{Q}_{22} \mathbf{G}_{22} = \left(\frac{1}{T} \mathbf{S}_{\bar{X}\bar{X}}^{21} T \mathbf{G}_{12} + \mathbf{S}_{\bar{X}\bar{X}}^{22} \mathbf{G}_{22} \right) \hat{\mathbf{R}}_2^2, \tag{4.8}$$

$$(T \mathbf{G}'_{12}, \mathbf{G}'_{22}) \begin{bmatrix} \frac{1}{T^2} \mathbf{S}_{\bar{X}\bar{X}}^{11} & \frac{1}{T} \mathbf{S}_{\bar{X}\bar{X}}^{12} \\ \frac{1}{T} \mathbf{S}_{\bar{X}\bar{X}}^{21} & \mathbf{S}_{\bar{X}\bar{X}}^{22} \end{bmatrix} \begin{bmatrix} T \mathbf{G}_{12} \\ \mathbf{G}_{22} \end{bmatrix} = \mathbf{I}. \tag{4.9}$$

Define \mathbf{H} by $T \mathbf{G}_{12} = \mathbf{H} \mathbf{G}_{22}$. Since $T^{-1} \mathbf{Q}_{11} \xrightarrow{p} \mathbf{0}$, $T^{-1} \mathbf{Q}_{21} \xrightarrow{p} \mathbf{0}$, and $T^{-2} \mathbf{S}_{\bar{X}\bar{X}}^{11} \xrightarrow{p} \mathbf{0}$, (4.7), (4.8), and (4.9) are asymptotically equivalent to

$$\mathbf{Q}_{12} \mathbf{G}_{22} = (\mathbf{I}_{11} \mathbf{H} + \mathbf{S}_{\bar{X}\bar{X}}^{12}) \mathbf{G}_{22} \hat{\mathbf{R}}_2^2, \tag{4.10}$$

$$\mathbf{Q}_{22} \mathbf{G}_{22} = \mathbf{S}_{\bar{X}\bar{X}}^{22} \mathbf{G}_{22} \hat{\mathbf{R}}_2^2, \tag{4.11}$$

$$\mathbf{G}'_{22} \mathbf{S}_{\bar{X}\bar{X}}^{22} \mathbf{G}_{22} = \mathbf{I}. \tag{4.12}$$

We can solve (4.10) for

$$\begin{aligned} \mathbf{H} &= \mathbf{I}_{11}^{-1} (\mathbf{Q}_{12} \mathbf{G}_{22} \hat{\mathbf{R}}_2^{-2} \mathbf{G}_{22}^{-1} - \mathbf{S}_{\bar{X}\bar{X}}^{12}) \\ &= \mathbf{I}_{11}^{-1} (\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{S}_{\bar{X}\bar{X}}^{22} - \mathbf{S}_{\bar{X}\bar{X}}^{12}). \end{aligned} \tag{4.13}$$

From

$$\mathbf{S}_{\Delta X, \bar{X}} \xrightarrow{d} \begin{bmatrix} \mathbf{J}_{11} & \mathbf{0} \\ -\Sigma_{WW}^{21} & \Upsilon_{22} \Sigma_{XX}^{22} \end{bmatrix} \tag{4.14}$$

and

$$\mathbf{S}_{\Delta X, \Delta X} \xrightarrow{p} \Sigma_{\Delta X, \Delta X} = \begin{bmatrix} \Sigma_{WW}^{11} & \Sigma_{WW}^{12} \\ \Sigma_{WW}^{21} & \Sigma_{WW}^{22} + \Upsilon_{22} \Sigma_{XX}^{22} \Upsilon'_{22} \end{bmatrix}, \tag{4.15}$$

we calculate

$$\begin{bmatrix} \mathbf{Q}_{12} \\ \mathbf{Q}_{22} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} [\mathbf{J}'_{11} (\Sigma_{\Delta X, \Delta X}^{-1})_{12} - \Sigma_{WW}^{12} (\Sigma_{\Delta X, \Delta X}^{-1})_{22}] \Upsilon_{22} \Sigma_{XX}^{22} \\ \Sigma_{XX}^{22} \Upsilon'_{22} (\Sigma_{\Delta X, \Delta X}^{-1})_{22} \Upsilon_{22} \Sigma_{XX}^{22} \end{bmatrix}. \tag{4.16}$$

