The asymptotic distribution of canonical correlations and vectors in higher-order cointegrated models

T. W. Anderson*

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

Contribution by T. W. Anderson, February 6, 2001

The study of the large-sample distribution of the canonical correlations and variates in cointegrated models is extended from the first-order autoregression model to autoregression of any (finite) order. The cointegrated process considered here is nonstationary in some dimensions and stationary in some other directions, but the first difference (the “error-correction form”) is stationary. The asymptotic distribution of the canonical correlations between the first differences and the predictor variables as well as the corresponding canonical variables is obtained under the assumption that the process is Gaussian. The method of analysis is similar to that used for the first-order process.

Cointegrated stochastic processes are used in econometrics for modeling macroeconomic time series that have both stationary and nonstationary properties. The term “cointegrated” means that in a multivariate process that appears nonstationary some linear functions are stationary. Many economic time series may show inflationary tendencies or increasing volatility, but certain relationships are not affected by these tendencies. Statistical inference is involved in identifying these relationships and estimating their importance.

The family of stochastic processes studied in this paper consists of vector autoregressive processes of finite order. A vector of contemporary measures is considered to depend linearly on earlier values of these measures plus random disturbances or errors. The dependence may be evaluated by the canonical correlations and corresponding canonical vectors. The smaller the canonical correlations different from 0, the better the process will be nonstationary. In this paper, we assume that n (0 < n < p) roots of |B(λ)| = 0 satisfy |λ| < 1, and the other pm − n roots satisfy |λ| < 1, i = n + 1, . . . , pm. The first difference of the process, the “error-correction” form, is

\[ Y_t - Y_{t-1} = \Delta Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).

The Model

The general cointegrated model is an autoregressive process \( \{Y_t\} \) of order \( m \) defined by

\[ Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).

\[ Y_t - Y_{t-1} = \Delta Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).

\[ Y_t - Y_{t-1} = \Delta Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).

\[ Y_t - Y_{t-1} = \Delta Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).

\[ Y_t - Y_{t-1} = \Delta Y_t = B_1 Y_{t-1} + \cdots + B_m Y_{t-m} + Z_t, \]

where \( Z_t \) is unobserved with \( \mathbb{E}Z_t = 0, \mathbb{E}Z_t'Z_t = \Sigma_Z \), and \( \mathbb{E}Y_{t-1}Z_t = 0 \).

The likelihood ratio test for the degree of cointegration that I found (1) is given in Asymptotic Distribution of the Smaller Roots; its asymptotic distribution under the null hypothesis was found by Johansen (4). To evaluate the power of such a test, one needs to know the distribution or asymptotic distribution of the sample canonical correlations corresponding to process canonical correlations different from 0. See ref. 5, for example.

For further background, the reader is referred to Johansen (6) and Reinsel and Velu (7).
where $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \tilde{Y}_{t-1} \tilde{Y}_{t-1}^T$, $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \tilde{Y}_{t-1} \tilde{X}_{t-1}$, and $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \tilde{Y}_{t-1} \tilde{X}_{t-1}$ are the residuals of $\tilde{Y}_{t-1}$ and $\tilde{X}_{t-1}$, respectively.

**The vectors $Y_t$ and $X_t$** are the sample residuals of $\Delta Y_{t-1}$ and $\Delta X_{t-1}$, respectively. Define $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta Y_{t-1} \Delta Y_{t-1}^T$, $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta Y_{t-1} \Delta X_{t-1}^T$, and $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta X_{t-1} \Delta X_{t-1}^T$, where $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta Y_{t-1} \Delta Y_{t-1}^T$, $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta Y_{t-1} \Delta X_{t-1}^T$, and $S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta X_{t-1} \Delta X_{t-1}^T$. The sample canonical correlations between $\Delta Y_{t-1}$ and $\Delta X_{t-1}$ and variances are defined by

$$
S_{2Y,XY} = T^{-1} \sum_{t=1}^{T} \Delta Y_{t-1} \Delta Y_{t-1}^T - \hat{r}^2 S_{2Y,XY}^T = 0.
$$

More information on canonical analysis is covered in chapter 12 of ref. 8. One form of the reduced rank regression estimator is

$$
\Omega_{(k)} = S_{2Y,XY} \Gamma_{(k)} S_{2Y,XY}^T, \quad \Gamma_{(k)} = (\gamma_{(k),1}, \ldots, \gamma_{(k),p})
$$

We shall assume that there are exactly $n$ linearly independent solutions to $\omega \Omega = 0$, that is, $\omega \Omega = 0$. Then the rank of $\Omega$ is $p - n = k$ and there exists a $p \times n$ matrix $\Omega$ of rank $n$ such that $\Omega \Omega^T = 0$. See Anderson (9). There is also a $p \times n$ matrix $\Omega$ of rank $k$ such that $\Omega \Omega = Y \Omega \Omega^T = Y \Omega \Omega^T$, where $Y$ is a matrix ($k \times k$) that is nonsingular, and $\Omega = (\Omega_1, \Omega_2)$ is nonsingular.