Hence

$$\begin{aligned} \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{S}_{\bar{X}\bar{X}}^{22} &\xrightarrow{d} [\mathbf{J}'_{11} (\Sigma_{\Delta X, \Delta X}^{-1})_{12} (\Sigma_{\Delta X, \Delta X}^{-1})_{22}^{-1} - \Sigma_{WW}^{12}] (\Upsilon'_{22})^{-1} \\ &= -[\mathbf{J}'_{11} (\Sigma_{WW}^{11})^{-1} \Sigma_{WW}^{12} - \Sigma_{WW}^{12}] (\Upsilon'_{22})^{-1} \end{aligned} \tag{4.17}$$

and

$$\mathbf{H} \xrightarrow{d} \mathbf{I}_{11}^{-1} \mathbf{J}'_{2,1} (\Upsilon'_{22})^{-1}, \tag{4.18}$$

where $\mathbf{J}_{2,1}$ is given by (4.5). Now we calculate $\hat{\mathbf{Y}}_k = \mathbf{S}_{\Delta X, \bar{X}} \mathbf{G}_2 \mathbf{G}'_2$, where

$$\mathbf{G}_2 = \begin{bmatrix} \mathbf{G}_{12} \\ \mathbf{G}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Gamma}_{22} \end{bmatrix} + \begin{bmatrix} \frac{1}{T} \mathbf{H} \mathbf{G}_{22} \\ \frac{1}{\sqrt{T}} \mathbf{G}_{22}^* \end{bmatrix}, \tag{4.19}$$

$$\mathbf{S}_{\Delta X, \bar{X}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{\Sigma}_{WW}^{21} & \mathbf{\Upsilon}_2 \mathbf{\Sigma}_{XX}^{22} \end{bmatrix} + \begin{bmatrix} \mathbf{S}_{W\bar{X}}^{11} & \frac{1}{\sqrt{T}} \mathbf{S}_{W\bar{X}}^{*12} \\ \mathbf{S}_{\Delta X, \bar{X}}^{21} + \mathbf{\Sigma}_{WW}^{21} & \frac{1}{\sqrt{T}} (\mathbf{\Upsilon}_2 \mathbf{S}_{\bar{X}\bar{X}}^{*22} + \mathbf{S}_{W\bar{X}}^{*22}) \end{bmatrix}. \tag{4.20}$$

Then $\hat{\mathbf{Y}}_k$ is

$$\begin{aligned} \mathbf{S}_{\Delta X, \bar{X}} \mathbf{G}_2 \mathbf{G}'_2 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_{22} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{T}} \mathbf{S}_{W\bar{X}}^{*12} (\mathbf{\Sigma}_{XX}^{22})^{-1} \\ \frac{1}{T} \mathbf{\Upsilon}_2 \mathbf{\Sigma}_{\bar{X}\bar{X}}^{22} \mathbf{\Gamma}_{22} \mathbf{G}'_{22} \mathbf{H}' & \frac{1}{T} \mathbf{\Sigma}_{WW}^{21} \mathbf{H} \mathbf{G}_{22} \mathbf{\Gamma}'_{22} + \frac{1}{\sqrt{T}} \{ \mathbf{S}_{W\bar{X}}^{*22} (\mathbf{\Sigma}_{XX}^{22})^{-1} \\ & + \mathbf{\Upsilon}_2 [\mathbf{S}_{\bar{X}\bar{X}}^{*22} (\mathbf{\Sigma}_{XX}^{22})^{-1} + \mathbf{\Sigma}_{XX}^{22} \mathbf{\Gamma}_{22} \mathbf{G}'_{22} + \mathbf{\Sigma}_{XX}^{22} \mathbf{\Gamma}_{22}^* \mathbf{\Gamma}'_{22}] \} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{0} & o_p(T^{-1/2}) \\ o_p(T^{-1}) & o_p(T^{-1/2}) \end{bmatrix}. \end{aligned} \tag{4.21}$$

From (4.12) we have

$$\begin{aligned} \mathbf{I} &= \mathbf{\Gamma}'_{22} \mathbf{\Sigma}_{XX}^{22} \mathbf{\Gamma}_{22} + \frac{1}{\sqrt{T}} (\mathbf{\Gamma}'_{22} \mathbf{S}_{\bar{X}\bar{X}}^{*22} \mathbf{\Gamma}_{22} + \mathbf{\Gamma}'_{22} \mathbf{\Sigma}_{XX}^{22} \mathbf{G}_{22}^* + \mathbf{G}_{22}^{*'} \mathbf{\Sigma}_{XX}^{22} \mathbf{\Gamma}_{22}) \\ &+ o_p(T^{-1/2}), \end{aligned} \tag{4.22}$$

implying

$$\mathbf{0} = (\mathbf{\Sigma}_{XX}^{22})^{-1} \mathbf{S}_{\bar{X}\bar{X}}^{*22} (\mathbf{\Sigma}_{XX}^{22})^{-1} + \mathbf{G}_{22}^* \mathbf{\Gamma}'_{22} + \mathbf{\Gamma}_{22} \mathbf{G}_{22}^{*'} + o_p(1). \tag{4.23}$$

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

$$\begin{aligned} \mathbf{S}_{\Delta X, \bar{X}} \mathbf{G}_2 \mathbf{G}'_2 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_{22} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \frac{1}{\sqrt{T}} \mathbf{S}_{W\bar{X}}^{*12} (\mathbf{\Sigma}_{XX}^{22})^{-1} \\ \frac{1}{T} \mathbf{\Upsilon}_2 \mathbf{H}' & \frac{1}{\sqrt{T}} \mathbf{S}_{W\bar{X}}^{*22} (\mathbf{\Sigma}_{XX}^{22})^{-1} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{0} & o_p(T^{-\frac{1}{2}}) \\ o_p(T^{-1}) & o_p(T^{-\frac{1}{2}}) \end{bmatrix}. \end{aligned} \tag{4.24}$$

This gives us (4.4). \square

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

$$\begin{bmatrix} T(\hat{\Upsilon}_{11} - \Upsilon_{11}) & \sqrt{T}(\hat{\Upsilon}_{12} - \Upsilon_{12}) \\ T(\hat{\Upsilon}_{21} - \Upsilon_{21}) & \sqrt{T}(\hat{\Upsilon}_{22} - \Upsilon_{22}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{J}_{11}\mathbf{I}_{11}^{-1} & \mathbf{S}_{WX}^{*12}(\boldsymbol{\Sigma}_{XX}^{22})^{-1} \\ \mathbf{J}_{21}\mathbf{I}_{11}^{-1} & \mathbf{S}_{WX}^{*22}(\boldsymbol{\Sigma}_{XX}^{22})^{-1} \end{bmatrix}. \tag{5.1}$$

The limiting distribution of $T(\hat{\Upsilon}_{11} - \Upsilon_{11})$ is the distribution of $\mathbf{J}_{11}\mathbf{I}_{11}^{-1}$, while the limiting distribution of $T(\hat{\Upsilon}_k - \Upsilon_k)$ is $\mathbf{0}$. The limiting distribution of $T(\hat{\Upsilon}_{21} - \Upsilon_{21})$ is that of $\mathbf{J}_{21}\mathbf{I}_{11}^{-1}$ while the limiting distribution of $T(\hat{\Upsilon}_k^{21} - \Upsilon_k^{21})$ is $\mathbf{J}_{2,1}\mathbf{I}_{11}^{-1}$. The distribution of $\mathbf{J}_{2,1}$ is more concentrated around $\mathbf{0}$ than that of \mathbf{J}_{21} . For each method the estimator of $\Upsilon_{\cdot 1}$ is superefficient.

In the original coordinate system we have

$$\begin{aligned} \hat{\mathbf{B}}_k - \mathbf{B} &= \frac{1}{T}(\boldsymbol{\Omega}')^{-1} \begin{bmatrix} \mathbf{0} \\ [-\boldsymbol{\Sigma}_{WW}^{21}(\boldsymbol{\Sigma}_{WW}^{11})^{-1}, \mathbf{I}]\mathbf{J}_{\cdot 1}\mathbf{I}_{11}^{-1}\boldsymbol{\Omega}'_1 \end{bmatrix} \\ &+ \frac{1}{\sqrt{T}}\mathbf{S}_{ZX}^*(\boldsymbol{\Sigma}_{XX}^{22})^{-1}\boldsymbol{\Omega}'_2 + [o(T^{-1}), o(T^{-1/2})] \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} \sqrt{T}(\hat{\mathbf{B}}_k - \mathbf{B}) &\xrightarrow{d} \mathbf{S}_{ZX}^*(\boldsymbol{\Sigma}_{XX}^{22})^{-1}\boldsymbol{\Omega}'_2 \\ &= \mathbf{S}_{ZY}^*\boldsymbol{\Omega}_2(\boldsymbol{\Omega}'_2\boldsymbol{\Sigma}_{YY}\boldsymbol{\Omega}_2)^{-1}\boldsymbol{\Omega}'_2. \end{aligned} \tag{5.3}$$

Note that the limiting distribution of $\sqrt{T}(\hat{\mathbf{B}}_k - \mathbf{B})$ is the same as the limiting distribution of $\sqrt{T}(\hat{\mathbf{B}} - \mathbf{B})$, 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 \mathbf{G}_2 , a factor of $\hat{\Upsilon}_k$, 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 $\mathbf{0}$ (Anderson, 1999). However, in the usual regression model the transformation to canonical variables $\mathbf{U} = \mathbf{A}'\mathbf{Y}$, $\mathbf{V} = \boldsymbol{\Gamma}'\mathbf{X}$ yields an error vector $\mathbf{W} = \mathbf{A}'\mathbf{Z} = \mathbf{U} - \boldsymbol{\Psi}\mathbf{V}$ with uncorrelated components whereas here the transformation $\Delta\mathbf{X}_t = \boldsymbol{\Omega}'\Delta\mathbf{Y}_t$, $\mathbf{X}_{t-1} = \boldsymbol{\Omega}'\mathbf{Y}_{t-1}$ yields an error vector $\mathbf{W} = \boldsymbol{\Omega}'\mathbf{Z}_t = \Delta\mathbf{X}_t - \boldsymbol{\Psi}\mathbf{X}_{t-1}$ in which $\boldsymbol{\Sigma}_{WW}^{12}$ may not be $\mathbf{0}$.