To distinguish between the stationary and nonstationary coordinates, we make a transformation of coordinates. Define

$$
\Omega \Psi_j = X_j = \begin{bmatrix} \Psi_1 \\ \vdots \\ \Psi_m \end{bmatrix}, \quad \Omega \Psi_j = W_j = \begin{bmatrix} W_{1j} \\ \vdots \\ W_{mj} \end{bmatrix},
$$

If we define $Y = \Psi_1 + \cdots + \Psi_m - \Omega = \Omega \Omega \Omega^{-1}, Y_j = -\sum_{j=1}^{m} \Psi_j = \Omega \Omega \Omega^{-1}, \quad Y = (Y_1, \ldots, Y_m-1)$, and $\Delta X_{t-1} = (\Delta X_{t-1}, \ldots, \Delta X_{t-m+1})^T$, the form 2 is transformed to

$$
\Delta X_{t-1} = YY_{t-1} + \Delta X_{t-1} + W_{t-1}.
$$

Note that $Y = \text{diag}(0, Y_{22})$. If we define $\Delta X_{t-1}, \Delta X_{t-2}, S_{2X,2X}, S_{2X,2X}, S_{2X,2X}, S_{2X,2X}, S_{2X,2X}$ and $S_{2X,2X}$ in a manner analogous to the definitions in the $Y$-coordinates, the reduced rank regression estimator of $Y$ is based on the canonical correlations and canonical variables between $\Delta X_{t-1}^*$ and $X_{t-1}^*$ defined by

$$
S_{2X,2X}^* S_{2X,2X}^* S_{2X,2X}^* - \tilde{r}^2 S_{2X,2X}^* = 0.
$$

The estimator of $Y$ of rank $k$ is $\hat{Y}(k) = \hat{S}_{2X,2X} \hat{G}_2 \hat{G}_2^T$, where $G_2 = (g_{m+1}, \ldots, g_k)$ and $\tilde{g}$ is the solution for $g$ in 8 when $r = r_{t+1}$, the solution to 7 and $r_1 \cdots r_p$. The rest of this paper is devoted to finding the asymptotic distribution of $(g, r)$. Note that $\hat{Y}(k) = \hat{\Omega} \hat{\Omega}^{-(k)} \hat{\Omega}^{-1}$.

The vectors $\Delta X_{t-1} = \Delta X_{t-1} - S_{2X,2X} S_{2X,2X} \Delta X_{t-1}$ and $\Delta X_{t-1} = X_{t-1} - S_{2X,2X} X_{t-1}$ are the residuals of $\Delta X_{t-1}$ and $\Delta X_{t-1}$, respectively. $\Delta X_{t-1}$ is the maximum correlation between $\Delta X_{t-1}$ and $\Delta X_{t-1}$, which is the correlation between $\Delta X_{t-1}$ and $\Delta X_{t-1}$ after taking account of the dependence “explained” by $\Delta X_{t-1}$.

The canonical correlations are the canonical correlations between $\Delta X_{t-1}$ and $\Delta X_{t-1}$ and $(X_{t-1} - \Delta X_{t-1})$ other than $\pm 1$.

**The Process**

The process $\{X_t\}$ defined by 5 can be put in the form of the Markov model

$$
\begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \Psi_1 & \Psi_2 & \ldots & \Psi_{m-1} & \Psi_m \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \end{bmatrix} + \begin{bmatrix} W_t \\ 0 \\ \vdots \end{bmatrix}.
$$

Yields a form that includes the error-correction form 6

$$
\begin{bmatrix} X_t \\ Y_t \\ \vdots \\ Y_{t-1} \\ \end{bmatrix} = \begin{bmatrix} Y_{t+1} - Y_t \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \end{bmatrix} + \begin{bmatrix} W_{t+1} \\ \vdots \\ \vdots \\ \vdots \\ \end{bmatrix}.
$$

The first $n$ components of 10 constitute

$$
X_t = X_{t-1} + \sum_{j=1}^{m-1} (Y_{j+1} - Y_{j-1}) + Y_{j+1}^{(2)}(X_{j+1} - X_{j-1}) + W_{t+1}.
$$

Write 12 as

$$
X_{t-1} = \Gamma \sum_{j=1}^{m-1} W_{t-j} + \hat{\Delta} X_{t-1}.
$$

where $\hat{\Delta} X_{t-1}$ is a linear combination of $Y_{1-1}, \ldots, Y_{m-1}$, and $\hat{\Delta} X_{t-1} = (X_{m}, \hat{\Delta} X_{m})$. [The matrix on the left-hand side of 12 is nonsingular because otherwise there would be a linear combination of the right-hand side identically 0.] The right-hand side of 13 is the sum of a stationary process and a random walk $(\hat{S}_{2X,2X} W_{t+1})$.