6. Higher-order processes

Now we consider an AR(m) process

$$\mathbf{Y}_t = \sum_{j=1}^m \mathbf{B}_j \mathbf{Y}_{t-j} + \mathbf{Z}_t. \tag{6.1}$$

If the roots $\lambda_1, \dots, \lambda_{pm}$ of

$$|\lambda^m \mathbf{I} - \lambda^{m-1} \mathbf{B}_1 - \dots - \mathbf{B}_m| = 0 \tag{6.2}$$

satisfy $|\lambda_j| < 1, j = 1, \dots, pm$, (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”

$$\Delta \mathbf{Y}_t = \mathbf{\Pi} \mathbf{Y}_{t-1} + \sum_{j=1}^{m-1} \mathbf{\Pi}_j \Delta \mathbf{Y}_{t-j} + \mathbf{Z}_t, \tag{6.3}$$

where $\mathbf{\Pi} = \sum_{j=1}^m \mathbf{B}_j - \mathbf{I}$ and $\mathbf{\Pi}_i = -\sum_{j=i+1}^m \mathbf{B}_j, i = 1, \dots, m - 1$.

Let the multiplicity of the root $\lambda_j = 1$ be n , and define $k = p - n$, so $\lambda_1 = \dots = \lambda_n = 1$ and $|\lambda_j| < 1, j = n + 1, \dots, pm$. Assume Condition A. Then $\Delta \mathbf{Y}_t$ and $\mathbf{\Omega}'_2 \mathbf{Y}_t$ can be given initial distributions so that $(\Delta \mathbf{Y}_t, \mathbf{\Omega}'_2 \mathbf{Y}_t)$ is stationary (Anderson, 2000), and the rank of $\mathbf{\Pi}$ is k . See also Johansen (1995).

The model (6.3) is of the form $\mathbf{Y} = \mathbf{A}_1 \mathbf{V}_1 + \mathbf{A}_2 \mathbf{V}_2 + \mathbf{Z}$ with \mathbf{Y} replaced by $\Delta \mathbf{Y}_t$, \mathbf{A}_1 by $\mathbf{\Pi}$, \mathbf{V}_1 by \mathbf{Y}_{t-1} , \mathbf{A}_2 by $(\mathbf{\Pi}_1, \dots, \mathbf{\Pi}_{m-1}) = \mathbf{\Pi}$ and \mathbf{V}_2 by $(\Delta \mathbf{Y}_t, \dots, \Delta \mathbf{Y}'_{t-m+1})' = \bar{\Delta} \mathbf{Y}_{t-1}$. Then the parameters can be estimated by reduced rank regression (Anderson, 1951). In the definitions of the LS estimator of $\mathbf{\Pi}$ (2.3) and the RRR estimator (2.6) $\Delta \mathbf{Y}_t$ and \mathbf{Y}_{t-1} are replaced by

$$\Delta \mathbf{Y}_t - \mathbf{S}_{\Delta Y, \bar{\Delta} Y} \mathbf{S}_{\bar{\Delta} Y, \bar{\Delta} Y}^{-1} \bar{\Delta} \mathbf{Y}_{t-1}, \tag{6.4}$$

$$\mathbf{Y}_{t-1} - \mathbf{S}_{\bar{Y}, \bar{\Delta} Y} \mathbf{S}_{\bar{\Delta} Y, \bar{\Delta} Y}^{-1} \bar{\Delta} \mathbf{Y}_{t-1}. \tag{6.5}$$

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

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351. [Correction, *Annals of Statistics* 8, 1980, p. 1400.]
 Anderson, T.W., 1999. Asymptotic distribution of the reduced rank regression estimator under general conditions. *Annals of Statistics* 70, 1–29.
 Anderson, T.W., 2000. The asymptotic distribution of canonical correlations and variates in cointegrated models. *Proceedings of the National Academy of Sciences* 97, 7068–7073.

- Anderson, T.W., 2001a. The asymptotic distribution of canonical correlations and variates in higher-order cointegrated models. *Proceedings of the National Academy of Sciences* 98, 4860–4865.
- Anderson, T.W., 2001b. A condition for cointegration, unpublished.
- Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63.
- Billingsley, P., 1968. *Convergence of Probability Measures*. Wiley, New York.
- Hansen, H., Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR-models. *Econometrics Journal* 3, 306–333.
- Johansen, S., 1995. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.