The last $pm - n = k + p(m - 1)$ components of 10 constitute a stationary process satisfying

$$
\hat{X}_t = \hat{\Delta} X_{t-1} + W_t.
$$
where $X_t = (X_t^0, \Delta X_t)$, $W_t = (W_t^0, W_t^s, 0)$, and $Y$ consists of the last $pm - n$ rows and columns of the coefficient matrix in 10. Note that the first $n$ columns and last $pm - n$ rows of that matrix consist of 0s. Because the eigenvalues of $Y$ are less than 1 in absolute value (9), $X_t = \sum_{s=0}^{\infty} Y^s W_{t-s}, \quad \Sigma = \sum_{s=0}^{\infty} Y^s \Sigma Y^*$. The covariance $\Sigma$ satisfies

$$\Sigma = Y \Sigma Y^* + \Sigma_{\text{partial}}.$$  

Given $Y$ and $\Sigma_{\text{partial}}$, 15 can be solved for $\Sigma$ [Anderson (10), section 5.5]. Further we write 13 as $X_t = \Gamma \sum_{s=0}^{\infty} W_{t-s} + H \sum_{s=0}^{\infty} Y^s W_{t-s}$. Then

$$\mathcal{E} X_t X_t' = tR \Sigma_{\text{partial}} \Gamma' + \Gamma \Sigma_{\text{partial}} \Gamma (I - Y^{-1} Y^*)^{-1} Y^* H' + H (I - Y^{-1} Y^*)^{-1} \Gamma'$$

since $I - Y^{-1} Y^* \rightarrow I$. Here $\mathcal{E} W_t W_t' = \Sigma_{\text{partial}}$ is the second set of rows in $\Sigma_{\text{partial}}$. Then $T^{-1} \Sigma_{\text{partial}} \rightarrow \Sigma_{\text{partial}}$, $\mathcal{E} X_t X_t' \rightarrow 2^{-1} \Gamma \Sigma_{\text{partial}} \Gamma'$ because $\Sigma_{t=1}^T = T (T+1)/2$. Further

$$\mathcal{E} X_t X_t' = \Gamma \left( \sum_{s=0}^{t-1} W_{t-s} + H \sum_{s=0}^{\infty} Y^s W_{t-s} \right) \sum_{s=0}^{\infty} W_{t-s} Y^s$$

$$= \Gamma \Sigma_{\text{partial}} (I - Y^{-1} Y^*)^{-1} + H \Sigma$$

Define

$$\Delta X_t = \Delta X_t - \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X}^{-1} \Delta X_t - 1, \quad \Delta X_t^- = X_t - 1 - \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X}^{-1} \Delta X_t - 1,$$

where $\Sigma_{\Delta X, \Delta X} = \Sigma_{\Delta X, \Delta X}$. $\Delta X_t$ and $\Delta X^-_t$ correspond to $\Delta X_t$ and $\Delta X^-_t$ of 13 with $\Sigma_{\Delta X, \Delta X}$. $\Sigma_{\Delta X, \Delta X}$ replaced by $\Sigma_{\Delta X, \Delta X}$, $\Sigma_{\Delta X, \Delta X}$ and $\Sigma_{\Delta X, \Delta X}$ respectively. Then 6 can be written as a regression model

$$\Delta X_t = \Sigma_{\Delta X, \Delta X} \Delta X_t^- + W_t$$

with $\Delta X_t^- W_t = 0$. Note that this model has the form of 2.10 in Anderson (2).

From 16 and 17 we calculate

$$\mathcal{E} \Delta X_t \Delta X_t' = \begin{bmatrix} 0 & 0 \\ 0 & Y^{22} \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} \end{bmatrix} = \Sigma_{\Delta X, \Delta X},$$

$$\mathcal{E} \Delta X_t^- \Delta X_t^-' = \begin{bmatrix} 0 & 0 \\ 0 & Y^{22} \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} \end{bmatrix} = \Sigma_{\Delta X, \Delta X},$$

$$\mathcal{E} X_t X_t^- = \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} \Delta X_t^- = \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} \Delta X_t^- = \Sigma_{\Delta X, \Delta X},$$

The process analogs of 7 and 8 are

$$\mathcal{E} X_t X_t^- = \Sigma_{\Delta X, \Delta X} \Delta X_t^- = \rho^2 \Sigma_{\Delta X, \Delta X} \Delta X_t^- = \mathcal{E} X_t X_t^-.$$  

These define the process canonical correlations and variances in the $X$-coordinates.

**Sample Statistics**

The canonical correlations and vectors depend on $\Sigma_{\Delta X, \Delta X}^+, \Sigma_{\Delta X, \Delta X}^+, \Sigma_{\Delta X, \Delta X}^+$, which in turn depend on the submatrices of $\Sigma_{\Delta X, \Delta X}$, $\Sigma_{\Delta X, \Delta X}$, and $\Sigma_{\Delta X, \Delta X}$ (equivalently $\Sigma_{\Delta X, \Delta X}$, $\Sigma_{\Delta X, \Delta X}$). The vector $X_t$ satisfies the first-order stationary autoregressive model 14. The sample covariances matrices $\Sigma_{\Delta X, \Delta X}$, $\Sigma_{\Delta X, \Delta X}$, and $\Sigma_{\Delta X, \Delta X}$ are consistent estimators of $\Sigma$, 0, and $\Sigma_{\Delta X, \Delta X}$, respectively. From 2 and 11.

Let $W(u)$ be the Brownian motion process defined by $T^{-1/2} \sum_{t=1}^{T} W_t \Rightarrow W(u)$. Define $J_{t1}$ by

$$T^{-1} W_t dW_t \rightarrow T^{-1/2} \sum_{t=1}^{T} \int_0^1 dW(t)W(t)du = I_{11}.$$  

See Anderson (2) and theorem B.12 of Johansen (6). Define $J_{1j}$ by

$$T^{-1} \sum_{t=1}^{T} W_t dW_t \rightarrow T^{-1/2} \sum_{t=1}^{T} \int_0^1 dW(t)W(t) = J_{1j}, \quad j = 1, 2.$$  

Then $T^{-1} \Sigma_{\Delta X} \rightarrow \Pi_{11} \Gamma'$ by 13, $T^{-1} \Sigma_{\Delta X} \rightarrow 0$, and the Cauchy-Schwarz inequality.

We shall find the limit in distribution of $S_{\Delta X}^+$ from the limit of $S_{\Delta X}^+$ by using equation B.20 of theorem B.13 of Johansen (6). A specialization to the model here is

$$T^{-1} \sum_{t=1}^{T} \sum_{s=0}^{t-1} \sum_{i=1}^{r} \int_0^1 dW(t)W(t) + \Sigma_{\text{partial}} = \Sigma_{\text{partial}}.$$  

where $W(u) = [W_2(u), W_1(u), 0]$. [In theorem B.13, let $\theta = (1, 0), \psi = (0, \bar{Y}), \bar{e}_i = (W_2, W_1)$, and $\Omega = \mathcal{E} e_i' e_i$, $V_i = X_i$.] Then

$$S_{\Delta X}^+ = \sum_{t=1}^{T} \sum_{s=0}^{t-1} \sum_{i=1}^{r} \int_0^1 dW(t)W(t) + \Sigma_{\text{partial}} \Gamma'$$

$$+ \Sigma_{\text{partial}} = (I - \bar{Y})^{-1} [J_{1j}, J_{1j}, J_{1j}, 0] + \Sigma_{\text{partial}} \Gamma' + \Sigma_{\text{partial}}.$$  

Because $\{X_i\}$ is stationary, $T^{-1} \Sigma_{i=1}^{r} W_i X_i \rightarrow 0$ and

$$S_{\Delta X}^+ = \begin{bmatrix} S_{\Delta X}^+ & S_{\Delta X}^+ & S_{\Delta X}^+ \\ S_{\Delta X}^+ & S_{\Delta X}^+ & S_{\Delta X}^+ \\ S_{\Delta X}^+ & S_{\Delta X}^+ & S_{\Delta X}^+ \end{bmatrix} \rightarrow [J_{1j} \Gamma', 0].$$

Now we wish to show that $\Delta X_t^+$ and $\Delta X_t^+$ lead to the same asymptotic results as $\Delta X_t^+$ and $\Delta X_t^+$, first note that $T^{-1} S_{\Delta X}^1 \rightarrow \Pi_{11} \Gamma'$ and $T^{-1} \times$ any other sample covariance converges in probability to 0. Hence $T^{-1} S_{\Delta X}^2 \rightarrow \Pi_{11} \Gamma'$ and $T^{-1} S_{\Delta X}^3 \rightarrow \Pi_{11} \Gamma'$. Because $\{X_i\}$ is stationary, $\{X_i\}$ is stationary, and $S_{\Delta X}^2 \rightarrow \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X}$ and $S_{\Delta X}^3 \rightarrow \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X} \Sigma_{\Delta X}$ have the limiting normal distribution. (See Asymptotic Distribution of the Larger Roots.) Finally, $T^{-1/2} S_{\Delta X}^1 \rightarrow 0$ and $T^{-1/2} S_{\Delta X}^2 \rightarrow 0$ because $S_{\Delta X}^1, S_{\Delta X}^1, S_{\Delta X}^1, S_{\Delta X}^1, S_{\Delta X}^1, S_{\Delta X}^1$ and $S_{\Delta X}^1$ have finite limits in distribution.

From 17 we find that $\text{plim}_{T \rightarrow \infty} S_{\Delta X, \Delta X} = S_{\Delta X, \Delta X}$ and

$$\Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X} = \rho^2 \Sigma_{\Delta X, \Delta X} \Sigma_{\Delta X, \Delta X}.$$
\[ S_{\DeltaX \DeltaX} = \begin{bmatrix} 0 & \gamma^{22} S_{\DeltaX \DeltaX}^{22} \\ \gamma^{22} S_{\DeltaX \DeltaX}^{22} & 0 \end{bmatrix} + S_{\DeltaX \DeltaX}' \]

where \( S_{\DeltaX \DeltaX}' = S_{\DeltaX} - S_{\DeltaX \DeltaX} \Sigma_{\DeltaX \DeltaX}^{-1} \Sigma_{\DeltaX \DeltaX} \), which converges in distribution to the right-hand side of (20).

As noted above, \( S_{\DeltaX \DeltaX}^{2k} \times S_{\DeltaX \DeltaX}^{21} \), which consists of the first \( k \) rows of the weak limit of \( S_{\DeltaX \DeltaX} \). Then

\[ S_{\DeltaX \DeltaX}^{+} \triangleq \begin{bmatrix} \Gamma_1 \nu' \\ \gamma^{22} \delta_{\DeltaX \DeltaX}^{12} + \Gamma_1 \nu' \end{bmatrix} \]

Asymptotic Distribution of the Smaller Roots

Let \( Q^{+} = S_{\DeltaX \DeltaX}^{-1} S_{\DeltaX \DeltaX} S_{\DeltaX \DeltaX}^{-1} \). The \( n \) smaller roots of \( |Q^{+} - t^{2} S_{\DeltaX \DeltaX}^{\delta}| = 0 \) converge in probability to 0 and the \( k \) larger roots converge to the roots of

\[ \left| S_{\DeltaX \DeltaX}^{+} - \Sigma_{\DeltaX \DeltaX}^{+} \right|^{2k} - r^{2} S_{\DeltaX \DeltaX}^{\delta} = 0 \]

by the analysis of ref. 2. Let \( d = T^{2} \). The \( n \) smaller roots of \( (Q_{\DeltaX \DeltaX}^{+} - T^{2} S_{\DeltaX \DeltaX}^{\delta}) \) converge in distribution to the roots of

\[ 0 = \left( \Gamma_{1}^{11}(\Sigma_{\DeltaX \DeltaX}^{11})^{-1} \Gamma_{1} + d \Gamma_{1}^{11} \Gamma_{1}' \right) \Gamma_{1}^{11}(\Sigma_{\DeltaX \DeltaX}^{11})^{-1} \Gamma_{1} - d_{1} \]

by the algebra used in ref. 2, section 5, resulting in the same distribution as in ref. 2. The likelihood ratio criterion for testing that the rank of \( Y \) is \( k \) (2) is \( -2 \log \lambda = -T \sum_{i=1}^{n} (1 - r_{i}^{2}) = \sum_{i=1}^{n} d_{i} + o_{p}(1) \), the limiting distribution of which was found by Johansen (4, 6) and was given in equation 5.4 in ref. 2.

Asymptotic Distribution of the Larger Roots

We now turn to deriving the asymptotic distribution of the \( k \) larger roots of \( |Q^{+} - t^{2} S_{\DeltaX \DeltaX}^{\delta}| = 0 \) and the associated vectors \( |Q^{+} - t^{2} S_{\DeltaX \DeltaX}^{\delta}| = 0 \) and the associated vectors solving \( Q^{+} = t^{2} S_{\DeltaX \DeltaX} \). First we show that the asymptotic distribution of \( r_{\DeltaX \DeltaX}^{+} \), \( \ldots \), \( r_{\DeltaX \DeltaX}^{+} \) is the same as the asymptotic distribution of the zeros of \( Q_{\DeltaX \DeltaX}^{+} \). Then we transform the \( \DeltaX \) coordinates to the coordinates of canonical correlations and vectors.

Let \( \hat{R}_{\DeltaX}^{2} = \text{diag}(r_{\DeltaX \DeltaX}^{+}, \ldots, r_{\DeltaX \DeltaX}^{+}) \) and \( \hat{G}_{\DeltaX} = (\hat{G}_{\DeltaX}, \DeltaX \DeltaX \hat{G}_{\DeltaX} \DeltaX \DeltaX) \) consist of the corresponding solutions to \( Q^{+} \hat{G}_{\DeltaX} = S_{\DeltaX \DeltaX} \hat{G}_{\DeltaX} \). The normalization of the columns of \( \hat{G}_{\DeltaX} \) is \( \hat{G}_{\DeltaX} \hat{G}_{\DeltaX} \hat{G}_{\DeltaX} = I \), that is,

\[ \hat{I} = \left( \sqrt{T} G_{\DeltaX \DeltaX}^{12}, \DeltaX \DeltaX \hat{G}_{\DeltaX}^{22} \right) \begin{bmatrix} \frac{1}{T} S_{\DeltaX \DeltaX}^{11} & \frac{1}{T} S_{\DeltaX \DeltaX}^{12} \\ \frac{1}{T} S_{\DeltaX \DeltaX}^{21} & S_{\DeltaX \DeltaX}^{22} \end{bmatrix} \begin{bmatrix} \sqrt{T} G_{\DeltaX \DeltaX}^{12} \\ \DeltaX \DeltaX \hat{G}_{\DeltaX}^{22} \end{bmatrix} \] \[ \text{[21]} \]

The probability limit of 21 shows that \( \sqrt{T} G_{\DeltaX \DeltaX}^{12} = O_{p}(1) \) and \( \hat{G}_{\DeltaX}^{22} = O_{p}(1) \). The submatrix equations in \( Q^{+} \hat{G}_{\DeltaX} = S_{\DeltaX \DeltaX} \hat{G}_{\DeltaX} \) can be written as

\[ \frac{1}{T} Q_{11}^{+} \sqrt{T} G_{\DeltaX \DeltaX}^{12} + \frac{1}{T} Q_{21}^{+} \sqrt{T} G_{\DeltaX \DeltaX}^{22} = \left( \frac{1}{T} S_{\DeltaX \DeltaX}^{11} \sqrt{T} G_{\DeltaX \DeltaX}^{12} + \frac{1}{T} S_{\DeltaX \DeltaX}^{12} \hat{G}_{\DeltaX}^{22} \right) \hat{R}_{\DeltaX}^{2}, \]

\[ \frac{1}{T} Q_{21}^{+} \sqrt{T} G_{\DeltaX \DeltaX}^{12} + \frac{1}{T} Q_{22}^{+} \hat{G}_{\DeltaX}^{22} = \left( \frac{1}{T} S_{\DeltaX \DeltaX}^{21} \sqrt{T} G_{\DeltaX \DeltaX}^{12} + \frac{1}{T} S_{\DeltaX \DeltaX}^{22} \hat{G}_{\DeltaX}^{22} \right) \hat{R}_{\DeltaX}^{2}. \]

Because \( T^{-1} Q_{11}^{+} \rightarrow 0, T^{-1/2} Q_{12}^{+} ightarrow 0, T^{-1/2} Q_{12}^{+} ightarrow 0, T^{-1/2} Q_{12}^{+} \rightarrow 0, \) \( \Gamma_{1}^{11} \Gamma_{1}' \Gamma_{1} \rightarrow \Gamma_{1}^{11} \Gamma_{1}' \Gamma_{1} \), and \( \hat{R}_{\DeltaX}^{2} \rightarrow \hat{R}_{\DeltaX}^{2} = \text{diag}(p_{1}^{2} + \ldots, p_{k}^{2}) \), the probability limit of the left-hand side of 22 is 0; this shows that \( \sqrt{T} G_{\DeltaX \DeltaX}^{12} = O_{p}(1) \). Then the asymptotic distribution of \( \hat{G}_{\DeltaX}^{22} \) is the asymptotic distribution of \( G_{\DeltaX}^{22} \) defined by

\[ Q_{\DeltaX \DeltaX}^{+} G_{\DeltaX}^{22} = S_{\DeltaX \DeltaX}^{22} \hat{G}_{\DeltaX}^{22} \hat{R}_{\DeltaX}^{2}, \quad \hat{G}_{\DeltaX}^{22} S_{\DeltaX \DeltaX}^{22} G_{\DeltaX}^{22} = I \]

where the elements of \( \hat{R}_{\DeltaX}^{2} \) are defined by \( Q_{\DeltaX \DeltaX}^{+} - r^{2} S_{\DeltaX \DeltaX}^{22} = 0 \). Note that when \( \sqrt{T} G_{\DeltaX \DeltaX}^{12} \rightarrow 0 \) is combined with 23, we obtain \( Q_{\DeltaX \DeltaX}^{+} G_{\DeltaX}^{22} = S_{\DeltaX \DeltaX}^{22} \hat{G}_{\DeltaX}^{22} \hat{R}_{\DeltaX}^{2} + o_{p}(T^{-1/2}) \).

We proceed to find the asymptotic distribution of \( G_{\DeltaX}^{22} \) and \( \hat{R}_{\DeltaX}^{2} \) defined by 24 in the manner of ref. 2. Let

\[ S_{\DeltaX \DeltaX}^{+} = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} J_{11}^{\DeltaX} - \frac{1}{\sqrt{T}} \sum_{i=1}^{T} J_{11}^{\DeltaX} \]

by the algebra used in ref. 2, section 5, resulting in the same distribution as in ref. 2. The likelihood ratio criterion for testing that the rank of \( Y \) is \( k \) (2) is \( -2 \log \lambda = -T \sum_{i=1}^{n} (1 - r_{i}^{2}) = \sum_{i=1}^{n} d_{i} + o_{p}(1) \), the limiting distribution of which was found by Johansen (4, 6) and was given in equation 5.4 in ref. 2.
\[ \ell \vec{S}_{XX}^{+2}(\vec{S}_{XX}^{+2})^* = (I + \mathbf{K})(C \otimes C) \]
\[ + \left\{ (I - (\vec{Y} \otimes \vec{Y})^{-1}) \left( \Sigma \otimes \Sigma \right) \right\} [(I - (\vec{Y} \otimes \vec{Y})^{-1})(\Sigma \otimes \Sigma)]^{-1} - (\Sigma \otimes \Sigma)(C' \otimes C') \]. \]  
\[ \text{Lemma 1.} \] If \( X \) is normally distributed with \( \mathcal{E}X = 0 \) and \( \mathcal{E}XX' = \Sigma \), then \( \mathcal{E}xx' = \mathcal{E}x'x = (I + \mathbf{K})(\Sigma \otimes \Sigma) + \sigma \Sigma(\vec{X}(\vec{X})' - \mathcal{E}x'x) \). If \( X \) and \( Y \) are independent, \( \mathcal{E}xx'(\vec{Y}') = \mathcal{E}xx' \otimes (\vec{Y}' \vec{Y})' = \mathcal{E}xx' \otimes \mathcal{E}yy' \otimes \mathcal{E}xx', \) and \( \mathcal{E}vec(YY') = \mathcal{E}vec(XX') \).

\[ \text{Proof of Theorem 1:} \] First 26 is equivalent to the first expression in Lemma 1. Next vec \( \mathcal{E}vec(S_{YY}^{+2}x) = T^{-1}S_{YY}^{+2}x \) implies 27 because \( X_{2,2-t} \) and \( W_{2,2} \) are independent for \( t = 1 \). Similarly 28 follows. To prove 29, 30, and 31, we use the following lemma.

\[ \text{Lemma 2.} \] vec \( \mathcal{E}vec(W_{2,2}^{+2}(vec S_{XX}^{+2})) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \).

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r \left( \frac{1}{\sqrt{T}} \right). \]

\[ \text{Proof of Lemma 2:} \] We have from \( \mathcal{E}vec(W_{2,2}^{+2}(vec S_{XX}^{+2})) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \).

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r \left( \frac{1}{\sqrt{T}} \right). \]

\[ \text{Proof of Lemma 3:} \] Write \( W_{2,2} = W_{2,2}^{+2}(vec S_{XX}^{+2}) \).

\[ \text{Then} \]

\[ \mathcal{E}vec(W_{2,2}^{+2}(vec S_{XX}^{+2})) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r \left( \frac{1}{\sqrt{T}} \right). \]

\[ \text{Proof of Theorem 1 Continued:} \] Then 29 follows from Lemma 2, 26, and 28, and 30 follows from Lemma 2, 27, and 28. To prove 31, use 26, 27, 28, and 29.

\[ \mathcal{E}vec(S_{XX}^{+2}(vec S_{XX}^{+2})) = (C \otimes C)(I - (\vec{Y} \otimes \vec{Y})^{-1}) \]

\[ + (I + \mathbf{K})(\vec{Y} \otimes \vec{Y})(\vec{X} \otimes \vec{X})(I + \mathbf{K}) \]

\[ + (I + \mathbf{K})(\vec{Y} \otimes \vec{Y})(I - (\vec{Y} \otimes \vec{Y})^{-1})(C' \otimes C') \]

\[ + (\vec{X} \otimes \vec{X})(I - (\vec{Y} \otimes \vec{Y})^{-1})(C' \otimes C'). \]

Then substituting \( \vec{X} \otimes \vec{X} = \mathcal{E} - \vec{Y} \vec{Y}' \) in 33 yields 31. Let \( \mathcal{E} \) be a k x k matrix such that \( \mathcal{E}(\mathcal{E}^{+2}(\mathcal{E}^{+2})) = \mathcal{E} \) and \( \mathcal{E}^{+2} = I \) where \( \mathcal{E} = \operatorname{diag}(\theta_{k+1}, \ldots, \theta_k) \).

\[ \mathcal{E} = \operatorname{diag}(\rho^2_1, \ldots, \rho^2_d), \text{ and } \rho^2_k \text{ is a root of } 18 \text{ with } 0 < \rho^2_{k+1} < \cdots < \rho^2_1. \]

Let \( U_{2,2} = \mathcal{E}X_{2,2}^\prime, V_{2,2} = \mathcal{E}X_{2,2}, \) and \( \mathcal{E} = \mathcal{E}(Y^{+2}(I + 1)(E)^{-1}, \) M_1 = \mathcal{E}Y^{+2}(E)^{-1}, \) and \( \mathcal{E} = \operatorname{diag}(\mathcal{E}, \mathcal{E}_{1,1} \otimes \mathcal{E}_{1,1} \otimes \mathcal{E}_{1,1} \otimes \mathcal{E}_{1,1} \otimes \mathcal{E}_{1,1}) \), \( \Lambda = \mathcal{E}(Y^{+2}(E)^{-1}, \) U_1 = \mathcal{E}X_{1,1}, C_2 = \mathcal{E}(C(E)^{-1}). \)

Then \( \{U_1\} \) is generated by \( \mathcal{U}_1 = \mathcal{U}_1 + V_1 \), where \( \mathcal{V}_1 = \mathcal{E}X_{1,1} \) and \( \mathcal{U}_2 \) satisfies \( \mathcal{U}_{2,2} = U_{2,2} + V_{2,2} + U_{2,2} + V_{2,2} + \mathcal{U}_{2,2} + V_{2,2} + \mathcal{U}_{2,2} + V_{2,2} \).

Multiplication of 25 on the left by \( \mathcal{E} \) and right by \( \mathcal{E} \) yields

\[ \mathcal{E}vec(S_{XX}^{+2})(vec S_{XX}^{+2}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{Theorem 2.} \] If the \( V_i \) are independently normally distributed, \( S_{VV}^{+2,1,2}, S_{VV}^{+2,2,1}, \) and \( S_{UU}^{+2,1,2} \) have a limiting normal distribution with means \( 0, 0, \) and 0 and covariances

\[ \mathcal{E}vec(S_{VV}^{+2,1,2})(vec S_{VV}^{+2,1,2}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{Then} \]

\[ \mathcal{E}vec(S_{VV}^{+2,2,1})(vec S_{VV}^{+2,2,1}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{Let} \]

\[ \mathcal{E}vec(S_{UU}^{+2,2})(vec S_{UU}^{+2,2}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{Then} \]

\[ \mathcal{E}vec(S_{LL}^{+2,2})(vec S_{LL}^{+2,2}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{The covariances of the limiting normal distribution of vec} \]

\[ \mathcal{E}vec(S_{VV}^{+2,1,2}), \text{vec} S_{VV}^{+2,2,1}, \text{and vec} S_{UU}^{+2,2} = (M_2 \otimes M_2) \vec{S}_{UU}^{+2,2} \text{ are found from Theorem 2. We write the transform of 35 as} \]

\[ \mathcal{E}vec(S_{LL}^{+2,2})(vec S_{LL}^{+2,2}) = (I + \mathbf{K})(\vec{Y} \otimes \vec{Y}) \]

\[ + vec \mathcal{E}vec(W_{2,2}) + o_r(1). \]

\[ \text{where} \]

\[ \Phi^+ = (M_2 C_t \otimes M_2 C_t)(I - (\mathcal{E} \otimes \mathcal{E}))(I, M_{1,1} \otimes M_{1,1}). \]
Let $H_{22} = (M_2)'^{-1}\Xi^{-1}I_2G_2 = \Xi^{-1}(Y^{22})^{-1}G_2$. Then $Q_{22}G_{22} = \Sigma_{22}^2G_{22}R_2^2$ and $G_{22}\Sigma_{22}^2G_{22} = I$ transform to

$$(S_{LL}^+M_{LL}^+\Sigma_{LL}^+M_{LL}^+)H_{22} = S_{LL}^+M_{LL}^+R_2^2, \quad H_{22}S_{LL}^+M_{LL}^+ = I. \quad [38]$$

Because $(S_{LL}^+M_{LL}^+\Sigma_{LL}^+M_{LL}^+)G_{22} = \Theta_2^2$, the probability limits of 38 and $h_2 > 0$ imply $H_{22} \sim \Theta_2^{-1/2}$.

Define $\Theta_2 = \sqrt{\vec{I}}(H_{22} - \Theta_2^{-1/2})$ and $\tilde{R}_2^2 = \sqrt{\vec{I}}(R_2^2 - \tilde{R}_2^2)$. Then we can write 38 as

$$\Theta_2^{-1/2} \Theta_2^{-1/2} + \phi_2(1), \quad [39]$$

$$H_{22}^{-1/2} \Theta_2 \Theta_2^{-1/2} + \Theta_2^{-1/2} \tilde{R}_2^2 + \Theta_2^{-1/2} + \phi_2(1), \quad [40]$$

where

$$P = -R_2S_{LL}^+M_{LL}^+S_{LL}^+M_{LL}^+R_2 + R_2S_{LL}^+M_{LL}^+\tilde{R}_2^2 + R_2S_{LL}^+M_{LL}^+\tilde{R}_2^2 + R_2S_{LL}^+M_{LL}^+R_2 - (I - R_2). \quad [41]$$

**Lemma 4.**

$$\mathcal{E}(I + K)(I - R_2) \otimes R_2^2 \vec{s}_{LL}^{2+1,1}$$

$$= [I - R_2] \otimes \vec{R}_2^2[I + K]$$

$$= -[R_2(I - R_2) \otimes \vec{R}_2^2][I + K] + \mathcal{E}(I - R_2) \otimes \vec{R}_2^2[I + K]. \quad [42]$$

**Proof of Lemma 5:** We use the facts that $M_2 = \Delta_2 - I$, $J_2M_2 = \Delta_2 - I$, and $I + R_2 = I + K + R_2$. Then the left-hand side of 42 is

$$[(I - R_2) \otimes \vec{R}_2^2][I + K](M_2C_2 \otimes M_2C_2)[(I - \Delta) \otimes \Delta]^{-1}$$

$$= [(I - R_2) \otimes \vec{R}_2^2][I + K](M_2C_2 \otimes M_2C_2)$$

$$= [I - R_2 \otimes \vec{R}_2^2][I + K](M_2C_2 \otimes M_2C_2)$$

$$= [I - \Delta \otimes \Delta][I - (I - R_2 \otimes \vec{R}_2^2)][I + K](M_2C_2 \otimes M_2C_2)$$

$$= -[I - R_2 \otimes \vec{R}_2^2][I + K](M_2C_2 \otimes M_2C_2)$$

which is the right-hand side of 42.

**Theorem 3.** If the $Z_i$ vectors are normally distributed and the roots of 18 are distinct,

$$\mathcal{E} vec P(\mathcal{E}(I - R_2) \otimes \vec{R}_2^2) = \mathcal{E} vec P(\Theta_2 \otimes \vec{R}_2^2) \approx \mathcal{E} vec P(\Theta_2 \otimes \vec{R}_2^2 \otimes \vec{R}^2 \Theta)$$
